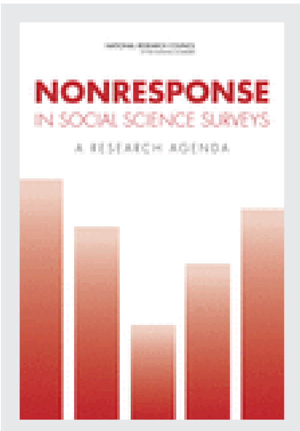

Nonresponse in Social Science Surveys: A Research Agenda

DETAILS

166 pages | 6 x 9 | PAPERBACK
ISBN 978-0-309-27247-6 | DOI 10.17226/18293

AUTHORS

Panel on a Research Agenda for the Future of Social Science Data Collection; Committee on National Statistics; Division on Behavioral and Social Sciences and Education; 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

NATIONAL RESEARCH COUNCIL
OF THE NATIONAL ACADEMIES

THE NATIONAL ACADEMIES PRESS
Washington, D.C.
www.nap.edu

THE NATIONAL ACADEMIES PRESS **500 Fifth Street, NW** **Washington, DC 20001**

NOTICE: The project that is the subject of this report was approved by the Governing Board of the National Research Council, whose members are drawn from the councils of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine. The members of the committee responsible for the report were chosen for their special competences and with regard for appropriate balance.

This study was supported by the Russell Sage Foundation (award number 97-10-04). Support for the Committee on National Statistics is provided by a consortium of federal agencies through a grant from the National Science Foundation (award number SES-1024012). Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the editors and do not necessarily reflect the views of the organizations or agencies that provided support for the project.

International Standard Book Number-13: 978-0-309-27247-6
International Standard Book Number-10: 0-309-27247-5

Additional copies of this report are available from the National Academies Press, 500 Fifth Street, NW, Keck 360, Washington, DC 20001; (800) 624-6242 or (202) 334-3313; http://www.nap.edu.

Printed in the United States of America

Suggested citation: National Research Council. (2013). *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.

THE NATIONAL ACADEMIES

Advisers to the Nation on Science, Engineering, and Medicine

The **National Academy of Sciences** is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. Upon the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Ralph J. Cicerone is president of the National Academy of Sciences.

The **National Academy of Engineering** was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. C. D. Mote, Jr., is president of the National Academy of Engineering.

The **Institute of Medicine** was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, upon its own initiative, to identify issues of medical care, research, and education. Dr. Harvey V. Fineberg is president of the Institute of Medicine.

The **National Research Council** was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both Academies and the Institute of Medicine. Dr. Ralph J. Cicerone and Dr. C. D. Mote, Jr., are chair and vice chair, respectively, of the National Research Council.

www.national-academies.org

PANEL ON A RESEARCH AGENDA FOR THE FUTURE OF SOCIAL SCIENCE DATA COLLECTION

Roger Tourangeau (*Chair*), Methodology Group, Westat, Rockville, MD
Nancy Bates, Research and Methodology Directorate, U.S. Census Bureau, Washington, DC
Suzanne M. Bianchi, Department of Sociology, University of California, Los Angeles
J. Michael Brick, Methodology Group, Westat, Rockville, MD
Douglas D. Heckathorn, Department of Sociology, Cornell University
Larry Hedges, Institute for Policy Research, Northwestern University
Arthur Kennickell, Board of Governors, Federal Reserve System, Washington, DC
Kristen Olson, Department of Sociology and Survey Research and Methodology Program, University of Nebraska–Lincoln
Nora Cate Schaeffer, Department of Sociology, University of Wisconsin–Madison
Frank Stafford, Economics Department and Population Studies Center, University of Michigan, Ann Arbor

Thomas J. Plewes, *Study Director*
Brian Harris-Kojetin, *Associate Study Director* (on detail from U.S. Office of Management and Budget)
Michael J. Siri, *Program Associate*

COMMITTEE ON NATIONAL STATISTICS
2012–2013

Contents

Preface

Nearly three decades have elapsed since the National Research Council (NRC) last convened a panel to undertake a comprehensive review of issues associated with nonresponse in sample surveys. The three-volume seminal study, *Incomplete Data in Sample Surveys* (National Research Council, 1983), reported the results of that early investigation. The 1983 panel focused mainly on statistical techniques that could illuminate and ameliorate the effects of nonresponse. Its study recommended a research agenda consisting of eleven far-reaching recommended programs, projects, and activities ranging from improvement of weighting methods to gathering and analyzing data on costs; these research recommendations are excerpted in Appendix B of this report. Many of these recommendations have been at least partially implemented.

Despite the significant improvements in general understanding of the causes and consequences of survey nonresponse and in methodology for compensating for the effects, the problems associated with the lack of response to surveys continue; in fact, nonresponse appears to be a growing issue. Response rates to government and privately sponsored household surveys that provide rich data for social science research have been falling throughout the richer countries of the world (see, e.g., De Leeuw and De Heer, 2002). To try to maintain response rates, sponsoring organizations have had to spend many more dollars in repeated efforts to contact sample units and address their concerns about participating. According to Curtin, Presser, and Singer (2005), the rapid decline in response rates has clearly increased survey costs (p. 97). Furthermore, this decline in response rates is challenging the underlying inferential assumption for estimation from sample surveys, which is that there is 100 percent response to a probability sample selected from a designated frame with nearly complete coverage of the target population.

These challenges threaten to undermine the validity of inferences obtained through the collection of information from subjects through surveys. Survey nonresponse affects validity in a number of ways. One way is through the introduction of bias into the survey results, but the issue of bias is quite complex. For example, a recent meta-analysis of 59 methodological studies (Groves and Peytcheva, 2008) concluded that large nonresponse biases can occur in surveys and, further, that nonresponse rates themselves are a poor predictor of the magnitude of the biases (p. 2). This study concluded that high response rates do not always reduce the risk of nonresponse bias. Various survey attributes, such as the method used to calculate bias, survey sponsorship, and the survey population, also play a role in determining bias (p. 25).

In early 2009, members of the board of the Russell Sage Foundation expressed concern to the Committee on National Statistics (CNSTAT) about the threats to statistical inference from the problems associated with declining response rates in traditional social science surveys and indicated their willingness to support a planning meeting that would help develop the plans for a useful project, such as a workshop, a series of workshops, or a full-scale panel study. The planning meeting was held in Washington, DC, on December 14, 2009. A distinguished roster of experts participated in the planning meeting, including experts in survey design; social scientists who use survey data; government, academic, and private-sector managers of surveys for research and policy analysis; and experts in alternative data sources and data collection methods.

Two papers were commissioned for the meeting, which summarized the research literature on what is known about the causes of survey nonresponse and the effects of the growing levels of nonresponse on inference. In addition, a panel session explored technologies and methods that could potentially mitigate nonresponse bias and other threats to the quality of data upon which social science relies. Such technologies and methods include mixed-mode surveys, the use of administrative records (e.g., retail scanner data, payroll data, or state tax and transfer program data) to replace some interviews or questions in a survey, automatic data capture methods (e.g., personal data assistants, global positioning system locators), and the use of geographic information systems to develop area-based sampling frames. The participants indicated the nature and scope of a project that could be of most value in addressing the problems in this area.

In concluding the planning meeting, the participants agreed that the first priority would be to develop a research agenda to capture information about causes, consequences, and remedies for nonresponse and to move forward the state of the science. As part of developing an agenda, it would be useful to identify short-term projects that would inform a larger, more comprehensive review of all ramifications of the problem and the solutions. This study derives from those outcomes of the planning meeting.

Statement of Task

A panel of experts under the National Research Council's (NRC's) Committee on National Statistics will conduct a study to develop a research agenda for addressing issues related to the deterioration in social science data stemming from the general decline in survey response by individuals and households. The panel will consider what is known about the causes and consequences of increasing nonresponse, the current state of survey methodology, and methods designed to improve response for surveys in the government, academic, and private sectors. The panel will identify high-priority research that can answer important unresolved questions about survey response and determine the most cost-effective ways to improve response and the quality of survey data for the advancement of knowledge in the social sciences. On the basis of its information-gathering activities, including a workshop, the panel will deliberate, make recommendations, and publish these recommendations along with supporting findings as an independent NRC report.

In November 2010, the Russell Sage Foundation commissioned the NRC's CNSTAT to assemble a panel of experts to develop a research agenda for addressing issues related to the impact on social science data of the general decline in survey response by individuals and households. In the statement of task (shown above), the panel was asked to consider what is known about the causes and consequences of increasing rates of non-response, the current state of survey methodology, and methods designed to improve response for surveys in the government, academic, and private sectors. The panel was asked to identify high-priority research that can answer important unresolved questions about survey response and determine the most cost-effective ways to improve response and the quality of survey data for the advancement of knowledge in the social sciences. For the most part, the panel has limited its purview to nonresponse in household surveys, both public and private, in keeping with the charge in the statement of task. Likewise, the report focuses largely on U.S. household surveys, although research and operational experience in several international surveys is discussed where it has a bearing on general nonresponse issues commonly confronted in the conduct of household surveys regardless of where they are done.

The panel engaged in wide-ranging information-gathering activities, including an extensive literature search. The literature review identified a number of recommendations for research on survey nonresponse topics, which are reproduced in Appendix B of this report. The panel also conducted two workshops to which experts in various aspects of nonresponse research were invited. The results of the literature review and the information gathered in the two workshops are summarized in Chapters 1 and 4 of the report, which focus on documenting response trends and identifying means of improving response, and in Chapters 2 and 3, which summarize the state of the science for understanding and adjusting for response bias.

Working with the information gathered from these activities, the panel deliberated in order to develop recommendations for a research agenda. These recommendations are presented in this report along with supporting findings and conclusions and are summarized in Chapter 5.

The panel especially and gratefully acknowledges the contributions of the many panel members and invited experts who participated in the two workshops and shared so freely of their knowledge. The findings of this report can be traced in large part to their input, although the guest experts bear no responsibility for the conclusions drawn by the panel.

In its first workshop on February 17–18, 2011, the panel focused on several topics that are basic to understanding nonresponse and its effects. Sessions featured reviews of the state of knowledge about the role of field operations in achieving high response rates, the current status of research on mode effects, evidence on effectiveness of incentives, research on post-survey adjustments for nonresponse, and new metrics for nonresponse. The presenters were asked to respond to questions about the state of the current knowledge on each topic.

In the first session, Cathy Haggerty and Nina Walker of NORC at the University of Chicago discussed recruiting, training, and managing field staff to achieve high-response levels, summarizing their extensive experience. A panel on mode effects featured presentations on the reports of the American Association for Public Opinion Research task forces on cell phone surveys by Paul Lavrakas, consultant, and online panels by Reg Baker of Market Strategies International. Rounding out that session was a presentation on self-administered modes by Mick Couper of the University of Michigan, Ann Arbor, and the Joint Program in Survey Methodology. Eleanor Singer of the University of Michigan, Ann Arbor, gave a presentation on what is known about incentives, and James Wagner, also of the University of Michigan, Ann Arbor, spoke on new metrics of survey nonresponse. The importance of collecting and analyzing paradata was discussed by Frauke Kreuter of the Joint Program in Survey Methodology, who described the state of the science on the use of paradata for post-survey adjustments. In the first of a series on federal statistical agency presentations, panel member Nancy Bates summarized the status and accomplishments of the U.S. Census Bureau research program on nonresponse. Panel member Mike Brick summarized the research and practice on using weighting to adjust for nonresponse.

Papers from this first workshop as well as from the planning meeting have been brought together in a volume of *The ANNALS of the American Academy of Political and Social Science,* "The Nonresponse Challenge to Surveys and Statistics," edited by Douglas S. Massey and Roger Tourangeau (Volume 645, January 2013). These papers contain an extensive literature review, which is not repeated in this report.

The second workshop, which took place on April 27–28, 2011, continued the review of ongoing research on nonresponse at federal agencies and took up several new topics, including international research on nonresponse; the state of knowledge on the role of interviewers in achieving high response rates; a discussion of models for survey costs; current issues and practices in mixed-mode survey research; and a discussion of issues of nonresponse in social network surveys and respondent-driven sampling methods.

The session on federal agency research on survey nonresponse featured John Dixon from the Bureau of Labor Statistics, Jaki McCarthy from the National Agricultural Statistics Service, Jennifer Madans from the National Center for Health Statistics, and Steven H. Cohen from the National Center for Science and Engineering Statistics. Two international guests, Ineke Stoop of the Netherlands Institute for Social Research and Lilli Japec of Statistics Sweden, discussed the status of international research and practice on survey nonresponse. The status of research on interviewer effects on nonresponse was summarized by panel member Nora Cate Schaeffer of the University of Wisconsin–Madison. Barbara O'Hare of the U.S. Census Bureau and François Laflamme from Statistics Canada led a session on survey costs, with the former discussing an interagency study coordinated by the Census Bureau and the latter summarizing important work in responsive design that is ongoing at Statistics Canada. Mixed-mode surveys were again a topic in this workshop and were discussed in a session featuring Don Dillman of Washington State University and Deborah Griffin of the U.S. Census Bureau. Douglas Heckathorn, a panel member from Cornell University, and Sandra Berry of RAND focused on nonresponse in the growing class of social network surveys.

