MATHEMATICA Policy Research

REPORT

FINAL REPORT—VOLUME I

# Background Paper

# Declining Response Rates in Federal Surveys: Trends and Implications

June 15, 2016

John L. Czajka

Amy Beyler

**Submitted to:**

Office of the Assistant Secretary for Planning and Evaluation

U.S. Department of Health & Human Services

200 Independence Avenue, S.W.

Washington, DC 20201

Project Officer: Amanda Cash

ASPE Task Leaders: Rachel A. Hornstein and Rashida Dorsey

Contract Number: HHSP23320095642WC/HHSP23337039T

**Submitted by:**

Mathematica Policy Research

1100 1st Street, NE

12th Floor

Washington, DC 20002-4221

Telephone: (202) 484-9220

Facsimile: (202) 863-1763

Project Director: Deborah Chollet

Reference Number: 40146.D4C

This report was prepared by Mathematica Policy Research under a contract for ASPE. The findings and conclusions are those of the authors and do not necessarily represent the views of ASPE or the U.S. Department of Health and Human Services.

# CONTENTS

## TABLES



## FIGURES

**This page has been left blank for double-sided copying.**

# EXECUTIVE SUMMARY

Household sample surveys depend upon their ability to reach and engage the individuals they select as potential respondents. The viability of such surveys is being challenged by declining response rates and related developments that affect not only the quantity but the quality of the information collected.

The Department of Health and Human Services (HHS) is the leader among federal agencies in the breadth and volume of the household surveys it sponsors. Recognizing the challenges facing departmental surveys, the Office of the Assistant Secretary for Planning and Evaluation (ASPE) is convening a Technical Expert Panel (TEP) on March 18, 2016, to examine the scope of the problem, discuss what the survey sponsors have found to be the most effective approaches to maintaining robust response rates and achieving representativeness in their surveys, and consider future steps.

This document was prepared to provide critical background for the TEP meeting. It reviews recent assessments of trends in response rates among primarily federal surveys and the reasons for declining response rates; documents response trends in the major HHS surveys from the mid-1990s through the most recent year available; examines what is known about the relationship between response rates and nonresponse bias; reviews approaches to addressing nonresponse and its effects; and summarizes key conclusions.

A recent report by the Committee on National Statistics of the National Academy of Sciences documented declines in response rates to household surveys since the late 1990s. Declines were not universal, and some surveys experienced greater declines than others, but the phenomenon of declining responses is sufficiently widespread that it has generated growing concerns about the potential for nonresponse bias.

While this review focuses on unit nonresponse, nonresponse to individual survey questions is a concern as well. Trends in item nonresponse are not nearly as well documented as trends in unit nonresponse, but it is readily apparent that item nonresponse among those who respond to surveys is increasing as unit response rates decline.

A common way of classifying the many reasons for nonresponse to a survey uses three categories: (1) non-contact, (2) refusal, or (3) other. The "other" category encompasses such reasons as language issues, poor health, and absence from home during the data collection period. Multiple societal and environmental factors have been suggested as contributing to the decline in survey response rates, either by making it more difficult to contact prospective respondents or by increasing the likelihood that, if contacted, they will refuse to participate. When non-interviews are apportioned among non-contacts, refusals, and other reasons, there is evidence that the distribution among these factors has changed little with growing nonresponse.

In general, the increase in nonresponse has been much greater in telephone surveys than in face-to-face surveys, which is consistent with the growing number of solicitations that households receive by telephone and the increasing use of answering machines and caller ID to screen calls. The observation that noncontacts do not appear to have grown as a share of nonresponse has led some researchers to suggest that the impact of technological barriers may be

less on preventing contact than on keeping the interviewer from fully communicating the survey request.

To document the trends in response rates to HHS surveys, we compiled trend data for seven surveys sponsored by the department. The seven surveys are national in scope and include those for which response rate trend data are most readily available. The seven surveys are:

- The National Health Interview Survey (NHIS), sponsored by the National Center for Health Statistics (NCHS) within the Centers for Disease Control and Prevention (CDC)
- The National Health and Nutrition Examination Survey (NHANES), also sponsored by NCHS
- The Behavioral Risk Factors Surveillance System (BRFSS), sponsored by CDC
- The National Immunization Survey (NIS), sponsored by the National Center for Immunization and Respiratory Diseases within CDC
- The National Survey of Drug Use and Health (NSDUH), sponsored by the Substance Abuse and Mental Health Services Administration (SAMHSA)
- The Medicare Current Beneficiary Survey (MCBS), sponsored by the Centers for Medicare & Medicaid Services (CMS)
- The Medical Expenditure Panel Survey (MEPS), sponsored by the Agency for Healthcare Research and Quality (AHRQ)

For comparison, we also included the Current Population Survey Annual Social and Economic Supplement (CPS ASEC), which is co-sponsored by the Bureau of Labor Statistics (BLS) and the U.S. Census Bureau, which performs the data collection.

While the response rates and trends differ, all eight surveys show some amount of decline in their response rates over the first half of the period, and six of the surveys—all but BRFSS and the NIS—exhibit accelerated rates of decline in the most recent years. Most striking is the NHIS, for which the response to the household module declined from about 92 percent in 1997 to about 74 percent in 2014, with most of that decline—13 percentage points—occurring in the last seven years.

In their review of declining response rates and their potential causes, Brick and Williams (2013) begin with the observation that "nonresponse in household surveys has been a topic of great interest for many years because the foundation of inference from probability samples is based on observing all the sampled units." In essence, anything less than a 100 percent response rate threatens the validity of statistical inference unless nonresponse is itself completely random. Response rates approaching 100 percent have been achieved by some of the major federal surveys, but response rates for even these surveys have fallen in the past two decades—in some cases dramatically. Other federal surveys have never achieved response rates approaching 100 percent; the declines in their response rates would seem to warrant even greater concern.

Understanding of the importance of the survey response rate as an indicator of potential bias has evolved as response rates have declined. This evolution has been driven in part by studies of

selected surveys showing no increase in response bias with declining response rates. A much larger body of research has demonstrated that the relationship between nonresponse and nonresponse bias is complex. This research has also altered thinking about the prospective respondent and this individual's inclination to respond to a given survey.

The current view of prospective respondents is that each one has a propensity to respond, which is influenced by personal characteristics and the "circumstances" of the survey, and can be changed—increased or decreased—by actions taken by the field staff and the survey sponsors. Across the sample, if the propensity to respond is correlated with any variable measured in the survey, then nonresponse will contribute to biased estimates. The higher the correlation, the greater the bias introduced by a given level of nonresponse.

Summarizing a meta-analysis of studies providing estimates of both nonresponse and nonresponse bias, Groves (2006) drew the following conclusions. First, there is no necessary connection between the nonresponse rate and nonresponse bias. Second, there is no minimum response rate below which a survey estimate is necessarily biased and no response rate above which it is never biased. Third, bias can vary across statistics in the same survey. A more extensive meta-analysis by Groves and Peytcheva (2008) underscored the earlier finding that the overall level of survey response, by itself, is not a good indicator of the overall or average level of nonresponse bias in a survey. Other conclusions included the fact that large nonresponse biases can occur in surveys and that the search for mechanisms linking nonresponse rates and nonresponse bias needs to focus on individual measures and not the survey level. In addition, there are multiple ways that nonresponse bias can be estimated, and how this is done can affect not only the estimated level of bias, on average, and how it affects individual characteristics.

Efforts to add respondents in order to increase response rates may not reduce bias and could actually increase it, depending on whether the additional respondents are more representative or less representative of sample members who are underrepresented among the existing respondents. There is also evidence that reluctant respondents may provide data of lower quality than respondents who participated more willingly.

While it has been shown that efforts to increase survey response rates do not guarantee a reduction in nonresponse bias, low response rates challenge the credibility of surveys, inviting efforts to increase them or least stem the tide of decline. A number of strategies are being used—or have been recommended for expanded use—in an effort to maximize survey response rates. These include the payment of incentives, reducing survey burden, the use of address-based sampling in combination with a mail survey mode, the use of multiple modes within the same survey, double or two phase sampling, and responsive design. Among strategies to maximize response rates, the payment of incentives has been shown to increase response rates most consistently. Reducing survey burden has less clear implications and consequences. A strategy for reducing questionnaire length is to divide major parts of the survey into segments and assign only one segment to each respondent—called matrix sampling. Sampling from lists of mailing addresses has developed as a viable and cost effective tool for conducting surveys by mail as the quality of available lists has improved. Mail surveys avoid some of the issues with in-person and telephone surveys but there is evidence that response rates decline with educational attainment. Mixed-mode designs seek to take advantage of differences across subpopulations in the likelihood of responding to alternative modes. Two phase sampling involves expending a low

effort to obtain an initial round of responses, subsampling the nonrespondents, and applying more intense methods to obtain a high response rate among the subsample. This approach is illustrated by the American Community Survey, which also employs mixed modes in sequence. Responsive design is a relatively new approach that involves monitoring key outcomes as data collection is in process and modifying aspects of the data collection—such as where to allocate contact attempts—to achieve particular response goals or cost control.

Data that are available for both respondents and non-respondents are invaluable in estimating nonresponse bias and making compensating adjustments to the survey weights. Administrative records and data contained in the survey frame provide two potential sources. Another data source attracting considerable interest is paradata, or data generated as a byproduct of the survey operations. Paradata include information on call attempts and outcomes and well as interviewer observations about the sample units. The latter have potential value for nonresponse adjustments.

Declining response rates encourage consideration of alternatives to reliance on probability surveys to address all of the HHS data needs. Strategies that have been suggested include making greater use of administrative records, learning how to extract useful information from "big data," and developing better ways to use the data obtained from nonprobability samples.

In summary, concerns about growing survey nonresponse and its potential impact upon nonresponse bias are not new, but federal household surveys have tended to enjoy high response rates until relatively recently. Response rate trends presented in this report highlight increased rates of decline among the major federal surveys in very recent years. They underscore the fact that declining response rates are not unique to particular surveys; they are widespread, although some surveys have experienced much greater declines than others.

Much has been learned about the nature of nonresponse bias. Two important findings are that nonresponse bias is generally unrelated to the overall response rate and that nonresponse bias tends to be item-specific. Some or even many items may exhibit no bias while others have substantial bias.

# I. INTRODUCTION

Household sample surveys depend upon their ability to reach and engage the individuals they select as potential respondents. The viability of such surveys is being challenged by declining response rates and related developments that affect not only the quantity but the quality of the information collected.

Concerns about survey nonresponse are hardly new. Such concerns underlay the founding of the International Workshop on Survey Nonresponse in 1990, and in 1992 Norman Bradburn devoted his presidential address to the American Association for Public Opinion Research (AAPOR) to the topic of growing survey nonresponse and what the profession could do about it (Bradburn 1992). An international conference on survey nonresponse was held in October 1999 (see Groves et al. 2002), and the *Journal of Official Statistics* in 1999 and 2001 and *Public Opinion Quarterly* in 2006 devoted special issues to the topic of survey nonresponse. In 2010, the Committee on National Statistics (CNSTAT) of the National Academy of Sciences assembled an expert panel to review a related set of issues. Groundwork for the Panel on a Research Agenda for the Future of Social Science Data Collection was laid at an all-day planning meeting that was held in Washington on December 14, 2009. The meeting featured two commissioned papers and a number of supplemental presentations. In addition, the panel conducted two workshops in early 2011 that included a number of participants from federal agencies, academia, and private research organizations. Papers from the planning meeting and the first workshop were published in a special issue of The Annals of the American Academy of Political and Social Science entitled, "The Nonresponse Challenge to Surveys and Statistics." The panel's final report, *Nonresponse in Social Science Surveys: A Research Agenda*, was published in 2013. Independently, the Census Bureau sponsored a CNSTAT workshop on the

Future of Federal Household Surveys, which discussed a number of challenges facing such surveys currently and in the near future (NRC 2011). The workshop was motivated by a concern among many in the federal statistical community that present methods of data collection may be unsustainable.

The Department of Health and Human Services (HHS) is the leader among federal agencies in the breadth and volume of the household surveys it sponsors. Recognizing the challenges facing departmental surveys, the Office of the Assistant Secretary for Planning and Evaluation (ASPE) is convening a Technical Expert Panel (TEP) on March 18, 2016, to examine the scope of the problem, discuss what the survey sponsors have found to be the most effective approaches to maintaining robust response rates and achieving representativeness in their surveys, and consider future steps.

This document was prepared to provide critical background for the TEP meeting. Chapter II reviews recent assessments of trends in response rates among primarily federal surveys and the reasons for declining response rates. Chapter III documents response trends in the major HHS surveys from the mid-1990s through the most recent year available and in so doing shows that in most cases the rate of decline in response rates has increased in the past few years. Chapter IV examines what is known about the relationship between response rates and nonresponse bias. Chapter V reviews approaches to addressing nonresponse and its effects, and Chapter VI summarizes key conclusions.

# II. SURVEY RESPONSE RATES: SETTING THE CONTEXT

This chapter begins with a review of how survey response is—or should be—measured. From there we proceed to a review of prior findings on recent trends in survey response rates among what are primarily federal surveys. The chapter closes with a discussion of possible reasons for declining survey response rates.

## A. Measuring survey response

Conceptually the response rate to a survey is the number of interviews divided by—or expressed as a percentage of—the number of units (or cases) eligible to be interviewed. While this concept is straightforward, there are challenges in applying it to an actual survey. For the numerator, respondents may answer some but not all of the questions; what, then constitutes an interview? For the denominator, the eligibility of some sample units that were not interviewed may not have been determined; how should these be treated?

The calculation of response rates begins with the assignment of disposition codes. A disposition code indicates the final outcome for a unit in the sample. To calculate a response rate, every sample unit must be assigned a disposition code. In an effort to standardize the assignment of disposition codes and introduce uniform practices in the calculation of survey response rates, AAPOR published a guide in 1998, *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys.* The AAPOR document was an extension of a 1982 report, *On the Definition of Response Rates*, issued by the Council of American Survey Research Organizations (CASRO). *Standard Definitions* has been revised and expanded numerous times since its initial publication; the 8th edition was published in April 2015 (AAPOR 2015). Survey organizations and other entities that conduct surveys are encouraged to use the disposition codes and formulas in *Standard Definitions* to calculate their response rates and to report which formula they used.

The AAPOR document provides a somewhat different set of disposition codes for each of four survey modes and universes: (1) random digit dialed (RDD) telephone surveys of households, (2) in-person household surveys, (3) mail surveys of specifically named persons, and (4) Internet surveys of specifically named persons (AAPOR 2015). While the set of disposition codes varies by mode, each code belongs to one of four main groups: (1) interviews, (2) eligible cases not interviewed (that is, non-respondents), (3) cases of unknown eligibility, and (4) cases not eligible. The AAPOR response rate formulas are expressed in terms of these four groups, but with interviews divided between complete and partial.[1]

AAPOR gives six alternative response rate formulations, each of which consists of the number of interviews divided by the number of eligible cases. The six response rates, shown in Table II.1, differ along two dimensions: (1) whether partial interviews are treated as interviews or as non-interviews and (2) how cases of unknown eligibility are handled. Response rate 1 (RR1) includes only complete interviews in the numerator while placing all cases of unknown eligibility in the denominator (along with both types of interviews and all eligible cases not interviewed). RR2 adds partial interviews to the numerator. RR3 and RR4 reduce the size of the denominator by including only the fraction of cases of unknown eligibility estimated to be eligible. This fraction can be estimated in different ways—for example, as the ratio of known eligible cases to the sum of known eligible and known ineligible cases. RR3 includes only complete interviews in the numerator while RR4 includes both complete and partial interviews. Of these first four, then, RR4 yields the highest response rate, and RR1 yields the lowest. The final two AAPOR response rates, RR5 and RR6, are used when there are no ineligible cases (or

[1] Examples of disposition codes within the other groups include refusals and noncontacts among eligible non-interviews; not attempted or an unsafe area among cases of unknown eligibility; and a business or vacant unit among cases not eligible.

when it is assumed that there are none). Again, RR5 includes only complete interviews in the numerator, and RR6 includes both complete and partial interviews. RR5 can be viewed as a special case of RR3, where the eligibility rate among cases of unknown eligibility is zero. Likewise, RR6 is a special case of RR4 and the lowest of the six AAPOR response rates.