Tom Plewes served as study director for the panel and ably supported its work. Michael Siri provided administrative support to the panel. The panel benefited greatly in the early phases of its work from the many contributions of Brian Harris-Kojetin who served as associate study director while on an Intergovernmental Personnel Act assignment on leave from the U.S. Office of Management and Budget. We are especially thankful for the personal participation of Constance F. Citro, director of CNSTAT, in the conduct of the workshops and in the preparation of this report. These people's hard work greatly benefited the report in numerous ways.

This report has been reviewed in draft form by individuals chosen for their diverse perspectives and technical expertise, in accordance with procedures approved by the Report Review Committee of the NRC. The purpose of this independent review is to provide candid and critical comments that assist the institution in making its reports as sound as possible and to ensure that the reports meet institutional standards for objectivity, evidence, and responsiveness to the study charge. The review comments and draft manuscript remain confidential to protect the integrity of the deliberative process.

The panel thanks the following individuals for their review of the report: Rachel A. Caspar, Center for Survey Methodology, RTI International; Frederick Conrad, Program in Survey Methodology, University of Michigan, Ann Arbor, and Joint Program in Survey Methodology, University of Maryland, College Park; John Dovidio, Department of Psychology, Yale University; Simon Jackman, Department of Political Science, Stanford University; Frauke Kreuter, Joint Program in Survey Methodology, University of Maryland, College Park; Tom W. Smith, Center for the Study of Politics and Society, NORC at the University of Chicago; and Kirk M. Wolter, Survey Research, NORC at the University of Chicago.

Although the reviewers listed above have provided many constructive comments and suggestions, they were not asked to endorse the conclusions or recommendations, nor did they see the final draft of the report before its release. The review of the report was overseen by Eleanor Singer, Survey Research Center, University of Michigan, Ann Arbor. Appointed by the NRC, she was responsible for making certain that the independent examination of this report was carried out in accordance with institutional procedures and that all review comments were carefully considered. Responsibility for the final content of the report rests entirely with the authoring panel and the NRC.

Roger Tourangeau, *Chair*
Panel on a Research Agenda for the Future of Social Science Data Collection

Summary

For many household surveys in the United States, response rates have been steadily declining for at least the past two decades. A similar decline in survey response can be observed in all wealthy countries. Efforts to raise response rates have used such strategies as monetary incentives or repeated attempts to contact sample members and obtain completed interviews, but these strategies increase the costs of surveys.

This review addresses the core issues regarding survey nonresponse. It considers why response rates are declining and what that means for the accuracy of survey results. These trends are of particular concern for the social science community, which is heavily invested in obtaining information from household surveys. The evidence to date makes it apparent that current trends in nonresponse, if not arrested, threaten to undermine the potential of household surveys to elicit information that assists in understanding social and economic issues. The trends also threaten to weaken the validity of inferences drawn from estimates based on those surveys. High nonresponse rates create the potential or risk for bias in estimates and affect survey design, data collection, estimation, and analysis.

The survey community is painfully aware of these trends and has responded aggressively to these threats. The interview modes employed by surveys in the public and private sectors have proliferated as new technologies and methods have emerged and matured. To the traditional trio of mail, telephone, and face-to-face surveys have been added interactive voice response (IVR), audio computer-assisted self-interviewing (ACASI), Web surveys, and a number of hybrid methods. Similarly, a growing research agenda has emerged in the past decade or so focused on seeking solutions to various aspects of the problem of survey nonresponse; the potential solutions that have been considered range from better training and deployment of interviewers to more use of incentives, better use of the information collected in the data collection, and increased use of auxiliary information from other sources in survey design and data collection. In addition, considerable effort has gone into developing weighting adjustments and adjustment models to compensate for the effects of nonresponse.

This report also documents the increased use of information collected in the survey process (paradata) in nonresponse adjustment. Some of this work is in early stages, while other work is more advanced. Two relatively new indicators of the nature and extent of nonresponse bias—representativity and balance indicators—may assist in directing focus on the core of the problem in ways that the traditional measures, such as overall nonresponse rates, cannot.

Several approaches to increasing survey response are being taken or have been proposed. Some of these approaches are aimed at increasing general knowledge about the conditions and motivations underlying response and nonresponse; others are focused on identifying techniques that change the interaction of interviewer and respondent or that could motivate respondents; still others employ paradata to identify possible survey design and management techniques that can be used to positively adjust the collection strategy to minimize the level or effects of nonresponse. As part of these efforts, survey researchers are enriching auxiliary information for both the reduction of nonresponse and adjustment for it, exploring matrix sampling ("planned missingness") and other strategies to reduce burden, exploring mixed-mode alternatives for data collection, and deploying responsive or adaptive designs.

The research agenda proposed in this report is needed to develop even better approaches to improve survey response and to improve our ability to use the data for analytical purposes even when response rates cannot be efficiently improved. The agenda should be multifaceted. In these times of increasingly constrained human and financial resources in the social science survey community, this agenda must be mindful of both costs and benefits.

Based on the panel's assessment of the state of knowledge about the problem of nonresponse in social surveys, the report suggests several key research areas in which the statistical community could fruitfully invest resources. Some of the recommended agenda items are designed to further advance our knowledge of the scope and extent of the problem, others to enhance our understanding of the relationship between response rates and bias, and still others to improve our ability to address the problems that come with declining response rates.

The recommendations for research include basic research that would help define the problem, develop appropriate measures, and expand our understanding of the scope and extent of the problem, such as:

- Research on people's general attitudes toward surveys and on whether these have changed over time.
- Research about why people take part in surveys and the factors that motivate them to participate.
- Research to identify the person-level and societal variables that have created the downward trend in response rates, taking into account changes in technology, communication patterns, and survey administration.

As a part of a research program that would illuminate why people take part in surveys, research is needed to clarify the factors that provide positive motivation (such as incentives) as well as those that provide pressure to participate. As specific examples:

- Research on the overall level of burden from survey requests and on the role that burden plays in an individual's decision whether to participate in a specific survey.
- Research on the different factors affecting contact and cooperation rates. In an era when more and more people are taking steps to limit their accessibility, research is needed on whether the distinction between contact and cooperation is still useful to maintain.

It is well-documented that the increase in nonresponse has led to increasing costs of conducting surveys. But cost measures are not standardized and are hard to come by. Research is needed on:

- The cost implications of nonresponse and how to capture cost data in a standardized way.

Likewise, it is important to periodically challenge the fundamentals that underlie our understanding of the statistical nature of nonresponse control and adjustment. This calls for a variety of research initiatives, including:

- Research on the theoretical limits of what nonresponse adjustments can achieve, given low correlations with survey variables, measurement errors, missing data, and other problems with the covariates.
- Research on and development of new indicators for the impact of nonresponse, including application of the alternative indicators to real surveys to determine how well the indicators work.

- Research on understanding mode effects, including ways in which mixed-mode designs affect both nonresponse and measurement errors and the impact of modes on reliability and validity.

The panel notes that there has been increasing appreciation of the role of nonresponse bias, but this only draws attention to the lack of a comprehensive theory of nonresponse bias. A more comprehensive theory would help further a basic understanding of the relationship between response rates and nonresponse bias, enhance the understanding of such bias, and aid in the development of adjustment techniques to deal with bias under differing circumstances. A unifying theory would assure that comparisons of nonresponse bias in different situations would lead to the development of standard nomenclatures and approaches to the problem. To assist in the development of such a theory, the report sugests a need for:

- Research on the relationship between nonresponse rates and nonresponse bias and on the variables that determine when such a relationship is likely.
- Research to examine both unit and item nonresponse bias and to develop models of the relationship between nonresponse rates and bias.
- Research on the impact of nonresponse reduction on other error sources, such as measurement error.
- Research to quantify the role that nonresponse error plays as an overall component of total survey error.
- Research on the differential effects of incentives offered to respondents (and interviewers) and the extent to which incentives affect nonresponse bias.

Finally, research that is needed to identify those plans, policies, and procedures that would assist in overcoming the problem:

- Research to establish, empirically, the cost–error trade-offs in the use of incentives and other tools to reduce nonresponse.
- Research on the nature (mode of contact, content) and the effects of the contacts that people receive over the course of a survey, based on data captured in the survey process.
- Research leading to the development of minimal standards for call records and similar data in order to improve the management of data collection, increase response rates, and reduce nonresponse errors.
- Research on the structure and content of interviewer training as well as on the value of continued coaching of interviewers. Where possible, support should be given to experiments designed to identify the most effective techniques.

- Research to improve the modeling of response as well as to improve methods to determine whether data are missing at random.
- Research on the use of auxiliary data for weighting adjustments, including whether weighting can make estimates worse (i.e., increase bias) and whether traditional weighting approaches inflate the variance of the estimates.
- Research to assist in understanding the impacts of adjustment procedures on estimates other than means, proportions, and totals.
- Research on how to best make a switch from the telephone survey mode (and frame) to mail, including how to ensure that the right person completes a mail survey.
- Research on the theory and practice of responsive design, including its effects on nonresponse bias, information requirements for its implementation, types of surveys for which it is most appropriate, and variance implications.
- Research on the availability, quality, and application of administrative records to augment (or replace) survey data collections.
- Research to determine the capability of information gathered by mining the Internet to augment (or replace) official survey statistics.

1

The Growing Problem of Nonresponse

The problem of nonresponse in social science surveys in the United States has been succinctly summarized by the sponsor of this study, the Russell Sage Foundation (2010), in a foreword to the foundation's description of this study:

> Household survey responses rates in the United States have been steadily declining for at least the last two decades. A similar decline in survey response can be observed in all wealthy countries, and is particularly high in areas with large numbers of single-parent households, families with young children, workers with long commutes, and high crime rates. Efforts to raise response rates have used monetary incentives or repetitive attempts to obtain completed interviews, but these strategies increase the costs of surveys and are often unsuccessful. Why is response declining and how does it increase the possibility of inaccurate survey results? Most important, with the advancement of reliable social science research at stake, what are effective methods for increasing the response rate for household surveys?

In examining the extent and effect of nonresponse in social science surveys and the steps that can be taken to reduce the effects of nonresponse, an overriding concern is that the validity of social science studies and their conclusions is indeed at stake. Apparent trends in nonresponse, if not arrested, are likely to weaken the validity of inferences drawn from estimates based on household surveys and undermine, perhaps fatally, their potential to elicit information that assists in understanding social and economic issues. There has always been incomplete acceptance that social science is *science*. Doubts about the validity of data from surveys give ammunition to critics of social science and other skeptics reluctant to accept the conclusions of social scientific investigations.

Moreover, these trends undermine researchers' (and the public's) confidence in the use of the survey as a tool in the development of public policy responses to social problems. This growing lack of confidence stands in stark contrast to the recent trends of very rapid growth in the use of surveys in social science and in the variety of methods used to conduct surveys. In documenting trends in government-sponsored social science survey research over a recent two-decade span, Presser and McCulloch (2011) found that the number of surveys increased by more than 50 percent over the period 1984 to 2004 and that the number of respondents approved for studies cleared by the Office of Management and Budget (OMB) increased by more than 300 percent. The authors point out that the relative increase in the number of surveys was many times greater than the relative increase in the size of the general population during the same time frame.

The modes employed by surveys in the public and private sectors have proliferated as new technologies and methodologies have emerged, but each of these new technologies and processes has its own set of methodological issues and its own nonresponse profile. To the traditional trio of mail, telephone, and face-to-face surveys, researchers have added interactive voice response (IVR), audio computer-assisted self-interviewing (ACASI), and Web surveys. In addition, new methods of sampling (such as address-based sampling) have emerged. The proliferation of modes of data collection has been accompanied by a growing public and private research agenda focused on seeking solutions to aspects of the problem of survey nonresponse, ranging from better training and deployment of interviewers, to increased use of incentives, to better use of information collected during the survey process, and to increased use of auxiliary information from other sources in survey design and estimation.

In this chapter, we set the stage for subsequent discussions of the issues surrounding survey nonresponse, its consequences, and proposed approaches to these issues by examining the nature and extent of nonresponse in social science surveys. We lay out evidence regarding the trends in nonresponse rates overall and by survey mode, and we consider evidence that increasing levels of nonresponse are an international phenomenon. We also explore some of the reasons for growing nonresponse rates. Finally, we discuss the current state of knowledge about some of the practical implications of growing nonresponse rates for survey costs and survey management. The evidence presented in this chapter leads to several recommendations for a research agenda aimed at filling some of the current gaps in our knowledge concerning trends in nonresponse as well as the reasons for those trends.