**Table II.1. AAPOR response rates**

| Response Rate Formula | Description |
|---|---|
| $RR1 = \frac{I}{(I+P)+(R+NC+O)+(UH+UO)}$ | RR1 is the minimum response rate calculated as the number of complete interviews divided by the number of interviews plus the number of non-interviews plus all cases of unknown eligibility. |
| $RR2 = \frac{(I+P)}{(I+P)+(R+NC+O)+(UH+UO)}$ | RR2 includes partial interviews as respondents in the numerator and is equivalent to RR1 in the denominator. |
| $RR3 = \frac{I}{(I+P)+(R+NC+O)+e(UH+UO)}$ | RR3 applies an assumed eligibility rate to the cases of unknown eligibility. |
| $RR4 = \frac{(I+P)}{(I+P)+(R+NC+O)+e(UH+UO)}$ | RR4 includes partial interviews as respondents in the numerator and is equivalent to RR3 in the denominator. |
| $RR5 = \frac{I}{(I+P)+(R+NC+O)}$ | RR5 either assumes no cases of unknown eligibility are eligible (e=0) or there are no cases of unknown eligibility. |
| $RR6 = \frac{(I+P)}{(I+P)+(R+NC+O)}$ | RR6 is the maximum response rate which includes partial interviews as respondents in the numerator and is equivalent to RR5 in the denominator. |

Source: The American Association for Public Opinion Research. 2015. *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys. 8th edition.* AAPOR.

RR = Response rate; I = Complete interview; P = Partial interview; R = Refusal and break-off; NC = Non-contact; O = Other; UH = Unknown if household/occupied HU; UO = Unknown, other; e = Estimated proportion of cases of unknown eligibility that are eligible.

Some surveys, including several of those examined in the next chapter, restrict the eligible persons within a household to those with certain characteristics (for example, by age) and may even sample the eligible individuals to identify a target respondent. In such multi-stage surveys, nonresponse may occur at either stage. That is, the household may respond to the initial, screening interview, but the selected respondent may refuse or otherwise fail to complete the

interview. Alternatively, the household may fail to respond to the screening interview, in which case no respondent is selected for the second stage. In a two-stage survey such as this one, three response rates can be calculated: (1) the household or screener response rate, (2) the response rate of the individual selected through the screener, conditional on having been selected, and (3) a final or unconditional response rate, representing the number of interviews in the second stage divided by the number of persons who would have been eligible to be interviewed if all households had been successfully screened. This last, unconditional response rate is commonly estimated as the product of the first two response rates. For example, if 90 percent of sample households were screened and 80 percent of the selected respondents were successfully interviewed, the unconditional response rate would be 72 percent. As AAPOR (2015) points out, however, this calculation assumes implicitly that the eligibility rate among the unscreened households is the same as the eligibility rate among screened households. AAPOR recommends that this assumption be tested, if possible.

## B. Prior findings on trends in survey response rates

Our review of prior findings on trends in response rates among major surveys begins with cross-sectional surveys, then moves to panel surveys and closes with a discussion of trends in item nonresponse.

### 1. Cross-sectional surveys

Brick and Williams (2013) reviewed response trends in four cross-sectional surveys beginning in the late 1990s and extending through the late 2000s:

- National Health Interview Survey (NHIS)
- The General Social Survey (GSS)
- The National Household Education Survey (NHES)
- The National Immunization Survey (NIS)

The CNSTAT panel summarized their findings and added response rate trends for three more surveys: the Behavioral Risk Factor Surveillance System (BRFSS), the Survey of Consumer Attitudes (SCA), and the Survey of Consumer Finances (SCF). The first two surveys in the list (NHIS and GSS) and the SCF utilize face-to-face interviewing while the other four are conducted by telephone. Also important, the NHIS, NHES, NIS, BRFSS, and SCF are sponsored by the federal government while the GSS and SCA are not. We discuss the NHIS, BRFSS, and NIS in the next chapter and in the Appendix, where we present updated response rates, so we will not discuss them here.

Citing Smith (1995), the CNSTAT report observes that GSS response rates held steady from 1975 to 1993, when they peaked at 82.4 percent. Response rates then dropped into the upper 70s, staying there through 1998 before declining to 70 percent in 2000, where they remained through at least 2006 (Brick and Williams 2013).[2]

The NHES, which is sponsored by the National Center for Education Statistics (NCES) and conducted by Westat, includes a household-level screening interview to identify members of the targeted subpopulation, who are then subsampled to determine potential respondents. This two-stage sample design is used in some of the HHS surveys whose response trends are discussed in the next chapter. Response rates to the second stage are reported two ways. The *conditional* response rate is the response rate among persons selected into the stage two sample. The *unconditional* (or final) response rate is the product of the response rate to the screening interview and the conditional stage two response rate. It represents the response rate among all persons who would have been selected into the stage two sample if all sample households had responded to the screening interview. Response rates to the screening interview declined from

[2] Beginning in 1994 the GSS has been conducted in alternate years. Prior to that and going back to its start in 1972, the GSS was conducted in all but three years (NRC 2013).

81.0 and 82.1 percent in 1991 and 1993, respectively, to 69.9 percent in 1996 and 52.5 percent in 2007, which was nearly 12 percentage points lower than the preceding survey in 2005 (NRC 2013). The unconditional or final adult response rates declined from almost 60 percent in 1996 to about 33 percent in 2007 (Brick and Williams 2013). Falling conditional adult response rates contributed to the decline in the final adult response rates, as the conditional rates dropped from 86 percent in 1996 to 63 percent in 2007.

Response rates to the SCA, a monthly survey with a rotating panel design (part of the sample in a given month was interviewed six months earlier), declined from 76 percent in 1979 to 60 percent in 2003, with the rate of decline being greater after 1996 than before 1997 (NRC 2013 and Curtin, Presser, and Singer 2005).

The SCF, which is sponsored by the Board of Governors of the Federal Reserve System, has been conducted every three years since 1992, with the University of Michigan and then NORC at the University of Chicago performing the data collection, which is done with face-to-face interviews. The sample design of the SCF differs from all of the other surveys in having both an area frame representing the entire population and a list frame selected from tax returns with patterns of income suggesting high asset levels. Response rates for the area frame sample fell off slightly from 68.0 percent in 1992 to around 66 percent in 1995 and 1998 but returned to around 68 to 69 percent from 2001 through 2010 (NRC 2013). Response rates for the list frame have been much lower historically and less stable than response rates for the area frame. In 2007, for example, the response rate for the list frame sample was 34.7 percent, with response rates varying from 12 to 49 percent across the seven strata (Kennickell 2010).

In addition to these results, the CNSTAT report cites an internal Census Bureau memorandum comparing rates of nonresponse in 1990 and 2009 in six household surveys

conducted by the Census Bureau (four for other agencies). In three surveys the nonresponse rate nearly doubled over the 20 years while it grew by somewhat more than that in two other surveys and increased nearly four-fold in a sixth survey. These point-to-point comparisons do not reveal how the growth in nonresponse was distributed over time, which is important for understanding not only its possible origins but where things may be headed in the near future. In the next chapter we take a close look at trends in eight federal surveys from the mid- to late 1990s through the most recent year available.

## 2. Panel surveys

Panel surveys conduct multiple interviews with the same units over a period of time. Open-ended panel surveys conducted by university-affiliated centers in the U.S. and by government agencies abroad continue to experience high rates of retention despite the long-term declines in response rates to cross-sectional surveys and short-term panel surveys. Schoeni et al. (2013) reviewed the experiences of three U.S. panel surveys plus British, German, and Australian panel surveys. The U.S. surveys were the Panel Study of Income Dynamics (PSID), the Health and Retirement Study (HRS), and the National Longitudinal Survey of Youth, 1979 (NLSY79). The other three surveys were the British Household Panel Study, the German Socio-Economic Panel, and the Household, Income, and Labor Dynamics in Australia Survey. For most of these surveys, reinterview rates, representing the percentage of respondents to one wave who were reinterviewed in the next wave, have not declined over time, remaining above 90 percent, generally.[3]

Some monthly surveys include a panel component in their sample designs, as we noted with the SCA. The Current Population Survey (CPS), the source of the monthly estimates of the

[3] The reinterview rate, as defined here, does not account for persons who were non-respondents to the prior wave but respondents in the current wave. Returning respondents reduce cumulative sample loss from attrition.

unemployment rate and labor force participation in the U.S., utilizes a design in which sampled addresses are included in the monthly survey for a total of 8 months. Specifically, they are in for 4 consecutive months, out for 8, and back in for another 4. The monthly sample is divided into 8 rotation groups such that 1/8 of the addresses are appearing for the first time, 1/8 for the second time, and so on. The 4-8-4 pattern ensures that 75 percent of the addresses appearing in two consecutive months are the same, and 50 percent of the addresses appearing in the same calendar month in consecutive years are the same. The overlap increases the precision of month-to-month and year-to-year estimates of change.

Atrostic et al. (2001) compared CPS response rates by rotation group and found a pattern of rising response rates across households as the number of months in the sample increased from 1 to 4 and from 5 to 8 but a reduction in response rates between months 4 and 5, where the interviews are separated by 8 months. This could reflect housing units changing hands over the 8 months, making the month 5 respondents newcomers to the survey, as well as former respondents becoming less enamored of the survey after such a long time away.

### 3. Item nonresponse

While this review focuses on unit nonresponse, nonresponse to individual survey questions is a concern as well. Trends in item nonresponse are not nearly as well documented as trends in unit nonresponse, but it is readily apparent that item nonresponse among those who respond to surveys is increasing as unit response rates decline.

An especially useful way to measure both the level and impact of item nonresponse to income questions when such questions encompass multiple sources of income that is of varying importance across respondents is to compute the percentage of total income (or major components of total income) that is imputed as opposed to reported. Czajka (2009) compared such estimates in 1993, 1997, and 2002 for the CPS March supplement and the Survey of Income

and Program Participation (SIPP). In the SIPP the fraction of total income that was imputed grew from 20.8 percent to 24.0 percent to 28.6 percent over the three years. In the CPS the fraction imputed grew from 23.8 to 27.8 to 34.2 percent. When total income was divided among 8 sources, nearly every source showed increases in imputation rates between each pair of years. The increases were generally more striking in the CPS than the SIPP, but the latter survey showed particularly large increases in the imputation rates for pension income, welfare income, and transfer income other than welfare, social security, and Supplemental Security Income. Between 1993 and 2002 the fraction of pension income imputed in the SIPP rose from 23.7 percent to 47.3 percent; the fraction of welfare income imputed rose from 13.8 percent to 32.8 percent; and the fraction of other transfer income imputed grew from 20.8 percent to 33.6 percent. It is noteworthy that these increases in item nonresponse were much greater than the observed increases in unit nonresponse over this same period.

Meyer et al. (2015) documented trends in item nonresponse rates for six transfer programs in the CPS supplement and the SIPP from 1991 through 2013. The six programs were Aid to Families with Dependent Children (AFDC)/Temporary Assistance to Needy Families (TANF), the Food Stamp Program/Supplemental Nutrition Assistance Program, Social Security, Supplemental Security Income (SSI), Unemployment Insurance, and Worker's Compensation. In 1991, item nonresponse rates for these six programs—expressed as the percentage of aggregate dollars imputed—ranged between 14 and 23 percent. By 2013 these rates had risen to between 24 and 35 percent.

## C. Reasons for declining response rates

A common way of classifying the many reasons for nonresponse to a survey uses three categories: (1) non-contact, (2) refusal, or (3) other. The "other" category encompasses such

reasons as language issues, poor health, and absence from home during the data collection period (Brick and Williams 2013; see also Groves and Couper 1998).

Multiple societal and environmental factors have been suggested as contributing to the decline in survey response rates, either by making it more difficult to contact prospective respondents or by increasing the likelihood that, if contacted, they will refuse to participate.

- Increasing number of two-worker households
- Longer average commuting time
- Increasing prevalence of caller ID on landline telephones
- Explosive growth of cellular telephones and cell phone-only households
- Growth in the number of federal surveys[4]
- Expansion of political polling and political calls to influence voting behavior
- Growth in telephone solicitations, contributing to people's reluctance to answer the phone
- Increasing fear or distrust of unknown callers, influenced in part by concerns about identity theft
- Increasing popularity of communication modes—specifically, through the Internet—that surveys have not used extensively

Arguably, some of these changes have accelerated in recent years.

When non-interviews are apportioned among non-contacts, refusals, and other reasons, there is evidence that the distribution among these factors has changed little with growing nonresponse. Brick and Williams (2013) compiled such data for the NHIS (a face-to-face survey) and the NHES (a phone survey) from the mid-1990s through 2007. For the NHIS, noncontacts were 26.4 percent of the total in 1997 and 27.0 percent in 2007, although they peaked at 35.7 percent in 1999. Refusals remained the dominant source at 65.3 percent in 1997 and 61.1 percent in 2007, never dropping below 59.0 percent or exceeding 67.5 percent. Other factors did show an upward trend, accounting for 6.9 percent in 1997 and between 10 and 12 percent in the final four

[4] Presser and McCullogh (2011) document the sharp growth in the number of household surveys administered by the federal government over the period 1984-2004.

years. In short, as total nonresponse grew, each of its main components kept pace. The NHES showed a somewhat different pattern. Between 1996 and 2007, refusals grew slightly, from 83.0 to 86.1 percent; noncontacts remained steady (10.0 versus 10.2 percent); and other reasons declined from 7.0 to 3.7 percent. During the intervening years, however, noncontacts were between 17.0 and 18.9 percent of the total, and refusals dropped to between 74.0 and 77.9 percent.

In general, the increase in nonresponse has been much greater in telephone surveys than in face-to-face surveys, which is consistent with the growing number of solicitations that households receive by telephone and the increasing use of answering machines and caller ID to screen calls. The observation that noncontacts do not appear to have grown as a share of nonresponse leads Brick and Williams (2013) to suggest that the impact of technological barriers may be less on preventing contact than on keeping the interviewer from fully communicating the survey request.

Survey design changes can also affect response rates. The introduction of computer-assisted interviewing into the CPS in January 1994, along with other changes (most notably, an increased reliance on telephone interviews), had a pronounced effect on response rates (U.S. Census Bureau 2006). In the 30 years preceding 1994 the average yearly response rate to the basic CPS ranged between 95 and 96 percent, with no clear trend. In 1994 the response rate fell below 94 percent and in the ensuing years showed evidence of a continuing small but steady decline. This secular decline may have started earlier (in 1992 or 1993) but was masked by the effects of the design changes.

The introduction of cell phone frames into RDD surveys—necessitated by the growth in households with only cell phone service—has contributed to more recent declines, as calls to cell

phones yield lower response rates than calls to landline phones. We show evidence of this in the next chapter.

It has also been suggested that declining response rates are in part a result of declining survey effort in the face of rising survey costs or a failure to increase the level of effort sufficiently to offset the greater difficulty in contacting prospective respondents and persuading them to participate. However, there is little evidence to support either position. In fact, Peytchev (2009) finds that call attempts for completed interviews doubled between 1976 and 1996 in the SCAs, yet the non-contact rate more than tripled over this period. The lack of sufficient cost data inhibits study of this issue, as several Census Bureau task forces convened to examine survey costs discovered. One particular area of need is how interviewers allocate their time. Paradata—information collected about the survey process (see Chapter V)—hold out some promise as a means of documenting contacts (and attempts) with sample households. Still, having a common framework for assessing the relationship between costs and response rates would be helpful—as would a quantitative model of their relationship.

After reviewing the available evidence, the NRC panel concluded that "response rates continue on a long-term downward path, but . . . solid evidence about the reasons for the decline is still elusive" (NRC 2013). That conclusion would probably be amplified in light of the findings presented in the next chapter, which update earlier trend data for the major HHS surveys and a flagship survey from the U.S. Census Bureau.

# III. RESPONSE TRENDS IN HHS SURVEYS

To document the trends in response rates to HHS surveys, we present trend data for seven surveys sponsored by the department. The seven surveys are national in scope and include those for which response rate trend data are most readily available. The seven surveys are:

- The National Health Interview Survey (NHIS), sponsored by the National Center for Health Statistics (NCHS) within the Centers for Disease Control and Prevention (CDC)
- The National Health and Nutrition Examination Survey (NHANES), also sponsored by NCHS
- The Behavioral Risk Factors Surveillance System (BRFSS), sponsored by CDC
- The National Immunization Survey (NIS), sponsored by the National Center for Immunization and Respiratory Diseases within CDC
- The National Survey of Drug Use and Health (NSDUH), sponsored by the Substance Abuse and Mental Health Services Administration (SAMHSA)
- The Medicare Current Beneficiary Survey (MCBS), sponsored by the Centers for Medicare & Medicaid Services (CMS)
- The Medical Expenditure Panel Survey (MEPS), sponsored by the Agency for Healthcare Research and Quality (AHRQ)

For comparison, we also include the CPS ASEC, which is co-sponsored by the Bureau of Labor Statistics (BLS) and the U.S. Census Bureau, which performs the data collection.

Descriptions of the eight surveys are provided in the Appendix, along with detailed trend data by survey component. Summary trends for seven of the surveys are presented here.

## A. Trends across surveys

To demonstrate recent trends across the surveys, Figure III.1 plots key response rates from 1995 or later through the present time for seven of the eight surveys: the NHIS, NHANES, NIS, NSDUH, MCBS, MEPS, and CPS. We exclude BRFSS from this figure because the data are collected by the states, so there is no national response rate, and trends in average state response rates can be influenced by the actions of a few states. In the next section we report median, minimum, and maximum state rates for BRFSS. For each of the other seven surveys we chose

one response rate to represent recent trends. These response rates reflect the following survey components:

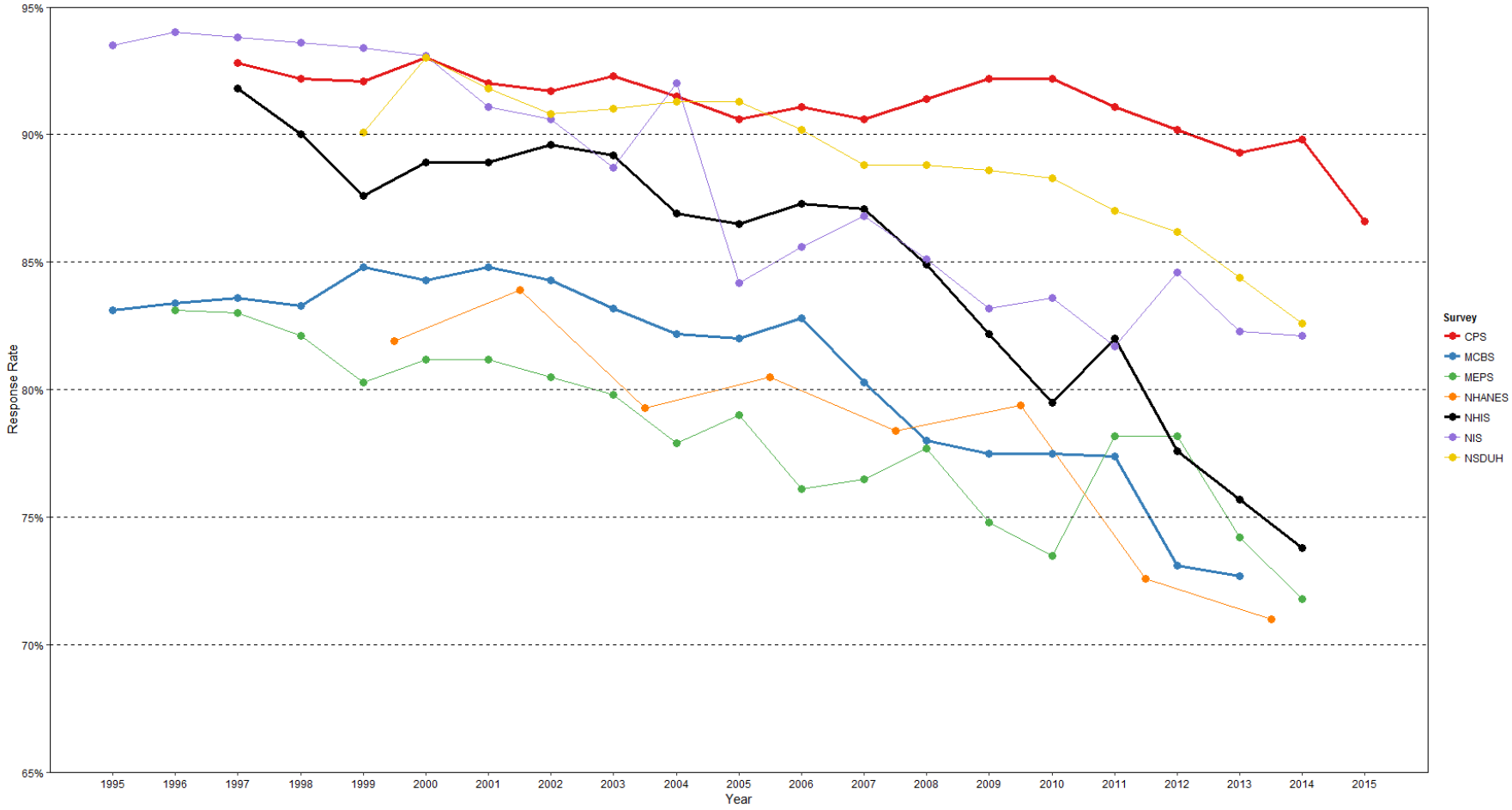
- For NHIS, the household module
- For NHANES, the personal interview, conditional on selection into the sample following completion of the household screening interview
- For NIS, the household interview, conditional on having an age-eligible child, as determined from a screening interview, for the landline telephone sample
- For NSDUH, the household screening interview
- For MCBS, the initial interview for each new panel
- For MEPS, the initial interview, conditional on completion of the prior year NHIS household interview
- For the CPS, the March labor force interview

Response rates for other components of each survey are presented in the Appendix.

While the response rates and trends differ, all seven surveys show some amount of decline in their response rates over the first half of the period, and six of the surveys exhibit accelerated rates of decline in the most recent years. In their initial and final response rates the surveys form two clusters of three—a higher response rate group, consisting of the CPS, NIS, and NSDUH; and a lower response rate group, including the MCBS, MEPS, and NHANES. The NHIS in dramatic fashion crosses from the high group to the low group. Its response rate, which started at about 92 percent in 1997, fell to around 74 percent in 2014.[5] Between 1997 and 2007 the response rate dropped about 5 percentage points, or one-half a percentage point per year. But over the next 7 years the NHIS response rate fell another 13 percentage points or nearly 2 percentage points a year.

[5] The response rate to the household module was 95.5 percent in 1990 (NRC 2013).

## Figure III.1 Selected survey response rates




Source: See the Appendix.

Among the seven surveys the smallest decline was recorded by the March CPS labor force survey, which is very brief and includes many repeat households in its sample. Its response rate dropped 2 percentage points between 1997 and 2007, but after a modest increase over the next few years the response rate declined more steeply than in the earlier years, ending up about 6 percentage points below where it stood in March 1997.

Almost from the start, the conditional response rate to the NIS interview began to decline, falling from 94 percent in 1996 to about 89 percent in 2003. Since then the response rate has moved up and down, but unlike the other six surveys the NIS response rate over the last six years does not show an increased downward trend relative to earlier years. The response rates in the landline sample in 2013 and 2014 (response rates to the cellular phone sample, introduced in 2011, are excluded) are just a percentage point lower than they were in 2009 and 2010 but about 11 percentage points lower than in 1996.

Completing the top cluster, the NSDUH response rate actually improved slightly between 1999 and 2005, rising from 90 percent to over 91 percent. From there it began a gradual decline, however, dropping by 3 percentage points between 2005 and 2010. The rate of decline increased after that point, reducing the response rate by nearly 6 percentage points over the next four years. The decline from its peak was about 10 percentage points.

In the lower cluster both the MCBS and MEPS started with response rates around 83 percent and ended up around 72 to 73 percent, although they followed different paths to arrive in the same place. MEPS showed a nearly 5 percentage point increase between 2010 and 2011 and maintained the higher rate for another year before dropping more than 6 percentage points in the next two years. The NHANES response rate, which represents a two-year average, declined from

82 percent between 1999 and 2000 to about 79 percent between 2009 and 2010 but then dropped another 8 percentage points in the next two years.

The magnitude of the decline in the NHIS response rate is puzzling, as the NHIS has attributes that are associated with high response rates. In particular, it is sponsored and conducted by the federal government, and it is conducted face-to-face.

## B. Other HHS surveys

Other notable national surveys sponsored by HHS agencies include:

- National Survey of Family Growth (NSFG)
- National Youth Tobacco Survey (NYTS)
- National Adult Tobacco Survey (NATS)
- Youth Risk Behavior Surveillance System (YRBSS)
- Monitoring the Future (MTF)

MTF is sponsored by the National Institute of Drug Abuse (NIDA) within the National Institutes of Health (NIH). All of the rest are sponsored by CDC, with NSFG being sponsored by NCHS within CDC. In addition to the national tobacco surveys there are state tobacco surveys of youth and adults as well as periodic, specialized surveys of select subpopulations. Response rate trend data are not readily available for these surveys. Also, some of the surveys do not start with a household frame. The NYTS and MTF draw first stage samples of schools and then sample students from the schools that agree to participate.

Not listed here are several surveys administered by CMS to capture the experience of patients with different components of the U.S. healthcare system. Most of these surveys are part of a family of surveys under the general title Consumer Assessment of Healthcare Providers and Systems (CAHPS).[6] Because of the variety of CAHPS surveys, which utilize different sample

[6] CAHPS is a registered trademark of the Agency for Healthcare Research and Quality, which developed the survey design in collaboration with CMS and funds and oversees the CAHPS program.

designs and employ differing modes to collect the data, response rates, if available, would not be very useful for monitoring trends.

# IV. RESPONSE RATES AND NONRESPONSE BIAS

In their review of declining response rates and their potential causes, Brick and Williams (2013) begin with the observation that "nonresponse in household surveys has been a topic of great interest for many years because the foundation of inference from probability samples is based on observing all the sampled units." In essence, anything less than a 100 percent response rate threatens the validity of statistical inference unless nonresponse is itself completely random. As reported in previous chapters, response rates approaching 100 percent have been achieved by some of the major federal surveys, but response rates for even these surveys have fallen in the past two decades—in some cases dramatically. Other federal surveys have never achieved response rates approaching 100 percent; the declines in their response rates would seem to warrant even greater concern.

Federal agencies that seek to administer a questionnaire to more than nine respondents must obtain formal approval from the Office of Management and Budget (OMB). This is done through an Information Collection Request (ICR), known more informally as an "OMB package," which provides a detailed description of the proposed data collection, including expected response rates. When the expected unit response rate falls below 80 percent, an agency must plan to conduct a nonresponse bias analysis to evaluate the actual or potential nonresponse bias on key survey estimates. Even if the unit response rate exceeds 80 percent, the agency should conduct a nonresponse bias analysis for any individual item for which the response rate falls below 70 percent. OMB suggests a higher threshold for key variables.

This chapter considers the relationship between response rates and nonresponse bias, discusses ways of evaluating nonresponse bias, and reviews recent research that examines the relationship between survey response rates and nonresponse bias.

## A. How nonresponse contributes to biased estimates

Understanding of the importance of the survey response rate as an indicator of potential bias has evolved as response rates have declined. This evolution has been driven in part by studies of selected surveys showing no increase in response bias with declining response rates. A much larger body of research has demonstrated that the relationship between nonresponse and nonresponse bias is complex. This research has also altered thinking about the prospective respondent and this individual's inclination to respond to a given survey.

If nonresponse were entirely random, nonresponse bias would not be an issue, regardless of the level of nonresponse. This was an implication of an early formulation of nonresponse bias, which expressed such bias as a function of the nonresponse rate (1 less the response rate, $R$) and how much the respondents ($r$) differed from the non-respondents ($nr$) on a given measure, in this case the mean of $Y$:

$$\text{bias}(\bar{Y}) = (1 - R) \times (\bar{Y}_r - \bar{Y}_{nr})$$

This formulation implied or at least suggested a view of sample cases as either inclined to respond or inclined to not respond. The current view of prospective respondents is that each one has a propensity to respond, which is influenced by personal characteristics and the "circumstances" of the survey, and can be changed—increased or decreased—by actions taken by the field staff and the survey sponsors. Across the sample, if the propensity to respond is correlated with any variable measured in the survey, then nonresponse will contribute to biased estimates. The higher the correlation, the greater the bias introduced by a given level of nonresponse.

## B. Methods of estimating nonresponse bias

In its guidance to federal agencies conducting or sponsoring surveys, OMB (2006) outlines several methods that can be used to estimate nonresponse bias. These include:

- Comparing respondents and non-respondents on variables available from the sampling frame
- Comparing respondents and non-respondents on external data that can be linked to the sampling frame—ideally at the individual level (possible for a survey of program participants) but, if not, then at the area level
- Conducting a follow-up survey of a sample of non-respondents with an abbreviated questionnaire and using more extensive and intensive methods to elicit responses and then comparing the sampled non-respondents with earlier respondents on the characteristics collected
- Comparing respondents and initial refusals who were later converted to respondents
- Comparing estimates of key characteristics by level of effort required to elicit responses (for example, number of contact attempts)
- Comparing estimates of key characteristics using different nonresponse adjustments to assess the sensitivity of the estimates to these alternative adjustments

For all of these methods, OMB stresses that, to the extent possible, comparisons should utilize variables that reflect the central purpose of the study. It is not sufficient, for example, to assess bias for estimates of demographic characteristics—which may be relatively easy to do for some surveys—if these are not particularly relevant to what the survey was developed to measure. OMB recommends that agencies consult with "professional statisticians and survey methodologists" so that any potential nonresponse bias can be addressed in how the study is designed, as the options become more limited once data collection is under way.

The CNSTAT report (NRC 2013) identified some additional approaches:

- Comparing respondents and non-respondents on characteristics observed by interviewers, including descriptions of housing units and neighborhoods
- Post-stratifying the sample using population totals and comparing weighted and unweighted distributions to determine where weighting has had the greatest impact, which may indicate variables with nonresponse bias; the results will be affected by undercoverage in the sampling frame as well, however
- Comparing estimates of respondent characteristics to estimates derived from much larger surveys and administrative sources, although such comparisons are affected by measurement differences between the sources as well as measurement error

The interviewer observations extend the frame data in allowing direct comparisons of responding and nonresponding households whereas the approaches using aggregate data compare totals based solely on respondents to totals from the complete population or what is believed to be a more complete representation of the population.

When surveys are longitudinal or multi-stage (that is, involving an initial screener interview with subsequent selection of respondents for more detailed interviews), there are additional methods available to estimate nonresponse bias. For longitudinal surveys, variables collected in the initial wave can be used to compare respondents and non-respondents to subsequent waves. Similarly, for multi-stage surveys, information collected in the screener interview can be used to compare respondents and non-respondents to the later stage, more detailed interviews. Ideally, the screener interview should be designed with such comparisons in mind, which implies including a selection of key variables or their correlates in the screener interview to allow later comparisons on important characteristics.