CONCEPTUALIZING AND DEFINING NONRESPONSE

Great strides have been made over the past three decades in conceptualizing and defining survey nonresponse, and this work has led to a growing consensus about the root causes and the definition and classification of types of survey nonresponse—a necessary first step in the development of metrics about nonresponse and in coming to grips with the consequences.

The decision of a person whether to respond to a survey involves several key factors, and an examination of those factors can be laid out as a conceptual framework for considering survey nonresponse issues. The elaboration of these factors provides a convenient conceptual point of departure for this review.

A main, overriding factor is a cost-benefit analysis: People respond to surveys when they conclude that the rewards outweigh the costs. Societal factors also enter in. If there is social disorganization or high crime rates, for example, some respondents will be more likely to refuse. There may be differences in the likelihood to respond among different sociodemographic groups, with some more or less prone to take part in surveys. The survey setting also may influence the decision to respond—which in most cases, particularly in a telephone environment, may be a near-instantaneous decision—and interviewers play a key role in determining cooperation, especially in face-to-face surveys. In the larger sense, a growing societal concern over intrusions on people's time and privacy may have an influence. Attitudes also play a part. The growth in the number of surveys has led to stronger attitudes about surveys in general, apart from attitudes about the topic. The salience of the topic plays a role, and participation may be partly an emotional decision. The mode of data collection may also affect response propensities. And finally, there may be a random component to the decision whether or not to cooperate. These factors are each explored in some depth in this report.[1]

The advances made toward arriving at a definitional consensus on nonresponse have been largely the result of ongoing work by the American Association for Public Opinion Research (AAPOR), which in the late 1990s began publication of its *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys*. This volume has been subsequently amended and updated, most recently in 2011.[2]

The survey research community has, under the leadership of AAPOR,

---

[1]The conceptual framework that is presented here was drawn largely from the survey methodology field, with the understanding that every field of social science has concepts and frameworks that might shed light on the problem of growing nonresponse, its causes, and its consequences.

[2]The AAPOR work in this field was preceded by a project to develop standard definitions by the Council of American Survey Research Organizations.

focused on four operational definitions of rates of survey participation: cooperation rates, contact rates, refusal rates, and response rates. These definitions differ in more than nuance. According to the definitions in the AAPOR *Standard Definitions* volume referred to above, cooperation rates are the proportion of all cases interviewed of all eligible units ever contacted; contact rates are the proportion of all cases in which some household member was actually reached; refusal rates are the proportion of all potentially eligible cases in which the housing unit or the sample member refuses to be interviewed or breaks off the interview;[3] and the most familiar rate, the response rate, represents the proportion of complete interviews, with responding units divided by the number of eligible units in the sample (American Association for Public Opinion Research, 2011). In view of the importance of response rate as a descriptor of the quality of a survey, AAPOR provides six related formulas for calculating response rates (see Box 1-1) and also defines different measures for different modes of survey collection.

There is now general concurrence with the overall AAPOR framework throughout the social science survey community. Indeed, several journals now require that one of the AAPOR response rates be reported when survey results are presented. Even with the rich variety of AAPOR definitions, however, some survey organizations have had to supplement these definitions with ancillary concepts and definitions based on their own needs and buttressed by their own research. Yet, by and large, the definitions have leveled the playing field and have held up well, except in one case: The distinction between contact and cooperation rates has been a source of confusion for some. Calculating household-level contact and cooperation rates requires obtaining information on contact or lack of contact with the household. The distinctions between them are often blurred, and in an era when more and more people are taking steps to limit their accessibility, research is needed on whether the distinction between contact and cooperation is still useful to maintain.

> **Recommendation 1-1: Research is needed on the different factors affecting contact and cooperation rates.**

The process of reaching a consensus definition of nonresponse rates is illustrated by the process within the federal government, which spon-

[3]Break-offs may occur at several points in the interview. As a rule of thumb, an interview is considered a break-off refusal when less than 50 percent of applicable questions are answered; those with 50–80 percent answered are considered partial responses; and those with 80 percent or more answered are considered completed responses (American Association for Public Opinion Research, 2011, p. 14).

BOX 1-1
What Is a Response Rate?

Response Rate 1 (RR1), or the minimum response rate, is the number of complete interviews divided by the number of interviews (complete plus partial) plus the number of non-interviews (refusal and break-off plus non-contacts plus others) plus all cases of unknown eligibility.

Response Rate 2 (RR2) counts partial interviews as respondents but otherwise is identical to RR1.

Response Rate 3 (RR3) estimates *e*, the proportion of cases with unknown eligibility that are actually eligible. In estimating *e*, researchers should be guided by the best available scientific information on what share eligible cases make up among the unknown cases and should not select a proportion in order to boost the response rate. The basis for the estimate must be explicitly stated and detailed.

Response Rate 4 (RR4) allocates cases of unknown eligibility as in RR3, but also includes partial interviews as respondents as in RR2.

Response Rate 5 (RR5) is either a special case of RR3 in that it assumes that $e = 0$ (i.e., that there are no eligible cases among the cases of unknown eligibility) or the rare case in which there are no cases of unknown eligibility.

Response Rate 6 (RR6) makes the same assumption as RR5 and also includes partial interviews as respondents. RR5 and RR6 are appropriate only when it is valid to assume that none of the unknown cases are eligible ones, or when there are no unknown cases. RR6 represents the maximum response rate.

SOURCE: Excerpted (without equations) from American Association for Public Opinion Research (2011, pp. 44–45).

sors and conducts some of the largest and most significant social science surveys in the United States. The federal government ventured into documenting and defining trends in unit nonresponse in reaction to studies conducted by the Federal Committee on Statistical Methodology (FCSM) in the 1990s (Johnson et al., 1994; Shettle et al., 1994). These studies found little consistency in how the federal surveys that they reviewed measured and reported nonresponse rates. These inconsistencies were found to have been due mainly to differences in sample design across the surveys. Still, as recently as 2000 it was observed that the official publication of many response rates from government surveys was “fragmented, sporadic, and non-standardized” (Bates et al., 2000). The lack of standardized reporting of response rates for federal surveys was again documented in FCSM Working Paper 31 (Office of Management and Budget, 2001). The comparability issues are exacerbated for international comparisons where refusals are handled differently.

OMB has subsequently issued policy guidance to standardize nonresponse rate definitions and to set targets for acceptable survey response rates—no easy task since government surveys vary widely in their mandate, sample design, content, interview period, mode, respondent rules, and periodicity. The main source of this guidance is the OMB Statistical and Science Policy Office release titled *Guidance on Agency Survey and Statistical Information Collections—Questions and Answers When Designing Surveys for Information Collections* (Office of Management and Budget, 2006). This document provides information on the OMB clearance process for surveys and other statistical information collections. Included is information about defining response rates. In summary, the document encourages agencies to use the AAPOR standard formulas in calculating and reporting response rates in clearance submissions to OMB, while permitting agencies to "use other formulas as long as the method used to calculate response rates is documented" (Office of Management and Budget, 2006, p. 57).

The existing formal definitions of nonresponse rates take the outcome of a case, response or nonresponse, as given. Although there is relatively little ambiguity about what constitutes response, nonresponse may cover a broad range of possibilities beyond "refusal" and "no contact." Some nonresponse may reflect refusals that are so adamant that conversion is a virtual impossibility, but in other cases there is a degree of judgment about the utility of following up, perhaps with another interviewer or a specialized refusal convertor in an interviewer-mediated survey. Respondents who are not contacted might not be reached because they are away for a prolonged period, because contact was poorly timed, or simply because contact is random and more attempts would have been needed. It is only relatively recently that it has become possible, with the use of electronic records to quantify attempts on individual cases, to make a broad exploration of effort and nonresponse. The availability of such electronic records opens the doors to new avenues of research, and additional work in this area is needed.

LONG-TERM TRENDS IN RESPONSE RATES

Evidence concerning trends in response rates in general population surveys has been accumulating for over three decades. In an early trend study, Steeh (1981) found that refusals were increasing in U.S. academic surveys conducted between 1952 and 1979. In 1989, Groves summarized the literature and concluded that "participation in social surveys by sample households appears to be declining in the United States over time." Later, in the early 1990s, the concern grew to a crescendo with new evidence offered by Bradburn (1992) that response rates were declining and had been declining for some time. It became apparent that this decline in response rates was becoming widespread. This concern was buttressed by the experience of the 1990 Census of Population and Housing, which saw an unexpected and significant decline in mailback responses (Fay et al., 1991). Brehm (1994) found that all sectors of the survey industry—academic, government, business, and media—were suffering from falling response rates. Groves and Couper (1998) likewise concluded that the decline in response rates was being experienced by government, academic, and commercial surveys. Their summary of data from other countries was based on surveys with long time series and found, for example, that while response to the Canadian Labor Force Survey appeared to be stable over the years, nonresponse appeared to be increasing in Holland, Japan, and Sweden. These repeated surveys are thought to be especially valuable indicators of time trends because their designs remain stable across years. Campanelli et al. (1997) also corroborated this trend, but they found that there were some surveys that had managed to maintain their response rates, although at the expense of a larger investment in time and money.

This direct evidence was accompanied by a growing literature exhibiting concern about the size and effects of nonresponse, evidenced by studies such as National Research Council (1983), Goyder (1987), Lessler and Kalsbeek (1992), Brehm (1993), Groves and Couper (1998), De Leuw and De Heer (2002), Groves et al. (2002), and Stoop (2005). In addition, the *Journal of Official Statistics* devoted special issues to nonresponse in 1999, 2006, and 2011. The most recent in this series evidenced the growing maturity in the field, emphasizing assessment of nonresponse bias and its impact on analysis (Blom and Kreuter, 2011).

In her introduction to the 2006 special edition of *Public Opinion Quarterly*, Singer (2006) cited multiple sources to support the consensus view that nonresponse rates in U.S. household surveys have increased over time. Although not discussed in this paper, household survey nonresponse is a matter of growing concern in many countries.[4]

[4]International comparative data on response rates and types of nonresponse in combination with data on design and collection strategies are scarce. While studies have shown that response rates differ across countries and over time, international data should be interpreted with caution. There are differences in survey design (such as mode of data collection and whether the survey is mandatory) and survey organization and implementation. Comparisons from the European Social Survey, a biennial face-to-face survey of attitudes, opinions, and beliefs in around 30 European countries, finds that, although the target response rate is 70 percent, in practice response rates are often lower and vary across countries (Stoop et al., 2010, p. 407). At our panel's workshop, Lilli Japec of Statistics Sweden reported that nonresponse

Based on the above evidence, we are able to conclude that response rates continue on a long-term downward path, but we are concerned that solid evidence about the *reasons* for the decline is still elusive.

Recommendation 1-2: Research is needed to identify the person-level and societal variables responsible for the downward trend in response rates. These variables could include changes in technology and communication patterns, which are also reflected in changes in methods of collecting survey data.

Recommendation 1-3: Research is needed on people's general attitudes toward surveys and on whether these have changed over time.

Recommendation 1-4: Research is needed about why people take part in surveys and the factors that motivate them to participate.

RESPONSE RATE TRENDS IN CROSS-SECTIONAL SURVEYS

This section reviews the recent experiences with response rates for several large and high-visibility repeated cross-sectional surveys. The surveys were selected to be illustrative of some of the more prominent surveys that produce information used in the social sciences. The discussion is abstracted from a paper prepared for this study by Brick and Williams (2013), who conducted an extensive literature review and examined four surveys from the late 1990s to the late 2000s. Two of them were face-to-face surveys—the National Health Interview Survey (NHIS) and the General Social Survey (GSS)—and two were list-assisted random digit dialing (RDD) telephone surveys—the National Household Education Survey (NHES) and the National Immunization Survey (NIS). We have added discussions of two additional telephone surveys: the Behavioral Risk Factor Surveillance System (BRFSS) and the Survey of Consumer Attitudes (SCA). We supplement the work of Brick and Williams with a discussion of results of analysis performed by the Census Bureau, the Bureau of Labor Statistics, and the Federal Reserve Board on response rates in the surveys they conduct.

in the Swedish Labor Force survey went from 2 percent to nearly 25 percent between 1963 and 2010. Among her findings were that response rates were higher when interviewers found the survey to be very interesting and among interviewers who had a positive attitude toward persuasion. At the sampled person level, there were lower response rates for non-Swedish citizens and persons 64 years of age and older and higher response rates for married persons.

National Health Interview Survey

The NHIS samples households as well as one adult and child within each sample household, so the rates shown in Table 1-1 reflect the initial household screening rate as well as the unconditional rate of response for the various survey target groups—the family and one sample adult and one child. The National Center for Health Statistics computes both conditional and unconditional response rates. We use the unconditional response rate (which is equivalent to RR1).

The NHIS underwent a major redesign, including changes in the survey questionnaire and moving from paper to computer-assisted personal interviewing (CAPI) administration in 1997. As a result, the response rate series begins with 1997, the first year of the new design. Nonresponse rose sharply at the beginning of this period but then fell slightly before rising again at the end of the period. Table 1-1 shows the unconditional response rates for the period 1997 to 2011.

General Social Survey

The GSS is a social survey that collects data on demographic characteristics and attitudes. It is conducted by NORC at the University of Chicago using face-to-face interviews of a randomly selected sample of non-

TABLE 1-1 National Health Interview Survey Unconditional Response Rates, 1997–2011 (in percentage)

Survey Year	Household Module	Family Module	Sample Child Module	Sample Adult Module
1997	91.8	90.3	84.1	80.4
1998	90.0	88.2	82.4	73.9
1999	87.6	86.1	78.2	69.6
2000	88.9	87.3	79.4	72.1
2001	88.9	87.6	80.6	73.8
2002	89.6	88.1	81.3	74.3
2003	89.2	87.9	81.1	74.2
2004	86.9	86.5	79.4	72.5
2005	86.5	86.1	77.5	69.0
2006	87.3	87.0	78.8	70.8
2007	87.1	86.6	76.5	67.8
2008	84.9	84.5	72.3	62.6
2009	82.2	81.6	73.4	65.4
2010	79.5	78.7	70.7	60.8
2011	82.0	81.3	74.6	66.3

SOURCE: National Center for Health Statistics (2012).

institutionalized adults (18 years of age and older). The survey was conducted annually every year from 1972 to 1994 except in 1979, 1981, and 1992, and since 1994 it has been conducted every other year. The 2010 national sample had 55,087 respondents.

Response rates for this survey are reported using the AAPOR RR5 definition. Smith (1995) reports that from 1975 to 1993, the GSS exemplified a survey that did not experience increased nonresponse. After peaking at 82.4 percent in 1993, however, the response rate declined for a number of years. Since 2000, the rate has held steady in the vicinity of 70 percent; it was 70.3 percent for the most recent collection in 2010. Although the rate has held steady, the reasons for the decline in response rates have varied by survey round. Refusals have risen most as the reason for nonresponse over this period (NORC, 2011).

National Household Education Survey

The NHES is a biannual set of surveys that, except for a screening survey, vary in content from year to year. These surveys are sponsored by the National Center for Education Statistics (NCES) and are conducted via CATI using RDD samples. The survey has changed significantly over the years, and completion and response rates varied in different administrations of NHES. These various versions of the NHES are described in a series of working papers and technical papers from NCES for the period 1999–2007 (Brick et al., 2000; Hagedorn et al., 2004, 2006, 2009; Nolin et al., 2004). In their study of response rate trends for the early years of the survey, Brick et al. (1997) looked at response rates, refusals, and cooperation rates from the four NHES surveys. Zuckerberg (2010) reported screener response rates for 1991–2007.

Table 1-2 presents the NHES screener survey completion rates for 1991 through 2007. Overall response rates for NHES screeners have fallen from 81 percent in 1991 to 52.5 percent in 2007. There was a decline of 11 percentage points in response rates overall from the 2005 cycle. NCES undertook a major redesign of the survey to boost the response rates, conducting an initial feasibility test in 2009. The test involved a two-phase mail-out/mail-back survey using an address-based sampling (ABS) frame instead of the CATI data collection and RDD sampling, which had been used for NHES from 1991 to 2007. A large-scale field test took place in 2011. The most recent cycle of NHES data collection took place in 2012. (See Chapter 4 for a discussion of the ABS field test.)

In their analysis of the response rates for the NHES screening and the conditional adult rate, Brick and Williams (2013, p. 41) concluded that NHES has experienced a much greater increase in nonresponse rates, with an annual increase of 2.3 percentage points per year over the period, than

TABLE 1-2 Number of Completed Interviews and Weighted Screener Response Rates in NHES, by Survey Year and Component, 1991–2007

Survey Year	Number of Completed Interviews	Screener Response Rate (%)
NHES:91	60,322	81.0
NHES:93	63,844	82.1
NHES:95	45,465	73.3
NHES:96	55,838	69.9
NHES:99	55,929	74.1
NHES:01	48,385	67.5
NHES:03	32,049	61.7
NHES:05	58,140	64.2
NHES:07	54,034	52.5

NOTE: NHES = National Household Education Survey.
SOURCE: Zuckerberg (2010).

either the GSS or the NHIS. Like the NHIS, the linear fit is not very descriptive of the pattern of nonresponse even though the R^2 is high ($R^2 = 0.80$). If the 2007 data point were excluded, then the annual increase in nonresponse would only be 1.7 percentage points (similar to that reported by Curtin et al. in 2005 for another RDD survey).

National Immunization Survey

The NIS is a national RDD survey of telephone households with children who are between 19 and 35 months old. It is sponsored by the Centers for Disease Control and Prevention. The first data collection effort took place in 1994, and the survey has been conducted on a quarterly basis since April of the same year. Each year, 10 million telephone calls are made to identify households with 35,000 age-eligible children (Ballard-LeFauve et al., 2003). Because the population of interest is children, it is necessary to screen households to find out whether they include children who are eligible for the survey. Information on a child's vaccination history is provided by a household respondent. Since 1995, data have been sought from the child's medical provider, as household respondents often have problems recalling a child's vaccination history. Zell et al. (1995) and Battaglia et al. (1997) pointed out that using data from two sources provides a more accurate estimate of vaccination coverage.

NIS response rates have fallen fairly consistently over a period of 16 years (see Table 1-3). Brick and Williams (2013, p. 41) noted that the in-

TABLE 1-3 National Immunization Survey, Response Rates by Survey Year, 1995–2010[a]

Survey Year	Resolution Rate (%)	Screening Completion Rate (%)	Interview Completion Rate (%)	CASRO Response Rate (%)	Children with Adequate Provider Data (%)
1995	96.5	96.4	93.5	87.1	50.6
1996	94.3	96.8	94.0	85.8	63.4
1997	92.1	97.9	93.8	84.6	69.7
1998	90.4	97.8	93.6	82.7	67.1
1999	88.6	97.0	93.4	80.2	65.4
2000	88.1	96.0	93.1	78.7	67.4
2001	86.8	96.2	91.1	76.1	70.4
2002	84.8	96.6	90.6	74.2	67.6
2003	83.6	94.0	88.7	69.8	68.9
2004	83.8	94.8	92.0	73.1	71.0
2005	83.3	92.8	84.2	65.1	63.6
2006	83.3	90.5	85.6	64.5	70.4
2007	82.9	90.2	86.8	64.9	68.6
2008	82.3	90.3	85.1	63.2	71.0
2009	82.9	92.4	83.2	63.8	68.7
2010	83.3	91.5	83.6	63.8	71.2

NOTE: CASRO = Council of American Survey Research Organizations.
[a]Excludes the U.S. Virgin Islands.
SOURCE: Centers for Disease Control and Prevention, National Center for Immunization and Respiratory Diseases, and National Center for Health Statistics (2011, Table H.1).

crease in nonresponse is much more linear for the NIS than for the other surveys; a temporary uptick in response in 2004 has been the only departure of NIS from a very linear pattern. Battaglia et al. (2008) suggested that the uptick was likely due to interventions using incentives in refusal conversion around this time. The average annual increase in nonresponse over the period was 2.1 percentage points ($R^2 = 0.96$). This rate of increase is comparable to that seen in the NHES, although the NIS level is lower than the level of the adult interview in the NHES.

There are several measures used by NIS to assess patterns of nonresponse for the survey. The resolution rate is the percentage of the total telephone numbers selected that can be classified as non-working, nonresidential, or residential. The screening completion rate is the percentage of known households that are successfully screened for the presence of age-eligible children. The interview completion rate is the percentage of households with one or more age-eligible children that complete the household interview. The Council of American Survey Research Organizations (CASRO) response rate equals the product of the resolution rate, the screening completion rate, and the interview completion rate among eligible households.[5] This rate declined from 87.1 percent in 1995 to 63.8 percent in 2010.

This decline in response rates is due to the fact that, from 1994 to 2010, the NIS used an RDD list-assisted landline phone sample frame. Blumberg et al. (2012) report that landline phone use decreased over time while cell phone use increased. In 2011, the NIS sampling frame was expanded from sampling landline phones to sampling landline and cell phones, creating a dual-frame sample. Differences between the dual-frame and landline data were documented. Using the dual-frame sampling, including cell phone–only survey participants, assures that the survey base reflects the U.S. population.

Behavioral Risk Factor Surveillance System

The BRFSS collects data on health-risk behaviors, clinical preventive health practices, and health-care access. It is the largest health survey conducted by telephone in the world.[6] Like other surveys, BRFSS is faced with the challenge presented by the trade-off between lower quality data and higher costs. Mokdad (2009) pointed out that use of telephone-based RDD methods to collect public health data had reached a crossroads. He documented the various experiments done by BRFSS to make improvements to the system, as recommended by a BRFSS expert panel. BRFSS conducted studies on the impact of advance letters, answering machine messages, and the Do Not Call registry on response rates; in addition, studies examined the use of real-time survey interpreters, mail and Internet collection, and address-based sampling. Advance letters proved to be more effective than leaving messages on answering machines. BRFSS's assessment of the impact of the Do Not Call registry on participation results showed that state-level responses were not affected. Using real-time survey interpreters improved the interpretation of questions and the understanding of responses provided by survey respondents. Results from Internet and mail survey experiments indicated that a combination of modes (self-administered data collection with telephone follow-up) improved response rates. BRFSS looked at the quality of address-based sampling by using a combination of RDD and mail surveys. The mail survey approach had the added advantage of including households that did not have landlines, which led to better responses in

[5]The description of resolution rate, screener completion rate, interview completion rate, and CASRO response rate is taken from Centers for Disease Control and Prevention, National Center for Immunization and Respiratory Diseases, and National Center for Health Statistics (2011, p. 11).

[6]The information is taken from the BRFSS Methodology Brochure. Available: http://www.cdc.gov/brfss/pdf/BRFSS_Methodology_Brochure.pdf [November 2011].

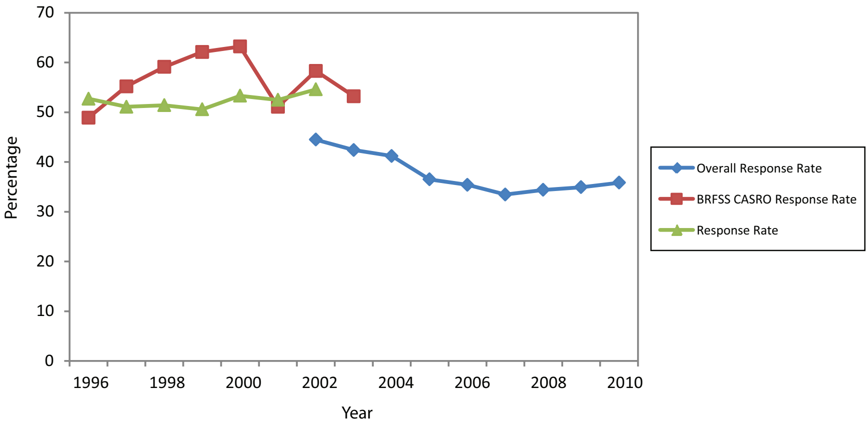

FIGURE 1-1 BRFSS median response rates.
NOTES: BRFSS = Behavioral Risk Factor Surveillance System, CASRO = Council of American Survey Research Organizations. Median rates calculated with respect to state-level results, excluding Guam, Puerto Rico, and the U.S. Virgin Islands.
SOURCE: Adapted from multiple annual issues of the Summary Data Quality Report, available by choosing individual years from http://www.cdc.gov/brfss/annual_data/annual_data.htm [May 2013].

low-response BRFSS states and therefore provided a good alternative to RDD.

BRFSS reports a response rate, a CASRO response rate, and an overall response rate (as depicted in Figure 1-1). The response rate is calculated by dividing the number of complete and partial interviews (the numerator) by an estimate of the number of eligible units in the sample (the denominator). The CASRO response rate assumes that the unresolved numbers contain the same percentage of eligible households as those records whose eligibility or ineligibility is determined. The overall response rate is more conservative in that it assumes that more unknown records are eligible and so includes a higher proportion of all numbers in the denominator (Centers for Disease Control and Prevention, 2011).