Studies using frame data or supplemental data are stronger, methodologically, than studies that base their estimates on attempts to interview non-respondents because both the availability and accuracy of frame or supplemental data are independent of response propensity and the characteristics of respondents versus non-respondents. Re-interview attempts rarely succeed in contacting and interviewing all non-respondents, so the estimates of bias that they yield may be misleading. Findings may suggest that there are no differences between the respondents and non-respondents or, worse, that the bias is in one direction when, in fact, it is in the opposite direction. There are additional concerns that reluctant respondents may not provide data of the same quality as more willing respondents.

## C. The relationship between response rates and nonresponse bias

To investigate the relationship between response rates and nonresponse bias requires estimates of nonresponse bias, which can be obtained using the methods described above. Here we present the findings from a landmark analysis of a large set of studies that provided estimates of both response rates and nonresponse bias. We then review a recent study that linked administrative records to the household sample frame of a major survey in order to estimate nonresponse bias across the characteristics present in the administrative data.

### 1. Meta-analysis of nonresponse bias studies

Groves and Peytcheva (2008) conducted a meta-analysis of 59 international studies that produced 959 estimates of nonresponse bias for population averages (means or medians) or percentages. The studies were identified through a search of the literature dating back to 1978, and they range in publication date from 1978 through 2005. To be included, a study had to have used one of five bases for estimating nonresponse bias:

- Sample frame data for respondents and non-respondents
- Supplemental data linked to both respondents and non-respondents
- Screener interview data, used to compare respondents and non-respondents to a later, larger interview
- Follow-up surveys of sample persons who were non-respondents to an earlier survey, comparing the data collected from respondents in the earlier survey with the data collected in the follow-up survey
- Reports of intentions to respond or not respond to a later survey among respondents to an earlier survey

Studies comparing early and late respondents to the same survey were not included.

Regarding the method of estimating nonresponse, the authors observed that estimates of nonresponse bias obtained from studies that compared screener versus full interviews or early versus follow-up surveys were higher, on average, than estimates obtained from studies comparing respondents and non-respondents with respect to either frame or supplemental data.

In addition, while there is evidence that different types of nonresponse (for example, noncontact versus refusal) affect bias differentially, the studies that were analyzed combine all types of nonresponse.

Summarizing an earlier, more limited meta-analysis, Groves (2006) drew the following conclusions about the relationship between nonresponse and nonresponse bias. First, there is no necessary connection between the nonresponse rate and nonresponse bias. Second, there is no minimum response rate below which a survey estimate is necessarily biased and no response rate above which it is never biased. Third, bias can vary across statistics in the same survey.

Building on the earlier meta-analysis, Groves and Peytcheva (2008) addressed three questions:

- Are there characteristics of survey design that are systematically related to nonresponse bias?
- Are the properties of target populations related to nonresponse bias?
- Are there characteristics of survey estimates that are systematically related to nonresponse bias?

The findings are summarized below.

**Attributes of the Survey Design.** Neither pre-notification nor incentives—techniques used to increase response rates—was significantly related to the magnitude of the response bias, but the authors note that we would not necessarily expect such techniques to have the same effect across different types of survey estimates. Prior involvement with the survey sponsor was associated with significantly lower nonresponse bias whereas government-sponsored surveys tended to have higher nonresponse bias than non-government surveys, although this latter result disappeared when the studies were restricted to those using frame or supplemental data. Interviewer-administered surveys tended to produce higher nonresponse bias than self-administered surveys although the reasons were not clear. One possible explanation is that

interviewers may express the goals of the survey more explicitly than is done in the material accompanying a self-administered survey. If response is affected by attitudes toward these goals, then nonresponse bias would be greater in the interviewer-administered surveys.

There is reason to believe that the potential for nonresponse bias varies with the survey topic. Groves and Peytcheva explored this to a limited degree, comparing surveys classified as health versus other, but they found no difference. When they restricted the comparison to studies using frame or supplemental data, however, health surveys had significantly lower nonresponse bias.

**Attributes of the Sample Population.** There is an impression that surveys of subpopulations that focus on issues specific to those subpopulations tend to generate higher nonresponse bias than surveys of the general population because people with particular views may be either more inclined or less inclined to respond. Contrary to expectation, surveys of the general population tended to have significantly larger estimates of nonresponse bias than surveys of subpopulations. Likewise, urban residents and members of minority subcultures tended to have lower response rates than other sample members, which could imply differences between those who respond or do not respond within these groups, but there were no significant differences in nonresponse bias by urbanicity or majority/minority status.

**Attributes of the Survey Estimates.** If most of the variation in nonresponse bias lies within rather than across surveys, particular types of items may be subject to greater nonresponse bias than other items. The findings from the meta-analysis support this to at least some degree. Attitudinal variables tended to have much larger nonresponse bias than behavioral or demographic characteristics. There were no significant differences between the latter two types of variables, however. When the topic of the survey is of particular relevance to the respondent,

response propensity may be associated with the respondent's views on the topic. The saliency of the topic was examined in two ways—at the survey level and at the item level (that is, for the item for which the estimate of nonresponse bias was produced). For the former, estimates of nonresponse bias for surveys that were likely to be of interest to the population being surveyed were compared to estimates for all other surveys. For the latter, the comparison was limited to items that reflected the general theme of the survey. In neither case were differences in response bias statistically significant, however.

Groves and Peytcheva also explored whether *differences* in subclass means tended to show less nonresponse bias than the subclass means themselves. The supposition here is that if two subclass means were biased in the same direction, the biases would partially cancel when the difference in means was calculated. The meta-analysis provided no support for this theory. Instead, the bias in the difference between subclass means tended to be at least as large as the bias in the subclass means.

**Conclusions.** In summarizing the results of their meta-analysis, Groves and Peytcheva listed their principal conclusions in order from strongest to weakest, noting potential confounding factors that could be present in the underlying studies. In their words:

- Large nonresponse biases can happen in surveys
- The search for mechanisms that link nonresponse rates and nonresponse bias should focus on the level of individual measures and not on the level of the survey
- Differences of subclass means do not, in general, enjoy lower nonresponse biases than their constituent subclass means
- How we estimate nonresponse bias may make a difference

Another conclusion that is implicit in their presentation of findings but not expressed here is that the overall level of survey nonresponse, in and of itself, is not a good indicator of the overall or average level of nonresponse bias.

### 2. Nonresponse bias in the CPS ASEC

A property of surveys that draw their samples from addresses is that any external dataset containing addresses can be linked to the sample. The linked data can then be used to study nonresponse bias, as both responding and nonresponding households will have such data. Bee et al. (2015) linked IRS tax records, by address, to the 2011 CPS ASEC sample and used the linked data to compare survey respondents and non-respondents on the characteristics contained in the tax records. This included demographic information from Social Security Administration records that had been linked to individual taxpayers. Bee et al. found little evidence of differences between the tax return income distributions of respondents and non-respondents—that is, little evidence of nonresponse bias on the adjusted gross income reported on the tax returns. Significant differences did exist, however, in the number of children (respondents tended to have more) and marital status (respondents were more likely to be married). In addition, respondents were more likely than non-respondents to have dividends and income from social security and less likely to have wage and salary income. Respondents were also less likely to itemize, more likely to have capital gains/losses, and more likely to have a profit from farming. These findings underscore a key observation from Groves and Peytcheva—that nonresponse bias is item-specific and that a survey with no nonresponse bias on one characteristic could have considerable bias on another characteristic.

## D. Response rates and data quality

If nonresponse is a potential source of bias, then adding respondents in order to increase response rates would seem to be a sure way to reduce potential bias. However, as we noted earlier, different methods of increasing response rates may attract different pools of respondents. The implication is that expanding the number of completed interviews through additional effort may not yield less biased estimates and could even result in more biased estimates. On this point,

a number of research studies have found adverse effects from efforts to recruit additional respondents. These effects include not only increased bias but reduced quality of the data collected.

Studies that increase effort (call attempts and in-person visits) without changing their approach have found that they attract more of the same kinds of respondents that they have without the additional effort (Curtin, Presser, and Singer 2000; Keeter et al. 2000). On the other hand, Peytchev et al. (2009) reported on a study in which an additional round of calls was made using the same protocol versus a revised protocol that increased incentives and reduced the survey demand. Comparison of the respondents added in each case showed that the respondents added with the revised protocol were different on a number of survey estimates from those added with the original protocol. Furthermore, the respondents added without changing the protocol resembled the respondents obtained in the initial round.

Another dimension that enters the conversation about the impact of adding new respondents via different protocols is quality. A study based on two surveys (the CPS and the American Time Use Study) found evidence that data quality decreased as a sample member's predicted probability of nonresponse increased (Fricker and Tourangeau 2010). Kreuter et al. (2010) found in a German labor panel that increased contact attempts were associated with a reduction in nonresponse bias but with some increase in measurement error. Kreuter et al. (2014), cited in Meyer et al. (2015), found in a German survey that "hard-to-recruit respondents provided less accurate reports of welfare benefit receipt than those" who were "easy to recruit." Bollinger and David (2001) found that those who responded to all waves of a SIPP panel reported Food Stamp Program participation more accurately than those who missed one or more waves.

While some of these studies suggest that adding respondents can have a negative effect on survey data quality, there is evidence as well that declining response rates—both unit and item—in selected surveys have been accompanied by reductions in the quality of data collected. Meyer et al. (2015) document the average bias in reporting of benefits received from seven transfer programs in five household surveys over the period 2000 to 2012 and provide coefficients for the regression of bias on the survey year over an extended period covering 30 or more years for most of the surveys. Most of the coefficients show a decline in the reporting of benefit dollars over time—a measure of quality for data on program participation.

These findings underscore the dilemma faced by survey organizations that are confronted with declining response rates on the one hand and whatever that may mean with respect to nonresponse bias, and the possibility, on the other hand, that whatever they do to increase response rates may only make matters worse. New strategies for survey design that may hold some promise for addressing this problem are discussed in the next chapter.

## E. Summary measures of nonresponse bias

The response rate cannot be dismissed, as it remains an important indicator of at least the potential for bias. Downplaying the value of the response rate can also become a reason to expend less effort and creativity in obtaining interviews. Nevertheless, the principal lesson from the material presented above is that one must be cautious in attributing too much importance to the response rate as an indicator of quality, as it is not directly linked to bias, and it is not variable-specific (NRC 2013). High response rate standards can be counter-productive, as they focus on a single potential source of error to the exclusion of others (Kreuter 2013).

In light of the diminished importance of the survey response rate as an indicator of nonresponse bias, new indicators are being developed that can measure nonresponse bias more directly or at least predict its potential more accurately. Wagner (2012) proposes a typology of

such indicators, discusses their strengths and weaknesses, and suggests directions for future research. His typology distinguishes: (1) indicators involving the response rate, (2) indicators involving the response rate and either frame data or paradata, and (3) indicators involving the response rate, either frame data or paradata, and survey data. Indicators of the second type include subgroup response rates for subgroups defined by frame data or paradata; coefficients of variation of these subgroup response rates; and representativity indicators (or R-indicators) as proposed by Schouten et al. 2009), which are intended to capture the variability of response propensities. Indicators of the second type are measured at the survey-level. Indicators of the third type include correlations between auxiliary variables and survey variables; comparisons of early and late responders; follow-up surveys of non-respondents; variation in variable means across deciles of the survey weights. Indicators of the third type are measured at the variable or item level.

Not included in Wagner's typology are balance indicators (B-indicators), which measure the degree of fit between respondent and population characteristics on frame variables—ideally, providing a rich set. B-indicators were proposed by Sarndal (2011). A dataset is considered to be perfectly balanced if the means of the measured auxiliary variables are identical for respondents and all of the population sampled.

# V. ADDRESSING SURVEY NONRESPONSE AND ITS EFFECTS

Given the potential for nonresponse bias, what can be done to lessen its chances or limit its consequences? We differentiate between strategies to improve response rates and strategies for mitigating the potential effects of survey nonresponse on bias. We conclude this chapter with a discussion of new approaches to data collection.

## A. Strategies to maximize response rates

While the preceding chapter underscored the point that efforts to increase survey response rates do not guarantee a reduction in nonresponse bias, low response rates challenge the credibility of surveys, inviting efforts to increase them or least stem the tide of decline. A number of strategies are being used—or have been recommended for expanded use—in an effort to maximize survey response rates. These include the payment of incentives, reducing survey burden, the use of address-based sampling in combination with a mail survey mode, the use of multiple modes within the same survey, double or two phase sampling, and responsive design. We discuss these alternate approaches below followed by a review of strategies for the special case of retaining respondents in panel surveys.

### 1. Incentives

Among strategies to maximize response rates, the payment of incentives has been shown to increase response rates most consistently. Much research has been done on types and levels of incentives but often in small experiments added to ongoing surveys. That major surveys continue to test alternative incentives is an indication that much remains to be learned about the potential impact of incentives and that the survey context is such an important mediator that it is difficult to generalize from one survey to another.

In their review of the use and effects of incentives in surveys, Singer and Ye (2013) present six general conclusions about the impact of incentives on survey response rates:

- *Incentives increase response rates to surveys in all modes, including the Web, and in cross-sectional and panel studies*
- *Monetary incentives increase response rates more than gifts, and prepaid incentives increase them more than promised incentives or lotteries, though they are difficult to implement in Web surveys*
- *There is no good evidence for how large an incentive should be. In general, though response rates increase as the size of the incentive increase, they do so at a declining rate.*
- *Relatively few studies have evaluated the effect of incentives on the quality of response. Most studies that have done so have found no effects, though the variables used to assess quality have generally been limited to item nonresponse and length of responses to open-ended questions. Research is needed on what effect, if any, incentives have on reliability and validity.*
- *Relatively few studies have examined the effect of incentives on sample composition and response distributions, and most studies that have done so have found no significant effects*
- *Incentives, thus, have clear potential for both increasing and reducing nonresponse bias. If they can be targeted to sample members who would otherwise fail to respond, they may reduce nonresponse bias. If they affect all sample members equally, however, they are unlikely to affect nonresponse bias; and if they bring into the sample more of those who are already overrepresented, they may increase whatever nonresponse bias exists.*

On this last point, Singer and Ye cite examples of incentives affecting nonresponse bias in all three ways.

While the effectiveness of incentives in increasing response rates has been demonstrated time and again, Singer and Ye (2013) caution that incentives cannot be expect to "do it all." Other ways to reduce the costs of responding to surveys must be identified and put into practice.

**2. Reducing survey burden**

Survey burden includes not only the time required to take a survey but the effort entailed in doing so—both before, as preparation, and during—and the stress associated with providing the requested information, which today may include not only verbal or written responses but biomarkers (NRC 2013). Evidence on the impact of survey length—the one quantifiable dimension of burden—as a deterrent to response is mixed. Bogen (1996) reviewed observational and experimental studies of the relationship between questionnaire length and the response rate

and found evidence both supporting and refuting a relationship. Nevertheless, the time required to take a survey is unquestionably a cost that potential respondents, if they are told or can otherwise gauge, are likely to weigh against the benefits that they incur by participating. Only if they perceive or receive some value from expending the time will this cost be offset. Efforts to reduce the burden have tended to focus on questionnaire length.

One strategy for reducing the length of a survey is to divide major parts of the survey into segments and assign only one segment to each respondent. This approach, matrix sampling, has been employed in various surveys over the years and is used currently in the National Assessment of Educational Progress (NAEP) to administer an expanded set of mathematics and reading questions nationally while limiting the burden imposed on students taking the test. Every child responds to questions that test the full range of mathematics and reading abilities, but the questions administered to each student are a subset of the full battery of mathematics and reading questions. Matrix sampling is one of the options for burden reduction being considered for the American Community Survey (ACS), a continuous survey that has replaced the decennial census long form.