Survey of Consumer Attitudes

The SCA is a monthly telephone survey, with 500 telephone interviews of men and women living in U.S. households completed each month. The Survey Research Center at the University of Michigan uses RDD to draw the national sample for SCA. It uses a rotating panel design. The monthly

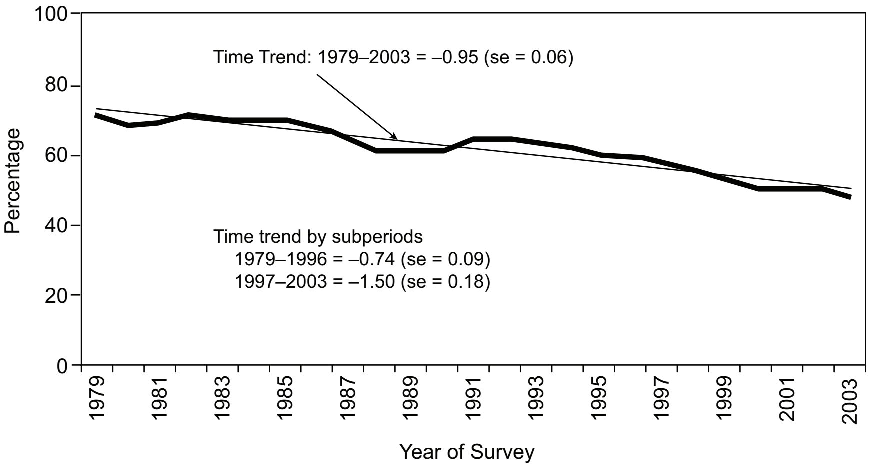

FIGURE 1-2 Changes in telephone survey response in the Survey of Consumer Attitudes.
SOURCE: Figure 1 in Curtin et al. (2005). Reprinted with permission. © Oxford University Press 2005.

samples include 300 fresh cases and 200 cases that were interviewed six months earlier.

Curtin et al. (2005) analyzed the response rates in SCA from 1979 to 2003. The response rate dropped from 76 percent in 1979 to 60 percent in 2003, an average decline of three-quarters of a percentage point per year. The rate of decline in SCA response rates is not smooth over this period, as seen in Figure 1-2, which is adapted from Figure 1 in Curtin et al. (2005).[7] There was a gradual decline from 1979 to 1989, followed by a plateau from 1989 to 1996 and then a steep decline from 1996 to 2003.

The authors reported opposite results for final refusals to SCA and non-contacts. Both of these factors contributed to the decline in response rates, with non-contacts driving the decline from 1979 to 1996, while the rise in refusals led to the steep decline in response rates from 1996 to 2003. The authors did not find evidence that lower unemployment or the increased use of call-screening devices in the 1990s contributed to the rise in refusals and non-contacts. They conjectured that the rapid growth in sales phone calls

[7]The response rate shown here uses every sampled phone number with the exception of those known to be ineligible in the denominator and uses partial interviews in the numerator (AAPOR response rate RR4) (Curtin et al., 2005).

and survey phone calls in the period of analysis might have contributed to the phenomenon of response rate decline in the SCA.

Census Bureau Household Surveys

The Census Bureau conducts several of the largest and most important household surveys in the federal government under the sponsorship of other federal agencies. The response rate trends that are discussed in this section pertain to the Consumer Expenditure Diary (CED) Survey, the Consumer Expenditure Quarterly (CEQ) Interview Survey, the Current Population Survey (CPS), the National Crime Victimization Survey (NCVS), the National Health Interview Survey (NHIS), and the Survey of Income and Program Participation (SIPP) during the period 1990–2008. Data for two other major surveys, the American Housing Survey–Metropolitan Sample (AHS–MS) and the American Housing Survey–National Sample (AHS–NS), are not available for the full period.

Table 1-4 shows the initial interview nonresponse rates for the years 1990 and 2009. Initial nonresponse rates are those for the first interview in multi-interview surveys. Initial nonresponse rates for all of these surveys increased, with the CED, CEQ, and CPS nonresponse rates almost doubling and the NCVS, NHIS, and SIPP nonresponse rates more than doubling. The refusal rates showed similar patterns of increase, although this category accounted for about the same or a smaller amount of overall nonresponse than it did 18 years earlier. The proportional decline in refusals has been offset by steady increases in the proportion not at home (Landman, 2009).

The nonresponse rates for the two housing surveys, which are conducted on a different time schedule than the surveys presented in Table 1-4, likewise increased. The nonresponse rates for the AHS–MS went up from 8.3 percent in 1998 to 12.7 percent in 2007, and the nonresponse rate for the AHS–NS increased from 7.8 percent to 9.6 percent over the same time span.

Bureau of Labor Statistics Household Surveys

In his presentation to the panel, John Dixon of the Bureau of Labor Statistics (BLS) compared trends in response rates for several household surveys that are sponsored by BLS; these are shown in Figure 1-3. Most of the surveys had response rates that were relatively stable over the past decade, such as the Consumer Price Index Housing Survey, which had a budget-related problem at the end of 2009 but has recovered. A notable exception is the Telephone Point of Purchase Survey (TPOPS), a commodity and services purchasing behavior survey, which has seen its response rates decline precipitously and has the worst response rates of the BLS

TABLE 1-4 Initial Nonresponse and Initial Refusal Rates in Selected Census Bureau Surveys, 1990 and 2009 (in percentage)

Survey	Outcome Category	1990	2009
Consumer Expenditure Diary	Nonresponse	16.3	29.7
	Refusal	8.4	10.8
Consumer Expenditure Quarterly	Nonresponse	12.0	23.6
	Refusal	9.5	14.3
Current Population Survey	Nonresponse	5.7	9.5
	Refusal	2.2	4.1
National Crime Victimization Survey	Nonresponse	4.3	9.5
	Refusal	n/a	4.9
National Health Interview Survey	Nonresponse	4.5	17.8
	Refusal	2.7	10.8
Survey of Income and Program Participation	Nonresponse	7.3	19.2
	Refusal	5.3	12.9

NOTE: n/a = not available.
SOURCE: Landman (2009).

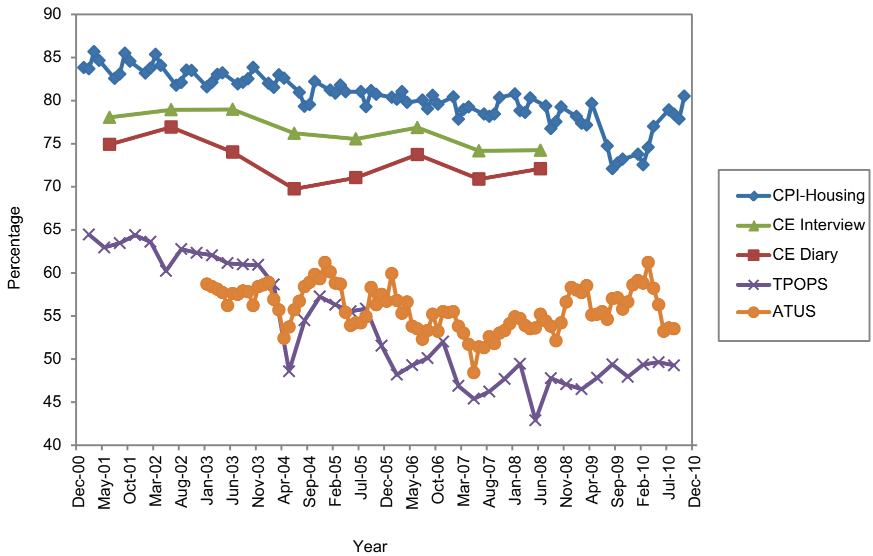

FIGURE 1-3 Response rate trends for major Bureau of Labor Statistics household surveys, December 2000–December 2010.
NOTE: ATUS = American Time Use Survey, CE Diary = Consumer Expenditure Diary, CE Interview = Consumer Expenditure Interview, CPI-Housing = Consumer Price Index-Housing, TPOPS = Telephone Point of Purchase Survey.
SOURCE: Based on figure in John Dixon presentation at panel's workshop (Dixon, 2011).

surveys (less than 50 percent recently). TPOPS, which is conducted as an RDD survey, had a dramatic decline in its response rate in 2008, at which point its rate was less than 45 percent, but it has stabilized since then in the range of 50 percent. The American Time Use Survey (ATUS)—a telephone survey of a specific Current Population Survey household member, conducted several months after the household has completed its last CPS interview—has shown relatively low but stable rates, generally in the range of 50 to 60 percent.

Survey of Consumer Finances

The Survey of Consumer Finances (SCF) is an example of a national survey that has managed to avoid a decline in response rates over the period in which other surveys have experienced declines (Kennickell, 2007, 2010). The SCF provides data to support the analysis of the financial behavior of U.S. households and their use of financial services. It is a complex survey, collecting detailed information on a wide variety of assets and liabilities as well as data on current and past employment, pensions, income, demographic characteristics, and attitudes, with a survey design that includes collection by telephone (about 47 percent of cases are completed by phone) and personal visits. The sample is selected from two sample frames, an area-probability sample and a list sample that covers wealthy households. The overall initial sample of approximately 10,000 cases is divided approximately evenly between the two subsamples. In 2007, the area-probability sample had a response rate of 67.8 percent, while the list sample had a rate of 34.7 percent, with substantial variation in rates across the list sample strata (Kennickell, 2010). The area-probability sample rates have recovered after a period of decline in the early 1990s (see Table 1-5).

Response Rate Trends by Survey Type

The experience of this illustrative set of surveys provides evidence that nonresponse rates continue to increase in all types of cross-sectional surveys, with little to suggest that the trend has plateaued. However, the data also clearly show that the recent rates of increase in nonresponse have been substantially greater for RDD telephone surveys than for face-to-face surveys. The absolute percentage point increase for the RDD surveys is roughly three times that of the face-to-face surveys.

RESPONSE RATE TRENDS IN PANEL SURVEYS

Response rates are often seen as measures of the quality of both cross-sectional and panel surveys. Although panel studies follow a sample of

TABLE 1-5 Response Rates for the Survey of Consumer Finances, Area-Probability Sample, 1992–2010 (in percentage)

	Year						
Sample	1992	1995	1998	2001	2004	2007	2010
Region							
Northeast	65.4	60.1	62.4	68.7	61.5	67.5	65.8
North Central	68.5	70.9	67.4	66.9	69.3	70.3	72.1
South	70.3	67.2	68.3	70.7	72.5	68.2	71.3
West	66.4	65.3	63.8	64.9	68.2	65.0	63.4
Area Type							
Self-represent PSU	61.8	58.9	62.3	63.2	65.8	62.2	65.3
Other MSA	67.4	66.6	66.6	69.7	70.4	74.4	74.4
Non-MSA	75.7	77.6	70.3	73.3	71.3	70.8	69.6
All Areas	68.0	66.3	65.9	68.1	68.7	67.8	68.7

NOTE: MSA = metropolitan statistical area, PSU = primary sampling unit.
SOURCE: Communication with Arthur Kennickell, U.S. Federal Reserve Board.

individuals over a long period of time and thus tend to suffer from attrition as well as initial nonresponse, the dropout rates from panel surveys, after the initial baseline entry interviews, tend to be smaller than the nonresponse rates in cross-sectional surveys. Of course, nonresponse rates in the baseline interview of panel surveys are often comparable to those in cross-sectional surveys. In this section, we discuss the response rate experience of three major panel surveys: the National Longitudinal Surveys of Youth (NLSY), the Health and Retirement Study (HRS), and the Panel Study of Income Dynamics (PSID). We use the term "retention" rates to refer to the response rates in later rounds of each survey, conditional on completing a baseline interview.

The research associated with these surveys has suggested several reasons for the generally low nonresponse rates in the later rounds of panel surveys.[8] One factor leading to lower nonresponse rates may be the respondent's familiarity with the interview format and interviewer. In a repeated survey, the respondent is aware of the kind of questions likely to be asked, the length of the interview, and the amount of effort required, because he or she has been through a similar questionnaire in the previous wave. Whatever commitment the respondent may feel to the survey may be reinforced if the same interviewer is assigned to that household

[8]Additional insight is contained in Lynn (2009), which reviewed evidence from previous research that modeled the response process within a multivariate framework and developed estimates of predictions of response from the Household, Income, and Labour Dynamics in Australia Survey.

or individual for several years. On the other hand, there is some evidence that increased prior interviewing contacts breed a familiarity with the interviewer that can depress reporting of some behaviors. Mensch and Kandel (1988) speculated that interviewer familiarity "increases salience of normative standards" and that responses are conditioned by that familiarity and also by the respondents' expectations of a future encounter with the interviewer (p. 100).

Another possibility, drawing on sociological research on blood donation by Piliavin and Callero (1991), is that people may develop an identity as a respondent to a particular survey over time. Although they may respond for a variety of reasons initially, the more they see their participation in identity terms, the more likely they will continue in the future.

In spite of high retention rates, as dropouts accumulate, panel surveys face the risk of attrition bias in many ways. Attrition in the sample over time can reduce the sample's representativeness and introduce bias; in addition, the loss of sample sizes over time increases the variance of the estimates. Sampling weights and refreshing the sample frame can ameliorate the effects of attrition bias, but the effectiveness of these approaches is still subject to question (Cellini et al., 2008).

There are many different response rates that can be calculated for panel surveys, and each has its separate purpose (Cheshire et al., 2011). For example, response rates can be evaluated at an aggregate (unit) level, at a wave-specific level, and as item-specific rates (Cho et al., 2004; Callegaro and DiSogra, 2008).

The National Longitudinal Surveys

The National Longitudinal Surveys are five surveys that cover different cohorts of men and women, each with the objective of collecting information on individual experiences with various life events, including employment (Bureau of Labor Statistics, 2001a, 2001b, 2002, 2006). We focus on the two most recent cohorts, which began in 1979 with a sample of people ages 14 to 22, and in 1997 with a sample of people ages 12 to 17.

The National Longitudinal Survey of Youth 1979 (NLSY79) interviews were initially conducted every year, but, starting in 1995, the survey has been conducted every other year. The interviews were conducted face to face from 1979 to 1986 and again from 1988 to 2000. In 1987, budget considerations dictated that most of the interviews be done by telephone. Telephone again became a major mode of data collection in the 2002 interview. A switch from paper-and-pencil interviewing (PAPI) to computer-assisted personal interviewing (CAPI) took place in 1993. In 2004, Web survey instruments were used for the first time in addition to the telephone interviews. The NLSY97 interviews were initially conducted annually but are now on a biannual schedule.

Table 1-6 presents the mode of data collection and completion rates for each round of the NLSY79 survey from 1979 to 2002. In spite of changes in interview modes, completion rates (100 minus the rate shown in the last column) for NLSY79 respondents considered eligible for an interview and interviewed were in the 80–90 percent range for most of the years from 1979 to 1996, although they slipped below that level in subsequent years. The cumulative retention rate across all rounds was around 80.9 percent for living respondents. Table 1-7 documents the decreasing response and

TABLE 1-6 Type of Survey and Completion Rate for NLSY79 by Survey Year

	Completed Interviews				
Year	Personal	Telephone	Mode Not Available	Not Interviewed	Completion Rate (%)
1979	11,863	548	275	—	—
1980	11,493	648	0	545	95.7
1981	11,541	654	0	491	96.1
1982	11,066	1,054	3	563	95.6
1983	11,897	324	0	465	96.3
1984	11,422	646	1	617	95.1
1985	9,941	953	0	713	93.9
1986	9,726	929	0	952	91.8
1987	1,126	8,998	362	1,122	90.3
1988	9,494	920	51	1,142	90.2
1989	PAPI: 8,832	1,469	3	1,002	91.4
	CAPI: 252	49			
1990	PAPI: 6,972	1,032	2	1,171	89.9
	CAPI: 2,145	285			
1991	7,773	1,241	4	946	90.5
1992	7,848	1,164	4	948	90.5
1993	7,917	1,081	13	953	90.4
1994	7,948	933	10	1,073	89.2
1996	7,594	1,042	0	1,328	86.7
1998	6,330	2,069	0	1,565	84.3
2000	5,420	2,613	0	1,931	80.6
2002	2,317	5,407	0	2,240	77.5

NOTES: PAPI interviews are those conducted with paper survey instruments and pencil-entered responses; CAPI interviews are administered using a laptop computer and an electronic questionnaire that captures respondent-, interviewer-, and machine-generated data. CAPI = computer-assisted personal interviewing; NLSY79 = National Longitudinal Survey of Youth 1979; PAPI = paper-and-pencil interviewing.
SOURCE: Adapted from Table 3.3 in Bureau of Labor Statistics (2005).

TABLE 1-7 Type of Survey and Retention Rates by Round for NLSY97

Round	Personal Interviews (%)	Telephone Interviews (%)	Total Interviewed	Retention Rate (%)
2	94.5	5.5	8,386	93.3
3	92.0	8.0	8,208	91.4
4	91.2	8.7	8,080	89.9
5	91.5	8.4	7,882	87.7
6	83.8	16.2	7,896	87.9
7	88.0	12.0	7,754	86.3
8	87.7	12.3	7,502	83.5
9	86.5	13.5	7,338	81.7
10	88.2	11.8	7,559	84.1
11	87.4	12.6	7,418	82.6
12	85.7	14.3	7,490	83.4
13	85.9	14.1	7,559	84.1
14	88.9	11.0	7,479	83.2

NOTES: Retention rate is defined as the percentage of all base-year respondents participating in a given survey. Deceased respondents are included in the calculations. NLSY97 = National Longitudinal Survey of Youth 1997.
SOURCE: See Table 1 at http://www.nlsinfo.org/content/cohorts/nlsy97/intro-to-the-sample/interviewmethods [May 2013].

retention rates for the rounds of the NLSY97. As might be expected, retention slipped a bit from year to year.

Health and Retirement Study

The HRS is a longitudinal database of a nationally representative sample of more than 20,000 individuals over the age of 50. It is a biennial survey. The baseline data collection took place in 1992, and the sample included individuals born between 1931 and 1941 and their spouses, irrespective of the spouses' ages. The Wave 1 response rate in 1992 was 81.6 percent. A second sample was generated to create an HRS auxiliary study known as Asset and Health Dynamics Among the Oldest Old (AHEAD). It consisted of respondents age 70 and above and their spouses. The HRS and AHEAD studies were merged in 1998. Two more cohorts—the War Baby (WB) cohort and the Children of the Depression Age (CODA) cohort—were added in the same year. The WB cohort sample consists of individuals born between 1942 and 1947, and the CODA cohort sample consists of those born between 1924 and 1930. In 2004, a new cohort was introduced, which is known as the Early Baby Boomer (EBB) cohort. The EBB sample incorporates individuals born between 1948 and 1953. The baseline response rates for AHEAD, WB, CODA, and EBB were 80.4 percent, 70 percent, 72.5 percent,

TABLE 1-8 Overall Retention Rates for Each Sample at Each Wave After the First (in percentage)

	Year(s) of Data Collection							
Sample	1994	1995/96	1998	2000	2002	2004	2006	2008
HRS	89.4	86.9	86.7	85.4	86.6	86.4	88.6	88.6
AHEAD		93.0	91.4	90.5	90.1	89.4	90.6	90.7
CODA				92.3	91.2	90.1	91.4	90.4
WB				90.9	90.6	87.9	88.1	87.0
EBB							87.7	86.3
Total (by year)	89.4	89.2	88.3	88.0	88.4	87.6	88.9	88.4

NOTE: AHEAD = Asset and Health Dynamics Among the Oldest Old Survey, CODA = Children of the Depression Age Sample, EBB = Early Baby Boomer Survey, HRS = Health and Retirement Study, WB = War Baby Survey.
SOURCE: National Institute on Aging (2011b).

and 68.7 percent, respectively.[9] Table 1-8 reports the retention rates from subsequent waves of the five samples. The baseline (first wave) response rate is based on completed interviews of all individuals deemed eligible for HRS. The wave-specific response rate is based on the number of individuals from whom an interview was sought in that wave. The declines in response rates over time that are apparent in cross-sectional surveys are not apparent here.

Most of the interviews in these five samples are conducted by telephone. Exceptions are made if the respondent has health issues or if the household does not have a telephone. Face-to-face interviews were conducted for HRS, WB, and CODA respondents in each sample's first wave and for AHEAD respondents age 80 years and above. Mode experiments were done on AHEAD respondents in the second and third wave and on HRS respondents in the fourth wave. The mode experiment in the second wave of AHEAD compared face-to-face and telephone interviews. A similar experiment was done on HRS respondents in the birth cohort of 1918–1920. These experiments assessed the impact of interview mode on measures of cognitive functioning. The average cognitive scores of telephone-mode respondents and face-to-face mode respondents did not differ significantly in the AHEAD- and HRS-mode experiments (Ofstedal et al., 2005). In a mail-out experiment conducted on the 1998 wave of the HRS, an advance letter was sent, followed by the mail questionnaire and a $20 incentive

[9]Descriptions of HRS, AHEAD, WB, CODA, and EBB are taken from National Institute on Aging (2011b).

check; the response rate achieved was 84 percent after excluding the exit cases (National Institute on Aging, 2011b).

Panel Study of Income Dynamics

The PSID is a longitudinal household survey directed by the University of Michigan Survey Research Center. The first round of data collection took place in 1968; the sample consisted of 18,000 individuals living in 5,000 families. From 1968 to 1997, data were collected every year. In 1999 PSID started collecting data biennially. The mode of data collection changed from in-person interview to telephone interview in 1973 to reduce costs. Exceptions are made for households with no telephone or in other circumstances that do not permit a respondent to give a telephone interview. The average in-person interview was around one hour long. Despite attempts at streamlining the survey over time, the cumulative response rate dropped from 76 percent in the baseline year, 1968, to 56 percent in 1988 (Hill, 1999). Further changes were made in 1993 when PSID moved from using paper-and-pencil telephone interviews to CATI. CATI has been used for PSID ever since.

Table 1-9 shows response rates for different segments of the PSID sample for the 2003 and 2005 waves.[10] A response rate of 97 percent is consistently achieved for the core segment and 88 percent for the immigrant segment. McGonagle and Schoeni (2006) documented the factors they believe are responsible for the high response rates in the PSID. These include incentive payments, payments for updating locating information during non-survey years, the use of experienced interviewers, persuasion letters sent to reluctant respondents, informing respondents about upcoming data collection waves, sending thank-you letters to responding households, and training interviewers to handle refusals. Even though response rates are exemplary in the case of the PSID relative to other panel surveys, it does suffer from high cumulative nonresponse rates due to an attrition rate of 50 percent (Cellini et al., 2008).

REASONS FOR NONRESPONSE

The reasons given by nonrespondents for not taking part in surveys have not changed much over time. A comparison of nonresponse reasons from two U.S. surveys—the 1978 National Medical Care Expenditure Survey (NMCES) and the 2008 NHIS—shows that the reasons reported

[10]The response rates are shown for two different family types—those that have remained stable (non-split-offs) and those in which family members have split off from the basic family unit to form a new family unit between waves.

TABLE 1-9 PSID Response Rates and Sample Sizes, 2003–2005

	Non-Split-Offs		Split-Offs		
Sample	Response Rate (%)	Number of Interviews	Response Rate (%)	Number of Interviews	Total Number of Interviews
2003 Actual					
Core reinterview	97	6,554	83	561	7,115
Core recontact	65	200	56	15	215
Core subtotal	95	6,754	82	576	7,330
Immigrant reinterview	94	459	61	36	495
Immigrant recontact	51	42	43	3	45
Immigrant subtotal	88	501	59	39	540
Total	95	7,255	80	615	7,870
2005 Actual					
Core reinterview	98	6,756	89	516	7,272
Core recontact	61	194	60	3	197
Core subtotal	96	6,950	88	519	7,469
Immigrant reinterview	93	500	74	45	545
Immigrant recontact	42	27	—	0	27
Immigrant subtotal	88	527	74	45	572
Total	95	7,477	87	564	8,041

SOURCE: McGonagle and Schoeni (2006). Reprinted with permission.

TABLE 1-10 Reasons for Nonresponse from Two Face-to-Face Surveys, 1978 and 2008, in Priority Order

NMCES Reasons for Nonresponse (1978)	NHIS Reasons for Nonresponse (2008)
Not interested	Not interested/does not want to be bothered
Unspecified refusal	Too busy
No time to give	Interview takes too much time
Poor physical health/mental condition of respondent	Breaks appointments
Antipathy to surveys in general	Survey is voluntary
Wants to protect own privacy	Privacy concerns
Third-party influences respondent to refuse	Anti-government concerns
Generalized hostility to government	Does not understand survey/asks questions about survey
Other reasons	Survey content does not apply
Objects to government invasion of privacy	Hang-up/slams door
	Hostile or threatens interviewer
	Other household members tell respondent not to participate
	Talk only to specific household member
	Family issues

NOTE: NHIS = National Health Interview Survey, NMCES = National Medical Care Expenditure Survey.
SOURCE: Brick and Williams (2013, p. 39).

and their relative order were similar in the two surveys, indicating that the increase in nonresponse rates over time cannot simply be attributed to a change in subjects' reasons for not responding (see Table 1-10).[11] This suggests in turn that there is no simple way of identifying mechanisms for the increase in nonresponse rates over time by examining the reasons given for nonresponse (Brick and Williams, 2013).