Ironically, the NHIS once employed matrix sampling to capture detailed health conditions for national estimates. Random segments of the sample were administered modules designed to capture broadly different types of conditions. This aspect of the design was eliminated in favor of capturing somewhat less detailed information from all respondents. With the matrix design, analysts could not estimate the frequency of combinations of conditions that crossed the different questionnaire segments, and statistical reliability for individual conditions was weakened as well. This situation differs from the design of NAEP, where each child taking the test receives

questions covering all of the subtopics, and the precision of national estimates at the question level is not important to analysts.

### 3. Address-based sampling

Sampling from lists of mailing addresses has developed as a viable and cost effective tool for conducting surveys by mail as the quality of available lists has improved. Aided by the availability of the U.S. Postal Service's Delivery Sequence File (DSF), address-based sampling (ABS) is contributing to a resurgence of mail surveys (Iannacchione 2011). With the growing difficulties encountered in contacting households by telephone, mail can provide a more effective and much less costly alternative. Following survey protocols established long ago, most major federal surveys that work from household frames utilize mail for the initial contact, even when the data collection is ultimately conducted using other modes. But RDD surveys do not have access to addresses, routinely, for an initial contact and, therefore, cannot make use of this approach. Another advantage of ABS relative to RDD sampling is that ABS can be conducted with a single sample frame whereas the growth of cell phone-only households, which accounted for 40 percent of all households in 2014 (AAPOR 2015), has made it necessary for RDD surveys to employ dual frames to adequately represent the household population, which greatly increases their complexity.

As the best source of addresses for ABS, however, the DSF does have a number of drawbacks, which Link et al. (2008) delineate. One, while the U.S. Postal Service compiles and maintains the DSF, it does not market the file directly, choosing instead to make it available to private vendors, who sell extracts or a complete copy of the file. Vendors differ in the extent to which they update their files, and they differ also in their ability to draw probability samples from the file. Many users are not interested in purchasing the entire file. There are coverage issues as well. The consensus from several evaluations is that coverage is quite good in urban

areas but much weaker in rural areas, rapidly growing areas, and areas with a lot of group quarters residences (Dohrmann et al. 2012). Furthermore, households that maintain both a street address and a post office box have dual exposure to selection. Post office boxes cannot be excluded to circumvent this problem because some households have only post office boxes.[7] Despite these limitations, the DSF holds substantial promise as a data source for creating frames of residential addresses (Link et al. 2008).[8]

Link et al. (2008) tested a mail survey with ABS against the RDD phone survey for BRFSS in six states, five of which had RDD response rates below 40 percent. The mail option employed a number of different experimental treatments. Across all treatments, the mail survey produced a slightly higher response rate in one state and was indistinguishable, statistically, from RDD in the four other low response rate states but was 15 percentage points lower than the RDD response rate in the high response rate state. However, when the mail sample was restricted to the cases that were assigned to the treatment that included a second mailing (if necessary), the response rate for the mail survey was 4 to 6 percentage points higher than RDD in all five of the low response rate states, and the gap in the high response rate state was reduced to 6 percentage points. The authors also compared the characteristics of the respondents to the two modes, with a third source, the CPS, presented as a gold standard. Both BRFSS modes had higher levels of attainment than the CPS, but the mail mode was substantially higher. Similarly, both of the BRFSS modes had a higher proportion of respondents who were white, non-Hispanic than the CPS, with the difference being much greater for the mail mode.

[7] An approach to dealing with multiple opportunities for selection in an address-based sample is to ask responding households if they have other addresses and, if so, adjust their base weights.

[8] Dohrmann et al. (2012) evaluate an approach to improving the coverage of the DSF.

While BRFSS has not adopted ABS (NRC 2013), the NHES switched from telephone to mail between 2007 and 2012. As part of the transition, pilot studies were conducted in 2009 and 2011 (Roth, Han, and Montaquila 2012).

Mail surveys do have notable limitations. The absence of interviewer assistance restricts the topics and the complexity of the instrument—especially relative to computer assisted interviewing. While assistance can be provided via a call-in help line, this does not substitute for the presence of an interviewer. Furthermore, while ABS may increase response rates overall, it is likely to reduce them among persons with lower levels of literacy—potentially increasing nonresponse bias.

**4. Multiple modes**

There is a growing recognition that to obtain responses from households that were non-interviews after the first attempts requires a change in methodology. Continuing to try the same methods may bring in more interviews, but they are unlikely to reduce nonresponse bias, as the additional respondents will tend to resemble the earlier respondents.

Mixed-mode designs seek to take advantage of differences across subpopulations in the likelihood of responding to alternative modes. Kreuter (2013) observes that "the core idea of mixed-mode surveys is to tailor response options to the convenience of individual survey respondents in specific subgroups." Multi-mode designs provide a way of addressing the deficiencies of individual modes.

There is little research evidence, however, that bears directly on the merits of mixed-mode versus single mode designs in reducing nonresponse. The CNSTAT report cites studies that compare modes within the same survey but none that evaluates multiple versus single modes. For instance, for the 2007 Health Information National Trends Survey (HINTS), which the National Cancer Institute sponsors, Westat designed and implemented a dual-frame design with

separate modes for each frame (Cantor et al. 2009). The RDD telephone frame used in the two previous surveys was complemented by an ABS frame, which was sent a mail questionnaire. The study also included experiments with incentives and delivery methods in an attempt to reduce the differential nonresponse generally associated with mail surveys (specifically, lower response by minorities and persons with lower levels of education). The overall response rate (combining the screener and extended interview) for the RDD sample was 24.2 percent whereas the comparable response rate for the mail sample was 31.0 percent. Beginning with the next survey in 2011, HINTS became a single-mode mail survey.[9]

Mixed-mode surveys currently use one of three approaches: (1) the modes are administered sequentially, (2) the modes are administered simultaneously, or (3) a primary mode is supplemented by a secondary mode (De Leeuw 2005 cited in Kreuter 2013). There is some evidence that administering different modes sequentially can improve response rates whereas giving respondents a choice of modes in a concurrent administration does not, but research continues on determining the optimal combination and sequence for different types of studies (Couper 2011).

The best example of a sequential mixed-mode design in the U.S. is provided by the ACS, which is discussed in the next section as it also utilizes a two phase design. Two other Census Bureau surveys, the CPS and SIPP, employ both CAPI and CATI for data collection in a way that represents a combination of the second and third approaches. Both surveys are designed as in-person surveys of households drawn from area frames, with clustering to minimize interviewer travel, but both allow interviews to be conducted by telephone as well as face-to-face. As panel surveys, both SIPP and the CPS conduct their initial interviews in person where

[9] See http://hints.cancer.gov/instrument.aspx [February 2016].

possible. The CPS then conducts the second, third, and fourth interviews (in consecutive months) primarily by phone. The fifth interview, four months later, is conducted in person in most cases while the remaining three interviews are conducted by phone. For SIPP there is no such formal allocation of interviews to phone versus in-person, and the Census Bureau's preference is to conduct the interviews face-to-face because of the complexity of the instrument. For both surveys, however, the principal purpose of the second mode is to reduce costs rather than boost response rates although it is possible that response rates do rise as a result.

With a mixed-mode survey, mode effects on subsets of respondents become an issue for analysis and a potential source of differential bias. Vannieuwenhuyze et al. (2010) discuss the issues and propose a method of calculating mode effects in mixed-mode surveys.

**5. Two phase sampling**

In his 1992 presidential address to AAPOR, Bradburn (1992) recommended that survey organizations look to a strategy introduced 50 years earlier to combat the growing problem of survey nonresponse. The basic idea of two phase sampling, which was proposed by Hansen and Hurwitz (1946), is to make an initial round of contacts, securing whatever level of response this generates with a low level of effort, and then subsample the non-respondents, who will be subjected to much more intense methods in order to elicit a very high response rate—ideally, close to 100 percent, effectively eliminating nonresponse.[10] This is possible, in principle, because the low cost of the first round of data collection combined with the subsampling of non-respondents allows a substantially greater expenditure of resources per case in the second round.

[10] In calculating the overall response rate, the subsampled cases are weighted by the inverse of the subsampling rate. Respondents are the sum of those cases responding in the first round plus the weighted number of cases responding in the second round. Non-respondents are the weighted number of eligible cases failing to respond in the second round. Cases not responding to the first round and not selected into the second round subsample make no contribution to the overall response rate.

Two phase sampling has been implemented on a very large scale with the ACS, where it is combined with a sequential multi-mode design. Households selected into the sample are first contacted by mail and given the opportunity to respond to the survey via the Internet, an option that was introduced in December 2012 (NRC 2013). Households that do not respond within two weeks are sent a paper questionnaire, which they are asked to complete and return by mail. Between 55 and 60 percent of the sample households respond by Internet or mail. Limited follow-up by telephone is conducted with all households that fail to respond by either of these modes. At the end of the telephone follow-up the Census Bureau draws a subsample of households, which are visited in person. Ultimately, data are collected from nearly all of these households, yielding a final, weighted response rate that has remained around 98 percent. The fact that responding to the ACS is mandatory by law contributes to the high response rate, although Census Bureau research has shown that in the absence of this legal requirement the Bureau could still achieve a very high response rate, albeit with a significantly higher cost.[11]

**6. Responsive design**

Responsive design (Groves and Heeringa 2006) is a relatively new approach to survey implementation that entails adjustments to the data collection process while data collection is in progress. Key indicators that are obtained in conjunction with the data collection (for example, completed interviews by strata defined from frame variables) are monitored to determine when interventions may be necessary (Kreuter 2013, Peytcheva 2013). Such interventions might include changing the instructions to field representatives on how and where to allocate their contact attempts. The goal of such interventions might be to equalize response rates across

[11] As reported by Navarro, King, and Starsinic (2011), Census Bureau research conducted in 2002 and 2003 found that if response to the ACS were voluntary, the mail response rate would drop by over 20 percentage points and the final, weighted response rate would fall by 4 percentage points. To achieve these rates while maintaining the same level of statistical reliability would increase the annual cost of the ACS by at least 38 percent.

subpopulations and, in so doing, reduce nonresponse bias. Cost control is another common objective of this active management of the survey process.

Wagner et al. (2012) illustrate the application of responsive design to the NSFG, which involved three types of interventions: (1) prioritizing cases for either the screener or main interview, (2) shifting effort to incomplete screener addresses, and (3) targeting selected subgroups to reduce variation in subgroup response rates. Priority cases for the first intervention were identified in 16 different ways, using information from different sources. The second intervention had the interviewers focus on incomplete screener addresses during a designated week early in the data collection process, which would allow more time to complete the main interviews. The third intervention was based on a review of preliminary response rates by subgroup; interviewers were asked to prioritize cases from subgroups with low response rates. All of the interventions were successful in altering interviewer behavior, but not all were successful in changing survey outcomes.

Responsive design is attracting considerable research interest although its use in the NSFG is one of few applications to a major federal survey at present. This may change. The Census Bureau is exploring the use of responsive design for the nonresponse follow-up phase of the 2020 Census (Walejko and Miller 2015). The focus of these efforts is to reduce census costs by improving the efficiency of field operations. In the 2013 Census Test in Philadelphia, adaptive case management was used to designate the number of contact attempts (one versus up to three) and to identify high priority cases for the field representatives each day. Fuller evaluation of these methods will require that the Census Bureau first address a number of implementation issues. Some of the instructions to field representatives run counter to practices they have learned over the years, so the instructions are not always followed fully.

### 7. Retaining respondents in panel surveys

The six panel surveys discussed in Chapter II employed a variety of measures to retain their respondents over time—that is, to minimize unit nonresponse after the first wave. These measures included incentive payments; communication with respondents between waves; a variety of strategies during fieldwork, such as sending advance notification letters, using informants and other tracking strategies to maintain contact with respondents, sending letters to respondents to address their concerns, maintaining interviewer continuity, securing public support from opinion leaders, providing interviewer incentives, and providing information to respondents; and employing survey design features such as limiting the length and frequency of interviews, using supplemental administrative data as an alternative to additional questions, using mixed modes, enhancing interview enjoyment, making use of dependent interviewing to reduce burden, re-contacting those who had become non-respondents, and allowing proxy reports. For the PSID, a variety of analyses conducted over the years have demonstrated that the panel has managed to retain its representativeness (aided by periodic sample additions).

## B. Strategies to compensate for nonresponse bias

Survey weights typically include a component to adjust for differential nonresponse—sometimes combined with a coverage adjustment. Growing concerns about declining response rates and their potential effects on bias have focused renewed attention on weighting adjustments. Under the stochastic model of nonresponse, where everyone has a nonzero propensity to respond, bias due to nonresponse can be eliminated (if nonresponse is random, conditional on one or more variables) if, within each weighting class, all cases have the same value of the outcome variable or all cases have the same response propensity. Either condition will eliminate the correlation between the survey outcome and the propensity to respond.

Because surveys tend to produce multiple outcomes, nonresponse adjustments tend to focus on equalizing the response propensities within weighting classes (NRC 2013).

A number of statistical methods can be used to adjust weights for nonresponse, but they all require variables that are highly correlated with either the outcome variables or the response propensities. Increasing attention is being focused on the identification and use of external or "auxiliary" data to perform these adjustments. We can differentiate between data for comparing non-respondents with respondents and data on the entire population.

**1. Data for comparing non-respondents and respondents**

Data that are available for both respondents and non-respondents are invaluable in estimating nonresponse bias. Such data include variables contained in the sample frame, variables from administrative records that can be matched to the frame, and paradata—that is, the data obtained as a byproduct or in conjunction with the collection of data in the survey.

Generally, the data contained in the sample frame are quite limited. They may include locational information and general housing characteristics. Administrative records that can be matched to the sample frame are a potentially rich source of additional variables for comparing respondents and non-respondents. Meyer et al. (2015) discuss several papers that show how linked survey and administrative data can be used to correct or supplement survey data. Access to such data, however, is typically highly restricted. For example, in the previous chapter we reviewed a study that linked tax records to the sample frame for the CPS ASEC and was able to compare responding and non-responding households on a number of income items and demographic characteristics. Access to the tax records used in the analysis is strictly controlled. Staff at the Census Bureau were able to use these data because the tax code grants the Census Bureau access to tax records for a number of specific uses. In addition, they could link the survey frame and tax data because they could access the addresses in both data sources. Researchers

external to the Census Bureau could obtain such access only through the Census Bureau and only after first submitting to a formal investigation through the Office of Personnel Management and taking required training on protecting the confidentiality of the survey and tax data and the privacy of the individuals who ultimately supplied it.

Paradata were once defined as "process data," generated as a byproduct of the survey operation. More recently the concept of paradata has been expanded to include all data captured by interviewers outside of administering the survey (Olson 2013). Some paradata are available for both respondents and non-respondents, making it useful for nonresponse adjustments. With computer-assisted interviewing and automated survey control systems, both the types and quality of paradata have improved greatly.

Olson (2013) identifies five categories of paradata. The first three include observations made by the interviewer about: (1) the sampled unit's neighborhood, (2) the sampled unit's housing unit, or (3) persons in the sampled housing unit. The next two involve data that can change over the observation period: (4) call record information collected in the sample management system and (5) observations recorded by the interviewer about their interaction with the sampled household at each interview. In considering how such data might be used for nonresponse adjustment, it should be noted that the first three categories and to some extent the fourth identify variables that could be observed for respondents and non-respondents alike whereas the fifth category is available only for respondents. It should also be noted that the quality of paradata is generally very high when it is generated automatically (for example, by a call attempt made through an interviewer's computer). Quality is less high for items that interviewers are asked to collect that fall well outside of the regular data collection process (NRC 2013). Items of paradata that are recorded for both respondents and non-respondents provide covariates for modeling

survey response. Krueger and West (2014) illustrate the use of paradata in nonresponse adjustments for the NSFG.

Interest in paradata has been growing for some time. In addition to their potential use for nonresponse adjustments, paradata provide key indicators that can be monitored in applications of responsive design. Paradata can also be used in research to improve our understanding of the behavior of survey participation. For example, paradata enable the identification of respondents who were difficult to convert, allowing comparison of their survey characteristics with those of other respondents (NRC 2013).