When the authors examined the relationship between survey nonresponse rates and nine selected characteristics that might be expected to influence response rates by affecting accessibility or cooperation, they were able to identify four variables that were highly correlated with nonresponse rates for the four surveys they studied: the percentage of families with children under the age of six; the percentage of single-person households; the violent crime rate; and travel time to work. However, it is unclear whether these trends just happened to coincide with the increase in nonresponse rates or whether they represent causal factors in that rise.

[11]The data on the NMCES are from Meyers and Oliver (1978) based on non-interview report forms completed by field interviewers and those for the NHIS are from Bates et al. (2008) utilizing automated contact history records of verbal and non-verbal interactions recorded by interviewers during contact with households.

Groves and Couper (1998) and Harris-Kojetin and Tucker (1999) conducted similar time-series analyses. The latter study found that outside influences also played a role. The authors found evidence that changes in presidential approval, consumer sentiments regarding the economy, and the unemployment rate were reliably related to changes in refusal rates in the Current Population Survey.

THEORETICAL PERSPECTIVES ON NONRESPONSE

Survey methodologists have proposed several theories to explain why people participate in surveys. According to Goyder et al. (2006), the development of a theory of survey response has centered on several key themes: People respond to surveys when they conclude that the rewards outweigh the costs; societal factors, such as social disorganization or high crime rates, cause some respondents to refuse; different sociodemographic groups are more or less prone to take part in surveys; factors in the survey setting may influence a near-instantaneous decision; there is growing concern over intrusions on people's time and privacy; the growth in the number of surveys has led to stronger attitudes about surveys in general, apart from topic; interviewers play a key role in gaining cooperation, especially in face-to-face surveys; the topic salience plays a role; participation is partly an emotional decision; response propensities can vary by mode of data collection; and there is a random component to these decisions. These ideas are captured to one degree or another in three main theories—social capital theory, leverage–saliency theory, and social exchange theory—which are summarized below.

Social Capital Theory

Social capital theory may help to explain the social and psychological underpinnings of the interpersonal relationships that promote trust and cooperation and thus promote the willingness to respond to surveys. According to Robert Putnam, who popularized this theory in his work (1995, 2001), social capital refers to the trust that people gain through productive interaction and that leads to cooperation. The cooperation is manifested in community networks, civic engagement, local civic identity, reciprocity, and trust in the community. Social capital can be measured in the prevalence of community organizations. The decline in association memberships in recent years is associated with less confidence in public institutions.

On an individual level, social capital is gained through education, which provides greater networking opportunities (Heyneman, 2000). Lower socioeconomic status is associated with reduced levels of trust and cooperative behavior (Letki, 2006). Other individual-level attributes affect the build-up of social capital; in particular, length of residence, home ownership, and the presence of spouses and children strengthen social capital.

In considering the effect that social capital has on the likelihood to respond to surveys, Brick and Williams (2013) point out two aspects of the theory that should be considered: Social capital is a collective rather than an individual attribute, and the loss in social capital is partly due to generational change (p. 54). To the extent that nonresponse is generational, for example, it would suggest "exploring predictive models for reasons of nonresponse that are time-lagged and smooth data, rather than relying on the simple models that are considered here" (Brick and Williams, 2013, p. 55). Although social capital theory remains a possible explanation for increasing nonresponse, considerable research must be done to establish a link between the theory and survey nonresponse. Brick and Williams state that they would find "a rigorous investigation of the relationship between social capital and nonresponse rates to be extremely helpful" (Brick and Williams, 2013, p. 55).

Leverage–Saliency Theory

The leverage–saliency theory (LST) provides a second perspective on survey participation. Groves et al. (2004) introduced the LST of survey participation in 2000. As summarized in Maynard et al. (2010), the LST is a theory of how potential respondents make decisions to participate in a survey interview.

The LST posits that people vary in the importance and value they assign to different aspects of a survey request (Groves et al., 2000). For example, for some individuals the topic may be important, for others the reputability of the organization conducting the survey may be significant, and for still others a chance to receive a cash reward may be of consequence. According to the theory, the influence of each component of the request depends both on the weight accorded it by a sampled individual (leverage) and on its prominence in the request protocol (saliency). One application of the LST has shown that when the survey topic is a factor in the decision to participate, noncooperation will cause nonresponse error on topic-related questions (Groves et al., 2004).

The LST assumes that a potential respondent has an expected utility for participating in a survey and agrees to participate if this expected utility is net positive. "Leverage" refers to a potential respondent's assessments (including valence and weight) of the survey's attributes that make participation more or less appealing. For example, a cash incentive might have a positive valence and a greater weight as the size of the incentive increases; a long interview might have negative valence and a weight that increases with the length of the interview (Dijkstra and Smit, 2002). Whether an attribute has a positive or a negative leverage varies across sample persons. A specific survey topic may have positive leverage for individuals who are more generally interested in talking about that topic and a negative leverage for those who are not (Groves et al., 2000). Likewise, different respondent groups may react differently. For example, because of different communication experiences people of different ages may be differentially sensitive to the length of a survey or to factors such as direct personal contact. There also may be strong ethnic and racial influences on a person's assessment of the survey attributes, perhaps related to distrust of mainstream institutions. To the extent that distrust of mainstream institutions is intensifying, rising nonresponse rates may be the result.

"Saliency"—or salience—refers to the prominence of different attributes of the survey request for a sample person who is deciding whether to participate. LST calls attention to the fact that sample persons may base their decisions on only a few attributes of the survey and also to how survey organizations and interviewers provide information to sample persons. A survey might provide a financial incentive and attempt to appeal to a sense of civic duty, for example, but an interviewer might emphasize only one of these aspects, making it more salient and potentially more influential as sample persons decide whether to participate. Consequently, requests for participation in a given survey might obtain different responses from the same person, depending on which attributes are made most salient. Box 1-2 indicates how topic saliency can improve response rates.

BOX 1-2
How Topic Saliency Improves Response Rates

Judging from behavior on the first contact containing the survey introduction, we found that persons cooperated at higher rates to surveys on topics of likely interest to them. The odds of cooperating are roughly 40 percent higher for topics of likely interest than for other topics, based on the four frames utilized in the experiment. Given these results and the deductions from leverage–salience theory, we suspect we could make the 40 percent much higher by making the topic a much more salient aspect of the survey introduction. It is important to note that the overall effects on total response rates of these effects are dampened by non-contact nonresponse, as well as by physical-, mental-, and language-caused nonresponse (Groves et al., 2004, pp. 25–26).

Social Exchange Theory

Social exchange theory posits that the decision to respond to a survey is based on the assessment of the costs versus the benefits of taking part. Exchanges occur in purest form in economic transactions, and little social relationship is required for successful economic exchanges. Social exchange theory attempts to cover non-economic exchanges, where the costs and benefits include such intangibles as the maintenance of tradition, conformity to group norms, and self-esteem.

The impetus for an exchange may be as simple as "liking" the interviewer by dint of personal interaction or prepaid incentives or as complex as establishing trust. One of the founders of social exchange theory, George Homans presents exchange as overlapping with psychological heuristics. Liking may even lead to a change in opinion in accordance with the liked person's views (Homans, 1958, p. 602).

Trust may be an important component in social exchanges. As Dillman (1978, 1991, 1999) has emphasized, in the survey setting it is important to minimize the costs to the respondents, clearly convey the nature and extent of the benefits, and establish trust so that the respondents are confident about the costs they can expect to incur and the benefits they can expect to receive. As demands on the time of respondents grow, increasing the cost of responding, and distrust of institutions multiplies, the social exchange theory would posit that nonresponse would grow unless perceived benefits of responding also increase.

IDENTIFYING COSTS ASSOCIATED WITH APPROACHES TO MINIMIZE NONRESPONSE

Costs and response trends have gone hand in hand as key driving factors in the selection of social science survey modes and in the design and conduct of surveys. As the cost of face-to-face data collection grew over the years, survey researchers moved to landline RDD data collection. As the coverage of landline RDD survey designs eroded, survey researchers shifted to dual-frame RDD designs or ABS designs. As the cost of dual-frame and ABS designs rose—and driven in part by the lack of timeliness of the latter—survey researchers were drawn to Web survey designs. It is possible that a significant portion of the downward trend in response rates is attributable to (a) survey budgets not keeping pace with rising costs even with (b) the increasing use of new frames and modes of interview to combat declining coverage. Combating declining response rates generally increases the cost of the survey, or, as stated quite succinctly in a 2004 study, "[S]urvey costs and survey errors are reflections of each other; increasing one reduces the other" (Groves, 2004, p. 49). The costs are incurred directly in increased interviewer time and also indirectly in the increased management costs associated with approaches such as responsive design, which requires greater resources for developing and using paradata and other information systems. (Responsive design is discussed in Chapter 4.)

Peytchev (2009) examines costs in the SCA, a telephone survey where a large part of the cost is interviewer time. The number of call attempts to complete an interview in the SCA doubled between 1979 and 1996, from 3.9 to 7.9 (Curtin et al., 2005); even so, the resulting non-contact rate more than tripled during this same period. Higher proportions of non-contacts and refusals also undoubtedly require more visits and more refusal conversion attempts in face-to-face surveys.

Unfortunately, there is a dearth of evidence that relates cost to nonresponse. In an attempt to rectify this knowledge gap, in conjunction with survey sponsors, the U.S. Census Bureau has launched a major effort to identify costs associated with survey activities.[12] Statistics Canada is exploring how responsive design initiatives can be employed to yield cost information that is useful for analysis and decision making.[13] Both of these initiatives are in their developmental stages, but they do constitute a serious attempt on the part of these agencies to understand the costs associated with response levels, and both are aided by the collection of data during the normal survey process.

Recommendation 1-5: Research is needed to establish, empirically, the cost–error trade-offs in the use of incentives and other tools to reduce nonresponse.

It is increasingly recognized that current trends in survey costs are unsustainable. The fact that, across the Census Bureau, data collection costs have been rising as survey participation has declined led to the cost savings identification effort. Census Bureau task forces were charged with identifying the most promising opportunities to improve the cost efficiency of data collection procedures in surveys that the Census Bureau conducts for other agencies under reimbursable agreements.

Three broad themes emerged from the task forces: (1) a need for better information on costs; (2) a need for less complex survey management; and (3) a need for continuous and cooperative cost management, especially during the data collection period. In regard to the first theme, it was pointed out that, to fully understand cost trends in the field, more detailed informa-

[12]Barbara O'Hare of the U.S. Census Bureau summarized this work in her presentation to the panel on April 28, 2011.

[13]François Laflamme of Statistics Canada summarized this work in his presentation to the panel on April 28, 2011.

tion about how field interviewers spend their time is needed. The Census Bureau does not ask interviewers to break out their survey hours, although other survey organizations capture travel and administrative time separately from actual data collection.

One method for monitoring costs is through information about the survey process via the collection of paradata (Couper, 1998). The Census Bureau has developed the contact history instrument (CHI), which captures the characteristics of household contact attempts and outcomes; other organizations have their own call record instruments. The information collected is used by field supervisors across organizations to manage cases. Call records, such as the CHI, provide valuable information on the level of effort by case and consequently can be a good proxy for cost. There is a need to go beyond call records alone and to consolidate and systematically monitor other paradata related to costs, such as daily work hours, travel, and case dispositions. When merged with information about nonresponse rates, missing data rates, and key indicator values, costs and data quality can be evaluated simultaneously, and data collection efforts may be managed more effectively.

Despite these pioneering efforts, there is no common framework for assessing the relationship between cost and response rates and no quantitative model of the relationship. A common framework would enable the development of common metrics that would be valid across modes so that comparisons could be made and so that information on costs in an increasingly mixed-mode environment can be generated. The literature suggests that some of the cost elements and metrics that would apply to interview surveys would include information on the nature and number of contacts that people receive and on production units, such as days of interviewer project-specific training, numbers of call attempts, the length of the field period, the number of interviewers, and remuneration and incentives—all of which can be captured in paradata collecting and generating systems (see, for example, American Association for Public Opinion Research, 2010a).

Recommendation 1-6: Research is needed on the nature (mode of contact, content) of the contacts that people receive over the course of a survey based on data captured in the survey process.

In the panel's discussion with field staff of a major ongoing survey, it became obvious that more significant costs are being incurred in the quest for high response rates.[14] More expensive senior interviewers are often assigned for difficult refusal cases. Much time is spent in gaining entry to

[14]Cathy Haggerty and Nina Walker of NORC at the University of Chicago summarized their experiences in a presentation to the panel on February 17, 2011.

gated and otherwise restricted communities in the hopes of gaining an interview with a sampled unit.

Recommendation 1-7: Research is needed on the cost implications of nonresponse and on how to capture cost data in a standardized way.