**2. Data on the entire population**

Sources of data on the entire population include population estimates by demographic characteristics, totals obtained from surveys with very large samples, and totals obtained from administrative records. Demographic estimates produced by the Census Bureau have provided the data for adjusting the weights for the CPS and other federal surveys. For SIPP, the Census Bureau has used the CPS ASEC to obtain approximate totals for a broader set of variables. Certain other federal surveys use the CPS ASEC in their weighting adjustments as well. The ACS, with its greater precision for a comparably broad set of variables, may be supplanting the CPS ASEC for this purpose. Administrative records can provide even more precise totals than the ACS but only for subsets of the population—some of them quite small.

## C. New approaches

Can HHS and other federal agencies reduce their reliance on probability surveys by developing other data sources or using surveys differently? Suggestions that have been offered include making greater use of administrative records, learning how to extract useful information from "big data," and applying statistical methods that treat survey data as derived from nonprobability samples.

### 1. Administrative records

Citro (2014) argues that some of the growing shortcomings of federal survey data can be addressed by substituting data drawn from administrative records for the data that respondents are unwilling to report or appear unable to provide with sufficient accuracy. As examples of how this might work she cites household income and housing characteristics, including plumbing facilities. She notes that in key surveys, Statistics Canada gives its respondents the option of skipping sections of their questionnaires and allowing the agency to substitute administrative records for the data that would have been collected in the survey. Such applications of administrative records are rare in the U.S., but they do exist. The Energy Information Administration (EIA) contracts with local electric utilities to provide data on household electrical energy usage and expenditures by its survey households instead of requesting this information from respondents. For internal use, researchers in the Social Security Administration (SSA) routinely substitute SSA data on social security and SSI benefit receipt for survey responses in the SIPP. Unlike the EIA application, however, this practice by SSA does not reduce the reporting burden on respondents.

Both the EIA and SSA applications represent examples where there is perfect conceptual alignment between the data maintained in administrative records and what the survey organizations would collect if they could. This is not always the case. For instance, the earnings reported to the Internal Revenue Service and the gross earnings typically requested in surveys differ, and the differences have grown over time as revisions to the tax code have allowed employees to exclude increasing amounts of their income from taxation. When agencies substitute administrative records for the items that they would otherwise collect in their surveys, they lose a certain degree of control over the content of the information they obtain (Czajka 2013).

**2. Big data**

Big data is mentioned in a number of contexts as a potentially rich source of information that can be acquired unobtrusively and at relatively low cost. Examples of the collection of such data include "Internet scraping" or data mining. The CNSTAT report (NRC 2013) provides several examples of the use of such techniques to capture economic and social statistics. While the cost of developing the extraction tools and models to effectively use such data may not be negligible, they very likely pale in comparison with the cost of conducting a large survey. Because of the many challenges in developing ways to extract information from big data, however, Citro (2014) recommends that federal statistical agencies be "'close followers' rather than leaders in using big data."

**3. Non-probability surveys**

If the problem created by unit nonresponse is that we are left with something less than a probability sample, then perhaps the correct reaction is to acknowledge that we have a nonprobability sample and treat it accordingly. Fahimi et al. (2015) review alternative practices that are being used to develop inferences from samples that violate the traditional assumptions of probability sampling with respect to coverage and response. Under these circumstances common methods of adjustment based on demographic and geographic characteristics are likely to be ineffectual. The authors expand the set of variables used to calibrate the survey weights to include behavioral and attitudinal measures obtained from federal surveys and other sources, such as election outcomes. Wang et al. (2014) provide an example of the use of a highly non-representative dataset—obtained from X-Box users who were solicited to respond through their gaming devices—to develop presidential election forecasts that were in line with those of top poll analysts, who combined hundreds of traditional election polls conducted during the lead-up to the 2012 election.

# VI. CONCLUSION

Concerns about growing survey nonresponse and its potential impact upon nonresponse bias are not new, but federal household surveys have tended to enjoy high response rates until relatively recently. The response rate trends presented in Chapter III highlight increased rates of decline in response rates among the major federal surveys in very recent years. They underscore the fact that declining response rates are not unique to particular surveys; they are widespread, although some surveys have experienced much greater declines than others.

Much has been learned about the nature of nonresponse bias. Two important findings are that nonresponse bias is generally unrelated to the overall response rate and that nonresponse bias tends to be item-specific. Some or even many items may exhibit no bias while others have substantial bias.

A troubling implication of the item-specific nature of nonresponse bias is potential need for item-specific nonresponse adjustments. Common methods of adjusting for nonresponse, which involve just a few dimensions, are not likely to be effective for the many items that are not highly correlated with these dimensions.

**This page has been left blank for double-sided copying.**

## REFERENCES

American Association for Public Opinion Research. *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys*. 8th Edition. AAPOR, 2015.

Atrostic, B. K., Nancy Bates, Geraldine Burt, and Adrian Silberstein. "Nonresponse in U.S. Government Household Surveys: Consistent Measures, Recent Trends, and New Insights." *Journal of Official Statistics*, vol. 17, no. 2 (2001), pp. 209-226.

Bee, C. Adam, Graton M. R. Gathright, and Bruce D. Meyer. "Bias from Unit Non-Response in the Measurement of Income in Household Surveys." In *Proceedings of the 2015 FCSM Research Conference.* Washington, DC: Federal Committee on Statistical Methodology, 2016.

Bollinger, Christopher and Martin H. David. "Estimation with Response Error and Nonresponse: Food-Stamp Participation in the SIPP." *Journal of Business and Economic Statistics*, vol. 19, no. 2 (2001), pp. 129-141.

Bradburn, Norman H. "A Response to the Nonresponse Problem." 1992 AAPOR Presidential Address. *Public Opinion Quarterly*, vol. 56, no. 3 (1992), pp. 391-397.

Brick, J. Michael and Douglas Williams. "Explaining Rising Nonresponse Rates in Cross-Sectional Surveys." *Annals of the American Academy of Political and Social Science*, volume 645 (January 2013), pp. 36-59.

Cantor, David, Kisha Coa, Susan Crystal-Mansour, Terisa Davis, Sarah Dipko, and Richard Sigman. "Health Information National Trends Survey (HINTS) 2007." Final Report. Rockville, MD: Westat, February 2009. Available at: http://hints.cancer.gov/docs/HINTS2007FinalReport.pdf [February 2016]

Center for Behavioral Health Statistics and Quality. *2013 National Survey on Drug Use and Health: Methodological Resource Book (Section 8, Data Collection Final Report).* Rockville, MD: Substance Abuse and Mental Health Services Administration, 2014. Available at: http://www.samhsa.gov/data/sites/default/files/NSDUHmrbDCFR2013.pdf [January 2016]

Center for Behavioral Health Statistics and Quality. *Behavioral Health Trends in the United States: Results from the 2014 National Survey on Drug Use and Health* (HHS Publication No. SMA 15-4927, HSDUH Series H-50). Rockville, MD: Substance Abuse and Mental Health Services Administration, 2015. Available at: http://www.samhsa.gov/data/sites/default/files/NSDUH-FRR1-2014/NSDUH-FRR1-2014.pdf [January 2016]

Citro, Constance F. "From Multiple Modes for Surveys to Multiple Data Sources for Estimates." *Survey Methodology*, vol. 40, no. 2 (2014), pp. 137-161.

Couper, Mick P. "The Future of Modes of Data Collection." *Public Opinion Quarterly*, vol. 75, no. 5 (Special Issue 2011), pp. 889-908.

Curtin, Richard, Stanley Presser, and Eleanor Singer. "The Effects of Response Rates Changes on the Index of Consumer Sentiment." *Public Opinion Quarterly*, vol. 64, no. 4 (2000), pp. 413-428.

Curtin, Richard, Stanley Presser, and Eleanor Singer. "Changes in Telephone Survey Nonresponse over the Past Quarter Century." *Public Opinion Quarterly*, vol. 69, no. 1 (2005), pp. 87-98.

Czajka, John L. "SIPP Data Quality." Appendix A in *Reengineering the Survey of Income and Program Participation*. National Research Council. Washington, DC: The National Academies Press, 2009.

Czajka, John L. "Can Administrative Records Be Used to Reduce Nonresponse Bias?" *Annals of the American Academy of Political and Social Science*, volume 645 (January 2013), pp. 171-184.

De Leeuw, Edith D. "To Mix or Not to Mix Data Collection Modes in Surveys." *Journal of Official Statistics*, vol. 21, no. 2 (2005), pp. 233-255.

Dohrmann, Sylvia, Graham Kalton, Jill Montaquila, Cindy Good, and Martha Berlin. "Using Address Based Sampling Frames in Lieu of Traditional Listing: A New Approach." *Proceedings of the Joint Statistical Meetings.* Alexandria, VA: American Statistical Association, 2012, pp. 3729-3741.

Fahimi, Mansour, Frances M. Barlas, Randall K. Thomas, and Nicole Buttermore. "Scientific Surveys Based on Incomplete Sampling Frames and High Rates of Nonresponse." *Survey Practice*, vol. 8, no. 5 (2015).

Fricker, Scott and Roger Tourangeau. "Examining the Relationship between Nonresponse Propensity and Data Quality in Two National Household Surveys. *Public Opinion Quarterly*, vol. 74, no. 5 (Special Issue 2010), pp. 935-955.

Groves, Robert M. and Steven Heeringa. "Responsive Design for Household Surveys: Tools for Actively Controlling Survey Errors and Costs." *Journal of the Royal Statistical Society Series A: Statistics in Society*, vol. 169, Part 3 (2006), pp. 439-457.

Groves, Robert M. and E. Peytcheva. "The Impact of Nonresponse Rates on Nonresponse Bias: A Meta-analysis. *Public Opinion Quarterly*, vol. 72, no. 2 (Summer 2008), pp. 167-189.

Iannacchione, Vincent G. "Research Synthesis: The Changing Role of Address-Based Sampling in Survey Research." *Public Opinion Quarterly*, vol. 75, no. 3 (Fall 2011), pp. 556-575.

Keeter, Scott, Carolyn Miller, Andrew Kohut, Robert M. Groves, and Stanley Presser. "Consequences of Reducing Nonresponse in a Large National Telephone Survey." *Public Opinion Quarterly*, vol. 64, no. 2 (2000), pp. 125-148.

Kreuter, Frauke. "Facing the Nonresponse Challenge." *Annals of the American Academy of Political and Social Science*, volume 645 (January 2013), pp. 23-35.

Kreuter, Frauke, Gerrit Muller, and Mark Trappmann. "A Note on Mechanisms Leading to Lower Data Quality of Late or Reluctant Respondents." *Sociological Methods and Research*, vol. 43, no. 3, (2014), pp. 452-464.

Kreuter, Frauke, Gerrit Muller, and Mark Trappmann. "Nonresponse and Measurement Error in Employment Research." *Public Opinion Quarterly*, vol. 74, no. 5 (2010), pp. 880-906.

Krueger, Brian S. and Brady T. West. "Assessing the Potential of Paradata and Other Auxiliary Data for Nonresponse Adjustments." *Public Opinion Quarterly*, vol. 78, no. 4 (Winter 2014), pp. 795-831.

Link, Michael W., Michael P. Battaglia, Martin R. Frankel, Larry Osborn, and Ali H. Mokdad. "Comparison of Address-based Sampling (ABS) versus Random-Digit Dialing (RDD) for General Population Surveys." *Public Opinion Quarterly*, vol. 72, no. 1 (Spring 2008), pp. 6-27.

Meyer, Bruce D., Wallace K. C. Mok, and James X. Sullivan. "Household Surveys in Crisis." *Journal of Economic Perspectives*, vol. 29, no. 4 (Fall 2015), pp. 199-226.

National Research Council. *The Future of Federal Household Surveys: Summary of a Workshop.* Krisztina Marton and Jennifer C. Karberg, Rapporteurs. Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press, 2011.

National Research Council. *Nonresponse in Social Science Surveys: A Research Agenda.* Roger Tourangeau and Thomas J. Plewes, Editors. Panel on a Research Agenda for the Future of Social Science Data Collection, Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press, 2013.

Navarro, Alfredo, Karen E. King, and Michael Starsinic. "Comparison of the American Community Survey Voluntary versus Mandatory Estimates." Final report. Washington, DC: U.S. Census Bureau, September 27, 2011.

Olson, Kristen. "Paradata for Nonresponse Adjustment." *Annals of the American Academy of Political and Social Science*, volume 645 (January 2013), pp. 142-170.

Peytchev, Andy. "Consequences of Survey Nonresponse." *Annals of the American Academy of Political and Social Science*, volume 645 (January 2013), pp. 88-111.

Peytchev, Andy, Rodney K. Baxter, and Lisa R. Carley-Baxter. "Not All Survey Effort Is Equal: Reduction of Nonresponse Bias and Nonresponse Error." *Public Opinion Quarterly*, vol. 73, no. 4 (2009), pp. 785-806.

Presser, Stanley and Susan McCullogh. "The Growth of Survey Research in the United States: Government-Sponsored Surveys, 1984-2004." *Social Science Research*, vol. 40, no. 4 (2011), pp. 1019-1024.

Roth, Shelley Brock, Daifeng Han, and Jill M. Montaquila. "The ABS Frame: Quality and Considerations." *Proceedings of the Joint Statistical Meetings.* Alexandria, VA: American Statistical Association, 2012, pp. 3779-3793.

Sarndal, Carl-Erik. "The 2010 Morris Hansen Lecture Dealing with Survey Nonresponse in Data Collection, in Estimation. *Journal of Official Statistics*, vol. 27, no. 1 (2011), pp. 1-21.

Schoeni, Robert F., Frank Stafford, Katherine A. McGonagle, and Patricia Andreski. "Response Rates in National Panel Surveys." *Annals of the American Academy of Political and Social Science*, volume 645 (January 2013), pp. 60-87.

Schouten, Barry, Fannie Cobben, and Jelke G. Bethlehem. "Indicators for the Representativeness of Survey Response." *Survey Methodology*, vol. 35, no. 1 (2009), pp. 101-113.

Singer, Eleanor and Cong Ye. "The Use and Effects of Incentives in Surveys." *Annals of the American Academy of Political and Social Science*, volume 645 (January 2013), pp. 112-141.

Smith, Thomas. "Trends in Nonresponse Rates." *International Journal of Public Opinion Research*, vol. 7, number 2 (1995), pp. 157-171.

U.S. Census Bureau. "Current Population Survey, Design and Methodology." Technical Paper 66. Washington, DC: U.S. Department of Commerce, October 2006.

U.S. Census Bureau. "Source and Accuracy Statement for the 2014 Income Consistent Annual Social and Economic Supplement Microdata File." Washington, DC: U.S. Department of Commerce, September 2015. Available at: http://www.census.gov/housing/extract_files/data%20extracts/cpsasec14/2014%20CPS%20ASEC%20Income%20Consistent%20Source%20and%20Accuracy.pdf [January 2016]

Vannieuwenhuyze, Jorre, Geert Loosveldt, and Geert Molenberghs. "A Method for Evaluating Mode Effects in Mixed-Mode Surveys." *Public Opinion Quarterly*, vol. 74, no. 5 (Special Issue 2010), pp. 1027-1045.

Wagner, James. "A Comparison of Alternative Indicators for the Risk of Nonresponse Bias." *Public Opinion Quarterly*, vol. 76, number 3 (Fall 2012), pp. 555-575.

Wagner, James, Brady T. West, Nicole Kirgis, James M. Lepkowski, William G. Axinn, and Shonda Kruger Ndiaye. "Use of Paradata in a Responsive Design Framework to Manage a Field Data Collection." *Journal of Official Statistics*, vol. 28, no. 4 (December 2012), pp. 477-499.

Walejko, Gina and Peter Miller. "The 2013 Census Test: Piloting Methods to Reduce 2020 Census Costs." *Survey Practice*, vol. 8, no. 5 (2015).

Wang, Wei, David Rothschild, Sharad Goel, and Andrew Gelman. "Forecasting Elections with Non-representative Polls." *International Journal of Forecasting*, 2014. Available at: http://dx.doi.org/10.1016/j.ijforecast.2014.06.001 [January 2016]

# APPENDIX A

# SURVEY DESCRIPTIONS AND RESPONSE RATES

**This page has been left blank for double-sided copying.**

## A. Overview of the surveys

The NHIS collects data on the health of the civilian non-institutional population in order to monitor the incidence and prevalence of illness and disability. Because illnesses occur at varying rates throughout the year, the NHIS is administered throughout the year as well. The survey is conducted by the U.S. Census Bureau using computer-assisted personal interviewing (CAPI), with computer-assisted telephone interviewing (CATI) follow-up. The survey consists of four components. The household component captures demographic and relationship information for all members in a household. The family component is administered separately to each family in a household and captures socio-demographic characteristics along with a range of health status indicators and utilization. One child and one adult in each household are selected at random for the sample child and sample adult surveys, respectively. These surveys collect information on healthcare service utilization and behavior. The NHIS underwent a major redesign in 1997, which is where we start our response rate series.

NHANES data collection, which is conducted by Westat, includes both a personal interview and a physical examination, with the latter carried out using mobile examination facilities. A household-level screening interview is conducted to identify members eligible for the interview and physical examination and record their demographic characteristics. From this list a sample is selected to meet targets defined by demographic characteristics (some groups are over-sampled, for instance). The sample is selected in such a way that households with at least one member selected tend to have more than one member selected. As a result, many households that complete the screener have no members selected. The personal interview is conducted using CAPI. Data for young children are collected from a parent. The questionnaire covers a wide range of health topics, which vary with the age of the respondent. For example, a component on cognitive functioning is administered only to persons 60 and older, whereas dietary information

is collected from respondents of all ages. The examination, which also includes a questionnaire, captures data from a number of medical tests, which have varied over the years. Biospecimens are collected and subjected to laboratory analysis as well.

BRFSS is administered by state health departments in all the states, the District of Columbia, and Puerto Rico, Guam, and the U.S. Virgin Islands. The purpose of the survey is to collect data on preventative health and risk behaviors for the adult population (18 and older). The BRFSS questionnaire includes a standard component, which is used in all areas; optional modules prepared by CDC; and questions that the states develop independently. BRFSS is conducted by telephone, using RDD methods. The survey work is carried out by contractors working for the states.

The NIS is an RDD telephone survey used to measure vaccination coverage among children at the ages of 19 to 35 months. Started in late 1994, the NIS covers the U.S., Puerto Rico, Guam, and the Virgin Islands. Parents or guardians who complete an interview about their age-eligible children are asked for the names of and permission to contact their children's vaccination providers. Questionnaires mailed to the providers collect details about the children's vaccinations and the facilities in which they were administered. NORC conducts the survey, which, like BRFSS, added a cell phone sample in 2011. NIS households are also screened for age-eligible children for two related surveys: the NIS-Teen and the NIS-Child Influenza Module. The teen survey, which samples adolescents 13 to 17 years of age, was first administered in the last quarter of 2006 and was expanded to the full year in 2008. The NIS-Child Influenza Module, samples children 6 to 18 months and 3 to 12 years of age who were not included in the NIS or

NIS-Teen.[12] This component of the NIS is administered between October and June. During those months, its questions on flu incidence are included in the NIS and NIS-Teen questionnaires as well.

Originally named the National Household Survey on Drug Abuse, NSDUH began in 1971 with the objective of collecting information on the use of alcohol, illegal drugs, and tobacco from the adult population (18 and older). Its content was later expanded to include mental disorders and the receipt of treatment for substance abuse and mental health. Beginning in 1999, the survey transitioned from a paper and pencil questionnaire to a computer-assisted interview. Because of their personal nature, most questions are administered using audio computer-assisted self-interviewing (ACASI). In 1999 the survey was expanded to include samples from all 50 states and the District of Columbia, and the name was changed to NSDUH to better reflect its content and potentially increase cooperation. Beginning in 2002, respondents were given a $30 incentive to encourage participation. From each sample household, NSDUH attempts to obtain from an adult resident a screening interview, from which zero, one, or two household members 12 and older are selected by computer for a detailed interview.

The MCBS is a panel survey of individual Medicare enrollees sampled from program administrative files and representative of persons enrolled as of January 1 of each year. The sample includes residents of both non-institutional and institutional quarters except for prisons and facilities for the criminally insane.[13] For each panel the survey collects three calendar years of data through interviews conducted at four-month intervals over four years—for 12 interviews

[12] While the nominal ages of children covered by the Child Influenza Module do not overlap those covered by the other two surveys, children are eligible for selection into the NIS if they are a few within a few weeks of the qualifying ages.

[13] Institutional quarters in the MCBS sample include a variety of facilities for long-term care, such as nursing homes, mental hospitals, intermediate care facilities, board and care homes, and group homes.

in all. The data were collected under contract by Westat until 2014, when NORC at the University of Chicago assumed that role. New panels are initiated annually to add new enrollees, replace participants expected to be lost or to have completed their data collection, and maintain the sample size and age stratification. Enrollees under 45 and 80 and older are over-represented, and the sample includes Puerto Rico. Interviews are conducted face-to-face. Cost and utilization data from Medicare claims files are linked to the survey data along with information on expenditures for non-covered medical services.

Like the MCBS, MEPS is a longitudinal survey. The sample for each panel is drawn from households that participated in the NHIS in the prior year. MEPS started in 1996, and the most recent panel, selected from the 2015 NHIS, is the 21st. The household component of the survey, MEPS-HC, which is our sole focus, involves five interviews that collect data for two full calendar years on topics that include health insurance coverage, employment, income, and, most importantly, on the type, frequency, and cost of healthcare. The MEPS-HC data collection is conducted by Westat using CAPI methods.

The CPS is a monthly survey whose primary purpose is to measure labor force participation. In that capacity it provides the official monthly estimate of the unemployment rate as well as a number of related statistics on the work force. The monthly labor force questionnaire is very brief, but in most months there is also a supplement devoted to a particular topic. The Annual Social and Economic (ASEC) Supplement, which is administered in March and to parts of the sample in February and April, collects data on household composition and income, migration, and health insurance coverage and is the official source of annual estimates of the poverty rate. Housing units (addresses) that are selected into the monthly CPS sample remain in the sample for four consecutive months, are out for the next eight, and then return for another four months.

In any given month of the year, one-eighth of the sample is in its first month, another eighth is in its second month, and so on. The Census Bureau attempts to conduct the first and fifth interviews in person and the remaining interviews by phone. Finally, as part of a sample expansion for the CPS ASEC, some households that have completed their eighth month or are between their fourth and fifth months are brought back into the sample to be interviewed for the supplement. These households are administered the monthly labor force questionnaire as well, but their responses do not contribute to the official labor force statistics for that month.

## B. Response rate trends by survey

Each of the seven surveys shown in Figure III.1 has more than one module (or, for MCBS and MEPS, more than one wave), for which separate response rates can be computed. These response rates, along with response rates for BRFSS, are presented and discussed below.

### 1. NHIS

Table A.1 reports response rates from 1997 through 2014 for the separate NHIS modules, beginning with the household module, which was shown in Figure III.1 and is used to identify respondents for the remaining portions of the survey.[14] The next three columns report the response rates for the family, sample child, and sample adult modules, conditional on having responded to the household module. The final three columns report the unconditional response rates for the three modules—that is, the product of the response rate to the household module and the conditional response rate to each module.

[14] In this table and in subsequent tables the response rates shown in Figure III.1 are bolded.

## Table A.1. Response rates for the National Health Interview Survey, 1997 to 2014

| Survey Year | Household Module | Conditional Family Module | Conditional Sample Child Module | Conditional Sample Adult Module | Unconditional Family Module | Unconditional Sample Child Module | Unconditional Sample Adult Module |
|---|---|---|---|---|---|---|---|
| 1997 | **91.8** | 98.4 | 93.1 | 89.0 | 90.3 | 84.1 | 80.4 |
| 1998 | **90.0** | 98.0 | 93.3 | 83.8 | 88.2 | 82.4 | 73.9 |
| 1999 | **87.6** | 98.3 | 90.8 | 80.8 | 86.1 | 78.2 | 69.6 |
| 2000 | **88.9** | 98.2 | 90.9 | 82.6 | 87.3 | 79.4 | 72.1 |
| 2001 | **88.9** | 98.5 | 92.0 | 84.2 | 87.6 | 80.6 | 73.8 |
| 2002 | **89.6** | 98.3 | 92.3 | 84.4 | 88.1 | 81.3 | 74.3 |
| 2003 | **89.2** | 98.5 | 92.3 | 84.5 | 87.9 | 81.1 | 74.2 |
| 2004 | **86.9** | 99.6 | 91.8 | 83.8 | 86.5 | 79.4 | 72.5 |
| 2005 | **86.5** | 99.5 | 90.1 | 80.1 | 86.1 | 77.5 | 69.0 |
| 2006 | **87.3** | 99.6 | 90.6 | 81.4 | 87.0 | 78.8 | 70.8 |
| 2007 | **87.1** | 99.4 | 88.4 | 78.3 | 86.6 | 76.5 | 67.8 |
| 2008 | **84.9** | 99.5 | 85.6 | 74.2 | 84.5 | 72.3 | 62.6 |
| 2009 | **82.2** | 99.3 | 89.9 | 80.1 | 81.6 | 73.4 | 65.4 |
| 2010 | **79.5** | 99.1 | 89.8 | 77.3 | 78.7 | 70.7 | 60.8 |
| 2011 | **82.0** | 99.2 | 91.8 | 81.6 | 81.3 | 74.6 | 66.3 |
| 2012 | **77.6** | 99.0 | 90.7 | 79.7 | 76.8 | 69.7 | 61.2 |
| 2013 | **75.7** | 99.0 | 92.1 | 81.7 | 74.9 | 69.0 | 61.2 |
| 2014 | **73.8** | 99.0 | 91.2 | 80.5 | 73.1 | 66.6 | 58.9 |

Source: Adapted from multiple annual issues of the NHIS Survey Description report, available by choosing individual years from ftp://ftp.cdc.gov/pub/Health_statistics/NCHS/Dataset_Documentation/NHIS/ [January 2016].

Given that the respondent to the household module is usually the respondent to the family module as well, the conditional response rate to the family module remains close to 100 percent throughout the period. The conditional response to the sample child module declines from just over 93 percent in 1997 to a low of 85.6 percent in 2008 but then rises to rates above 90 percent for the final four years. The conditional response to the sample adult module declines sharply between 1997 and 1999, falling from 89.0 to 80.8 percent. The response shows some volatility after that but remains around 80 percent on average. Notably, while the household response rate shows a steep decline between 2011 and 2014, the conditional response rates to the three modules are essentially flat over this period.

The unconditional response rates are largely driven by the downward trend in the household response rate, although the decline in the conditional adult response rate over parts of the period amplifies the decline in the unconditional adult response rate, which falls from a peak of 80.4 percent in 1997 to a low of 58.9 percent in 2014.

**2. NHANES**

NCHS reports NHANES response rates for both the personal interview and examination conditional on being selected into the sample. NCHS does not report response rates for the household screening interview and does not report unconditional response rates for either the interview or examination.

Prior to 1999, each NHANES was conducted over a three to four-year period, for which a single set of response rates was calculated. The survey became annual in 1999, but because the annual samples are too small to provide satisfactory precision for some estimates, the data have been released in two-year cycles, and response rates have been calculated combining the two years. From 1999-2000 to 2013-2014, the conditional response rate to the interview declined

from 81.9 to 71.0 percent while the response rate for the examination declined from 76.3 to 68.5 percent (Table A.2). For both rates, nearly all of the decline occurred after the 2009-2010 period.

**Table A.2. Response rates for the National Health and Nutrition Examination Survey, 1999 to 2014**

| Survey Year | Conditional Interview | Conditional Examination |
|---|---|---|
| 1999-2000 | **81.9** | 76.3 |
| 2001-2002 | **83.9** | 79.6 |
| 2003-2004 | **79.3** | 75.6 |
| 2005-2006 | **80.5** | 77.4 |
| 2007-2008 | **78.4** | 75.4 |
| 2009-2010 | **79.4** | 77.3 |
| 2011-2012 | **72.6** | 69.5 |
| 2013-2014 | **71.0** | 68.5 |

Source: Obtained from http://www.cdc.gov/nchs/nhanes/response_rates_cps.htm [January 2016].

Note: Response rates are conditional on being selected into the within-household sample, done in the household screening interview. Response rates for the screening interview are not reported, so unconditional response rates cannot be calculated.

### 3. NIS

As an RDD telephone survey, the NIS includes a first stage to determine if each sampled phone number connects to a household. Phone numbers are resolved if they are found to either represent households or not represent households (for example, they have been disconnected or they connect to businesses). For phone numbers that connect to households the next stage involves a screener interview to determine if a child 19 to 35 months is included in the household. If so, an interview is sought with a parent or guardian. Conditional rates for these three stages are reported in Table A.3 followed by a combined, CASRO rate, representing the unconditional response rate to the interview (the product of the three rates).

All three conditional rates declined between 1995 and 2014, with the resolution rate declining the most, falling from 96.5 percent to 82.7 percent. Nearly all of this decline occurred over the first nine years. The screener completion rate declined by only 4 percentage points over the period while the conditional interview completion rate declined by 12 percentage points. The

combined, CASRO rate declined by 22 percentage points between 1995 and 2005. Since then, however, it has declined only modestly, reaching levels between 62 and 63 percent in 2013 and 2014 compared to 87 percent in 1995.

**Table A.3. Landline response rates for the National Immunization Survey, 1995 to 2014**

| Survey Year | Resolution Rate | Screener Completion Rate | Interview Completion Rate | CASRO Response Rate |
|---|---|---|---|---|
| 1995 | 96.5 | 96.4 | **93.5** | 87.1 |
| 1996 | 94.3 | 96.8 | **94.0** | 85.8 |
| 1997 | 92.1 | 97.9 | **93.8** | 84.6 |
| 1998 | 90.4 | 97.8 | **93.6** | 82.7 |
| 1999 | 88.6 | 97.0 | **93.4** | 80.2 |
| 2000 | 88.1 | 96.0 | **93.1** | 78.7 |
| 2001 | 86.8 | 96.2 | **91.1** | 76.1 |
| 2002 | 84.8 | 96.6 | **90.6** | 74.2 |
| 2003 | 83.6 | 94.0 | **88.7** | 69.8 |
| 2004 | 83.8 | 94.8 | **92.0** | 73.1 |
| 2005 | 83.3 | 92.8 | **84.2** | 65.1 |
| 2006 | 83.3 | 90.5 | **85.6** | 64.5 |
| 2007 | 82.9 | 90.2 | **86.8** | 64.9 |
| 2008 | 82.3 | 90.3 | **85.1** | 63.2 |
| 2009 | 82.9 | 92.4 | **83.2** | 63.8 |
| 2010 | 83.8 | 91.5 | **83.6** | 63.8 |
| 2011 | 83.0 | 90.7 | **81.7** | 61.6 |
| 2012 | 84.1 | 90.7 | **84.6** | 64.5 |
| 2013 | 83.2 | 91.0 | **82.3** | 62.3 |
| 2014 | 82.7 | 92.2 | **82.1** | 62.6 |

Source: NORC at the University of Chicago and Centers for Disease Control and Prevention, A User's Guide for the 2014 Public-Use Data File, Table G.1. Available at: ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Dataset_Documentation/NIS/NISPUF14_DUG.PDF [February 2016].

Note: A cellular phone frame was added in 2011. The response rates for 2011 to 2014 reflect the landline sample only. The CASRO response rate is calculated as the product of the response rates in the first three columns. Because of rounding, the reported CASRO rate may differ from a computation based on the values reported in this table.

### 4. NSDUH

Prior to 1999, all NSDUH interviews were conducted with paper and pencil. In 1999, part of the sample was interviewed with a computer-assisted instrument, which fully replaced paper and

pencil in 2000. Response rates to the screening interview and the detailed interview, both conditional and unconditional, are reported in Table A.4, beginning with the computer-assisted interview rates in 1999 and continuing through 2014. While response rates to the household screening interview declined from the low 90s in the early years to a low of 82.6 percent in 2014, conditional response rates to the detailed interview showed no net change over the period. The rates rose to a peak of 84.5 percent in 2002 before falling back to 74 percent. The unconditional response rate to the interview peaked in 2002 as well, reaching 76.8 percent, but declined to 61.2 percent in 2014 or nearly 6 percentage points below the rate in 1999.

**Table A.4. Response rates for the National Survey on Drug Use and Health, 1999 to 2014**

| Survey Year | Screening | Conditional Interview | Unconditional Interview |
|---|---|---|---|
| 1999 | **90.1** | 74.2 | 66.8 |
| 2000 | **93.0** | 78.0 | 72.6 |
| 2001 | **91.8** | 76.8 | 70.5 |
| 2002 | **90.8** | 84.5 | 76.8 |
| 2003 | **91.0** | 83.0 | 75.6 |
| 2004 | **91.3** | 82.7 | 75.4 |
| 2005 | **91.3** | 81.5 | 74.4 |
| 2006 | **90.2** | 79.7 | 71.9 |
| 2007 | **88.8** | 79.0 | 70.2 |
| 2008 | **88.8** | 79.3 | 70.4 |
| 2009 | **88.6** | 80.2 | 71.0 |
| 2010 | **88.3** | 79.8 | 70.4 |
| 2011 | **87.0** | 79.2 | 68.9 |
| 2012 | **86.2** | 77.9 | 67.1 |
| 2013 | **84.4** | 76.4 | 64.5 |
| 2014 | **82.6** | 74.1 | 61.2 |

Source: Year 1999 response rates obtained from Gfroerer et al. (2002), available at: http://media.samhsa.gov/data/NHSDA/redesigningNHSDA.pdf [January 2016]. Year 2000-2012 response rates obtained from multiple annual issues of the NSDUH Data Collection Final Report, available by choosing individual years from http://media.samhsa.gov/data/Methodological_Reports.aspx [January 2016]. Year 2013-2014 response rates obtained from Center for Behavioral Health Statistics and Quality (2014, 2015).

Note: In 1999, the sample was split between paper and pencil--the sole mode in all prior years--and computer-assisted interviews, the only mode in subsequent years. Only the computer-assisted interview response rate is reported for 1999.

## 5. MCBS

In panel surveys, the response rate in a given round among those who responded in the previous round provides an important measure of retention. Table A.5 reports conditional response rates to the 2nd, 6th, and 12th MCBS interviews in addition to the response rate to the initial interview. Response rates to the initial interview stabilized at between 83 and 84 percent for 12 years after the survey's debut in 1991 with an initial response rate of 87.2 percent. After 2003 the initial round response rates began to decline, reaching 73 percent in 2012 and 2013.

Like the panel surveys discussed in the previous chapter, the MCBS has shown much greater stability in its conditional response rates to the later rounds. Conditional response rates to the 2nd interview were 94 to 95 percent for the survey's first decade.[15] Since then they have declined gradually to recent levels around 88 percent. For the 6th round interview, the conditional response rates started at around 98 percent and remained there through 2005. They have declined only slightly, to 94.5 percent in 2012. Conditional response rates to the 12th and final interview have been above 99.0 percent since the second year of the survey.

[15] CMS does not have a description of how Westat calculated these conditional response rates. Sample members who miss an isolated interview are retained for future interviews, which means that the respondents to rounds four and later will include individuals who did not respond to the immediately preceding round. How such respondents are handled in the conditional response rate calculations is not clear.

**Table A.5. Response rates for the Medicare Current Beneficiary Survey, 1991 to 2013**

| Survey Year | Initial Interview | Conditional 2nd Interview | Conditional 6th Interview | Conditional 12th Interview |
|---|---|---|---|---|
| 1991 | 87.2 | 93.6 | 98.2 | 98.7 |
| 1992 | 84.3 | 95.1 | 98.6 | 99.4 |
| 1993 | 82.8 | 95.4 | 97.7 | 99.8 |
| 1994 | 82.8 | 94.6 | 98.6 | 99.9 |
| 1995 | **83.1** | 94.0 | 98.5 | 99.9 |
| 1996 | **83.4** | 95.0 | 98.4 | 99.7 |
| 1997 | **83.6** | 94.8 | 98.4 | 99.6 |
| 1998 | **83.3** | 94.5 | 98.0 | 99.7 |
| 1999 | **84.8** | 94.3 | 98.3 | 99.8 |
| 2000 | **84.3** | 94.1 | 97.8 | 99.7 |
| 2001 | **84.8** | 93.1 | 97.9 | 99.8 |
| 2002 | **84.3** | 92.9 | 97.9 | 99.8 |
| 2003 | **83.2** | 92.9 | 97.7 | 99.9 |
| 2004 | **82.2** | 92.8 | 98.3 | 99.8 |
| 2005 | **82.0** | 91.7 | 97.9 | 99.9 |
| 2006 | **82.8** | 92.2 | 97.1 | 99.9 |
| 2007 | **80.3** | 90.5 | 97.6 | 99.9 |
| 2008 | **78.0** | 90.7 | 97.7 | 99.8 |
| 2009 | **77.5** | 89.8 | 97.3 | 99.7 |
| 2010 | **77.5** | 89.5 | 95.5 | 99.6 |
| 2011 | **77.4** | 89.0 | 96.1 | |
| 2012 | **73.1** | 88.0 | 94.5 | |
| 2013 | **72.7** | 87.5 | | |

Source: Data provided to the Office of Enterprise Data and Analytics of CMS by Westat.

## 6. MEPS

As we have noted, the sample for each new MEPS panel is drawn from households responding to the NHIS in the prior year. Between 1996 and 2010 the conditional point-in-time year 1 response rate for households selected into the MEPS—that is, the initial or round 1 response rate—declined from 83.1 to 73.5 percent, with some fluctuation over the period (Table A.6). As we saw in Figure III.1, this decline in the conditional response rate was interrupted when the rate rose to 78.2 percent for 2011 and 2012. But the decline resumed after that point,

with the rate dropping to 71.8 percent by 2014. With the steep decline in the NHIS response and the more modest decline in the conditional MEPS response rate over the period, the unconditional MEPS response rate fell from 78.0 percent in 1996 to 54.7 percent in 2014.

**Table A.6. Response rates for the Medical Expenditure Panel Survey, Household Component, 1996 to 2014**

| Survey Year | NHIS Prior Year | Conditional Point-in-Time Year 1 | Unconditional Point-in-Time Year 1 |
|---|---|---|---|
| 1996 | 93.9 | **83.1** | 78.0 |
| 1997 | 93.8 | **83.0** | 77.9 |
| 1998 | 93.7 | **82.1** | 76.9 |
| 1999 | 92.2 | **80.3** | 74.0 |
| 2000 | 92.2 | **81.2** | 74.9 |
| 2001 | 89.9 | **81.2** | 73.0 |
| 2002 | 89.7 | **80.5** | 72.2 |
| 2003 | 90.1 | **79.8** | 71.9 |
| 2004 | 90.3 | **77.9** | 70.3 |
| 2005 | 87.9 | **79.0** | 69.4 |
| 2006 | 87.3 | **76.1** | 66.4 |
| 2007 | 88.1 | **76.5** | 67.4 |
| 2008 | 87.4 | **77.7** | 67.9 |
| 2009 | 85.2 | **74.8** | 63.7 |
| 2010 | 84.0 | **73.5** | 61.7 |
| 2011 | 80.6 | **78.2** | 63.0 |
| 2012 | 82.9 | **78.2** | 64.8 |
| 2013 | 78.0 | **74.2** | 57.9 |
| 2014 | 76.2 | **71.8** | 54.7 |

Source: Agency for Healthcare Research and Quality, MEPS-HC Response Rates by Panel, Table 2, available at: http://meps.ahrq.gov/mepsweb/survey_comp/hc_response_rate.jsp [January 2016].

Note: The NHIS prior year response rate is the response rate from the NHIS for the year preceding the indicated survey year. Households interviewed in the prior year's NHIS provide the frame for MEPS.

## 7. CPS ASEC

The ASEC supplement is substantially longer than the monthly labor force or basic CPS questionnaire, which it follows, and the field staff understand that completing the labor force questionnaire has a higher priority than completing the ASEC supplement. A fraction of

respondents who complete the labor force questionnaire break off the interview before or during the ASEC supplement. Nevertheless, the conditional ASEC response rate remained between 91 and 93 percent from 1997 through 2011 while the unconditional ASEC response rate deviated little from 84 percent (Table A.7). Like the response rate to the basic CPS, however, the conditional ASEC response rate declined after 2011, dropping from 92.0 to 85.8 percent in 2015. With this decline in both the basic CPS and conditional ASEC response rate, the unconditional ASEC response rate fell from 83.8 percent in 2011 to 76.3 percent in 2015.

**Table A.7. Response rates for the Current Population Survey Annual Social and Economic Supplement, 1997 to 2015**

| Survey Year | Basic CPS | Conditional ASEC | Unconditional ASEC |
|---|---|---|---|
| 1997 | **92.8** | 90.8 | 84.3 |
| 1998 | **92.2** | 92.8 | 85.6 |
| 1999 | **92.1** | 91.1 | 83.9 |
| 2000 | **93.0** | 92.0 | 85.6 |
| 2001 | **92.0** | 91.5 | 84.2 |
| 2002 | **91.7** | 91.4 | 83.8 |
| 2003 | **92.3** | 92.0 | 85.0 |
| 2004 | **91.5** | 91.8 | 84.0 |
| 2005 | **90.6** | 91.2 | 82.6 |
| 2006 | **91.1** | 91.5 | 83.3 |
| 2007 | **90.6** | 91.5 | 82.9 |
| 2008 | **91.4** | 92.3 | 84.4 |
| 2009 | **92.2** | 93.0 | 85.8 |
| 2010 | **92.2** | 93.1 | 85.9 |
| 2011 | **91.1** | 92.0 | 83.8 |
| 2012 | **90.2** | 89.6 | 80.8 |
| 2013 | **89.3** | 89.2 | 80.0 |
| 2014 | **89.8** | 88.7 | 79.6 |
| 2015 | **86.6** | 85.8 | 76.3 |

Source: Adapted from multiple annual issues of the CPS ASEC Technical Documentation, available by choosing individual years from http://www.nber.org/data/current-population-survey-data.html [January 2016]. Year 2014 response rates obtained from U.S. Census Bureau (2015).

Note: The Basic CPS response rate is calculated from the households responding to the labor force portion of the survey in March.

## 8. BRFSS

Because the BRFSS data collection is administered by the state health agencies and response rates vary markedly across the states, CDC reports the median state response rate along with the minimum and maximum state response rates in each year. The series presented in Table A.8 starts in 1997 as that is the first year that all 50 states and DC participated in the survey. BRFSS also extends to Guam, Puerto Rico, and the Virgin Islands. In some years one or more of these may be included in the median, minimum and maximum statistics, as noted in the table.

With its RDD sample, BRFSS was affected by the decline in landline telephone use and the increasing number of households with only cell phones. Prior to 2011, BRFSS sampled only landline phones. The median state response rate was 62.1 percent in 1997, but this rate has declined since that time, reaching 49 percent in recent years.

In 2011, BRFSS added a cell phone frame to its sample. Calls to cell phones generally achieve lower response rates than calls to landline phones. When the cell phone frame was introduced, the median state response rate for this mode was only 27.9 percent versus 53.0 percent for landline phones. However, the median cell phone rate has risen each year since 2011, reaching 40.5 percent in 2014, while the median landline response rate has dropped by nearly 4 percentage points. The median combined response rate dropped by 4.5 percentage points between 2011 and 2012 but has risen nearly 2 percentage points since 2012.[16] BRFSS departs from the other surveys in showing a rise in its combined response rate in the past three years and no decline in its landline response rate.

The maximum state landline response rate has declined much more substantially than the median rate. In 1997 the maximum state landline response rate was 88.9 percent, but the

[16] Within each state the combined response rate is calculated as a weighted sum of the landline and cell phone response rates. The weights are the sample proportions allotted to each mode. The median combined rate is the median of the individual state combined rates.

maximum rate has fallen to 61.6 percent since then. Minimum landline rates have fallen from 41.3 percent in 1997 to 26.7 percent in 2014.

## Table A.8. Response rates for the Behavioral Risk Factor Surveillance System, 1997 to 2014

| Survey Year | Landline Median | Landline Minimum | Landline Maximum | Cellular telephone Median | Cellular Telephone Minimum | Cellular Telephone Maximum | Combined Landline & Cellular Telephone Median | Combined Landline & Cellular Telephone Minimum | Combined Landline & Cellular Telephone Maximum |
|---|---|---|---|---|---|---|---|---|---|
| 1997 | 62.1 | 41.3 | 88.9 | -- | -- | -- | -- | -- | -- |
| 1998 | 59.1 | 32.5 | 76.7 | -- | -- | -- | -- | -- | -- |
| 1999 | 55.2 | 36.2 | 80.8 | -- | -- | -- | -- | -- | -- |
| 2000 | 48.9 | 28.8 | 71.8 | -- | -- | -- | -- | -- | -- |
| 2001 | 51.1 | 33.3 | 81.5 | -- | -- | -- | -- | -- | -- |
| 2002 | 58.3 | 42.2 | 82.6 | -- | -- | -- | -- | -- | -- |
| 2003 | 53.2 | 34.4 | 80.5 | -- | -- | -- | -- | -- | -- |
| 2004 | 52.7 | 32.2 | 66.6 | -- | -- | -- | -- | -- | -- |
| 2005 | 51.1 | 34.6 | 67.4 | -- | -- | -- | -- | -- | -- |
| 2006 | 51.4 | 35.1 | 66.0 | -- | -- | -- | -- | -- | -- |
| 2007 | 50.6 | 26.9 | 65.4 | -- | -- | -- | -- | -- | -- |
| 2008 | 53.3 | 35.8 | 65.9 | -- | -- | -- | -- | -- | -- |
| 2009 | 52.5 | 37.9 | 66.9 | -- | -- | -- | -- | -- | -- |
| 2010 | 54.6 | 39.1 | 68.8 | -- | -- | -- | -- | -- | -- |
| 2011 | 53.0 | 37.4 | 66.5 | 27.9 | 20.2 | 54.0 | 49.7 | 33.8 | 64.1 |
| 2012 | 49.1 | 28.2 | 62.7 | 35.3 | 16.4 | 55.7 | 45.2 | 27.7 | 60.4 |
| 2013 | 49.6 | 28.0 | 62.0 | 36.5 | 19.1 | 62.6 | 46.4 | 29.0 | 59.2 |
| 2014 | 48.7 | 26.7 | 61.6 | 40.5 | 22.2 | 60.0 | 47.0 | 25.1 | 60.1 |

Source: Adapted from multiple annual issues of the BRFSS Summary Data Quality Report, available by choosing individual years from http://www.cdc.gov/brfss/annual_data/annual_data.htm [January 2016].

Note: The estimates for 1997 through 2003 include Puerto Rico. The estimates for 2001 through 2003 also include Guam and the U.S. Virgin Islands.

**This page has been left blank for double-sided copying.**

**This page has been left blank for double-sided copying.**

www.mathematica-mpr.com

**Improving public well-being by conducting high quality, objective research and data collection**

PRINCETON, NJ ■ ANN ARBOR, MI ■ CAMBRIDGE, MA ■ CHICAGO, IL ■ OAKLAND, CA ■ WASHINGTON, DC

MATHEMATICA
Policy Research