These "cost" data may be based on proxy variables (such as total interviewer hours per case or other measures of effort), but at a minimum it would be useful if these or similar aggregate cost measures were routinely published. Cost information is needed if researchers are to make informed decisions about whether additional efforts to combat nonresponse are worth the cost.

2

Nonresponse Bias

The trends toward declining survey response rates that are documented in Chapter 1 have consequences. One key consequence is that high nonresponse rates undermine the rationale for inference in probability-based surveys, which is that the respondents constitute a random selection from the target population. Most important, nonresponse creates the *potential* for bias in estimates, in turn affecting survey design, data collection, estimation, and analysis. We discuss the issue of nonresponse bias in this chapter as well as the relation of nonresponse bias to nonresponse rates. We also present evidence of the impact of nonresponse on the variance of the estimates and discuss the effectiveness of conventional adjustment tools to tackle nonresponse bias. Finally, we document the need to come to a better understanding of the causes and consequences of nonresponse bias.

RESPONSE RATES MATTER, BUT . . .

In a paper initially prepared for the planning meeting for this review, Peytchev (2013) observes that drawing "unbiased inference from probability-based surveys relies on the collection of data from all sample members—in other words, a response rate of 100 percent" (p. 89). If surveys are able to register a 100 percent response rate, there is no need for adjustments in developing estimates. However, when there is nonresponse, the probability of the respondent's inclusion is determined by both the initial probability of selection and the probability of responding.

The beauty of the response rate, most often stated as a single value for the entire survey, is that it reflects "the degree to which this goal of preserving the respondent's inclusion probabilities is achieved" (p. 89). Peytchev contends that this fact has inarguably contributed to the widespread interpretation of the response rate as a summary measure of a survey's representativeness. However, response rates can be misleading as measures of survey representativeness. The fact that response rates have fallen (as documented in Chapter 1) means only that the potential for nonresponse bias has increased, not necessarily that nonresponse bias has become more of a problem. That is because nonresponse bias is a function of both the nonresponse rate and the difference between respondents and nonrespondents on the statistic of interest, so high nonresponse rates could yield low nonresponse errors if the difference between respondents and nonrespondents is quite small or, in survey methodology terms, if nonresponse in the survey is ignorable and the data can be used to make valid inferences about the target population.

Moreover, it would be a relatively simple matter to overcome the problem of bias in the estimates brought about by nonresponse if there were a linear relationship between response rates and nonresponse bias across surveys. If it were so, one could theoretically reduce nonresponse bias by taking actions to increase response rates, and more effort, cost, training, and management control of the survey operation would solve the problem. This is not the case, however, as shown by Curtin et al. (2000), Groves et al. (2006), and Groves and Peytcheva (2008). The 2008 Groves and Peytcheva compilation of the results of 59 specialized studies found very little correlation between nonresponse rate and their measures of bias. Likewise, there is no proof that efforts to enhance response rates within the context of a survey will automatically reduce nonresponse bias on survey estimates (Curtin et al., 2000; Keeter et al., 2000; Merkle and Edelman, 2002; Groves, 2006).

Recommendation 2-1: Research is needed on the relationship between nonresponse rates and nonresponse bias and on the variables that determine when such a relationship is likely.

It is possible that extraordinary efforts to secure responses from a reluctant population may even increase bias on some survey estimates (Merkle et al., 1998). A 2010 study by Fricker and Tourangeau suggested that efforts to increase response can lead to quality degradation. The authors examined nonresponse and data quality in two national household surveys: the Current Population Survey (CPS) and the American Time Use Survey (ATUS). Response propensity models were developed for each survey. Data quality was measured through such indirect indicators of response error as item nonresponse rates, rounded value reports, and interview–reinterview response inconsistencies. When there was evidence of covariation between response propensity and the data quality indicators, potential common causal factors were examined. The researchers found that, in general, data quality, at least as they measured it, decreased for some variables as the probability of nonresponse increased. The study concluded that efforts to reduce nonresponse can lead to poorer quality data (Fricker and Tourangeau, 2010). Other work in this field is under way and may shed additional light on this important issue.

Recommendation 2-2: Research is needed on the impact of nonresponse reduction on other error sources, such as measurement error.

Recommendation 2-3: The research agenda should seek to quantify the role that nonresponse error plays as a component of total survey error.

EFFECTS OF NONRESPONSE BIAS

Although the exact relationship between nonresponse and bias is not yet clear, it is still important to understand the effects of nonresponse bias because bias jeopardizes the accuracy of estimates derived from surveys and thus the ability of researchers to draw inferences about a general population from the sample. The interactions are complex because nonresponse exists at the item level as well as at the interview level, and item nonresponse contributes to bias at the item-statistic level so that bias is a function of both unit and item nonresponse.

Recommendation 2-4: Research is needed to test both unit and item nonresponse bias and to develop models of the relationship between rates and bias.

Examples cited in Peytchev (2009) and other sources (e.g., Groves, 2006; Groves and Peytcheva, 2008) tend to show that the effect of nonresponse bias on means and proportions can be substantial:

- Nonrespondents to the component of the National Health and Nutrition Examination Survey III that measured glucose intolerance and diabetes through eye photography were 59 percent more likely to report being in poor or fair health than respondents to the main survey (Khare et al., 1994).
- The Belgian National Health Interview Survey with a response rate of 61.4 percent obtained a 19 percent lower estimate for reporting poor health than did the Belgian census, which had a response rate of 96.5 percent (Lorant et al., 2007).

- A comparison of the results of a five-day survey employing the Pew Research Center's usual methodology (with a 25 percent response rate) with results from a more rigorous survey conducted over a much longer field period and achieving a higher response rate of 50 percent found that, in 77 out of 84 comparisons, the two surveys yielded results that were statistically indistinguishable. Among the seven items that manifested significant differences between the two surveys, the differences in proportions of people giving a particular answer ranged from 4 percentage points to 8 percentage points (Keeter et al., 2006).
- An analysis of data from the ATUS—the sample for which is drawn from the CPS respondents—together with data from the CPS Volunteering Supplement was undertaken to demonstrate the effects of survey nonresponse on estimates of volunteering activity and its correlates (Abraham et al., 2009). The authors found that estimates of volunteering in the United States varied greatly from survey to survey and did not show the decline over time common to other measures of social capital. They ascribed this anomaly to social processes that determine survey participation, finding that people who do volunteer work respond to surveys at higher rates than those who do not do volunteer work. As a result, surveys with lower response rates will usually have higher proportions of volunteers. The result of the decline in response rates over time likely has been an increasing overrepresentation of volunteers. Furthermore, the difference shows up within demographic and other subgroups, so conventional statistical adjustments for nonresponse cannot correct the resulting bias.
- In a study of nonresponse bias for the 2005 National Household Education Survey (NHES), Roth et al. (2006) found evidence of a potential bias in the survey on adult education, in which females were more likely to respond than males were. The problem was resolved by a weighting class adjustment that used sex in forming the weighting classes.
- In another study of nonresponse bias in the 2007 NHES (Van de Kerckhove et al., 2009), results indicated undercoverage-related biases for some of the estimates from the school readiness survey. However, nonresponse bias was not a significant problem in the NHES data after weighting.

Recommendation 2-5: Research is needed on the theoretical limits of what nonresponse adjustments can achieve, given low correlations with survey variables, measurement errors, missing data, and other problems with the covariates.

Nonresponse can also affect the *variance* of any statistic, reducing confidence in univariate statistics, and it can also bias estimates of bivariate and multivariate *associations*, which could bias results from substantive analyses. These effects can be significant. The factors that could affect the variances are the underlying differences between respondents and nonrespondents in levels of variability, differences in respondent and nonrespondents means, uncertainty introduced by adjustments, and the variability in the number of interviewed respondents. All things considered, it is not clear whether nonresponse will generally cause variability to be understated or overstated.

The effect of nonresponse bias on *associations* is not clear, and, in Peytchev's (2009) view, it has been understudied. Peytchev observed that, to the extent that it has been studied, nonresponse bias in associations appears to be different from bias in means and proportions in that, even when there is substantial bias in means, there may be no nonresponse bias in bivariate associations.

There is a strong relationship between survey costs and nonresponse. On one hand, survey managers have intensified data collection activities in order to improve response rates. On the other hand, the trend toward higher survey costs has led to shortcuts, shortened collection periods, and cheaper modes that have had effects on survey response rates. The costs have been monetary and have also appeared in terms of data utility. Addressing nonresponse through a two-phase sample, for example, can increase the design effect in weighted estimates, producing a lower effective sample size than a comparable single-phase design. The problem of nonresponse can lead to survey designs of increasing complexity because accounting for nonresponse is becoming essential for the measurement and reduction of nonresponse bias. This is evident in the elaborate and dynamic responsive designs (Groves and Heeringa, 2006) that are now being implemented, in which design decisions are informed by deliberate variations in early recruitment protocols and careful monitoring of cost and error indicators during data collection. Some designs attempting to reduce nonresponse and keep costs under control are also incorporating multiple modes, as sample members exhibit different response propensities for different modes, and modes vary in cost structure. Some modes require separate sampling frames. Since combinations of sampling frames and modes yield higher response rates, the use of multiple sampling frames is also increasingly being considered. These designs are further discussed in Chapter 4.

Another consequence of growing nonresponse and the cost associated with improving response rates is an increase in the reliance of probability-based survey estimates on models that use auxiliary data. Some external data already exist on individuals in the population, but these data may be underutilized in surveys. External data can be found in administrative databases, in databases compiled by commercial vendors, and in the public domain. Some pioneering efforts to increase the use of auxiliary data for reducing the effect of nonresponse are discussed in Chapter 4. Likewise, the need to better understand the sources of nonresponse has led to an explosion in efforts to collect and analyze paradata—that is, data that document the survey process (call records, timing of call attempts, interviewer–respondent interactions, interviewer performance measures, and the like). Such measures can be collected by interviewers or directly by survey systems. Current challenges associated with paradata are the identification of what data elements to collect and how to organize such data structures in order to aid data collection and the creation of post-survey adjustments.

One use of paradata has been to gain a better understanding of the characteristics of "converted" respondents (i.e., those who were persuaded to take part after refusing initially), other respondents who were difficult to include in the survey, partial completers, and those who break off their survey responses versus those who complete interviews. The purpose is to test the assumption that reluctant sample members and those who do not complete the interview are more similar to nonrespondents than to respondents (Lin and Schaeffer, 1995; Bose, 2001). The advances in understanding and improving response rates through use of paradata are further discussed in Chapter 4.

NONRESPONSE BIAS IN PANEL SURVEYS

The problem of nonresponse bias in longitudinal panel surveys can be framed differently from the problem of nonresponse bias in non-repeated surveys. Longitudinal surveys suffer from attrition as well as initial nonresponse. Attrition is a form of nonresponse—that is, losing sample members from one collection wave to subsequent waves can produce nonresponse biases, just as in cross-sectional surveys.

Although still a substantial challenge, the characteristics of panel surveys are more amenable to understanding (and perhaps adjusting) for nonresponse. While bias in the first wave of data collection may persist in future rounds of data collection (Bose, 2001) and be exacerbated by attrition in later rounds, after the first wave of data collection, much is known about the characteristics, survey experiences, and response patterns of those who fail to respond to subsequent rounds. In a longitudinal study, once data have been collected in the base year from respondents, nonrespondents in subsequent rounds can be compared to respondents in those rounds using the base-year data as well as the frame data. Using this information, it is possible to understand, and adjust for, nonresponse bias. Still, understanding the amount and impact of nonresponse bias from survey attrition and from baseline nonresponse remains a challenge.

ANALYZING NONRESPONSE BIAS

As nonresponse has grown, so has interest in developing methods to estimate nonresponse bias and to adjust for it (Billiet et al., 2007; Stoop et al., 2010). For the surveys sponsored by the U.S. government, this interest has been driven, in part, by Office of Management and Budget requirements that "sponsoring agencies conduct *nonresponse bias analyses* when unit or item response rates or other factors suggest the potential for bias to occur" (Office of Management and Budget, 2006, p. 8, italics added). The thresholds that trigger a nonresponse bias analysis are an expected unit response rate of less than 80 percent or an item response rate of less than 70 percent.

The purpose of the analysis is to ascertain whether or not the data are missing completely at random. There are several approaches for determining this:

- Multivariate modeling of response using respondent and nonrespondent frame variables to determine if nonresponse bias exists.
- Weighting the final sample using population figures for background variables (i.e., using post-stratification) and comparing the weighted results with unweighted results to identify those variables that might be particularly prone to bias.
- Comparison of the characteristics of the respondents to known characteristics of the population in order to provide an indication of possible bias. However, a Federal Committee on Statistical Methodology study concluded that these comparisons should be used with caution because some of the differences between the respondent and population data may be due to measurement differences (including differences in coverage and content), true changes over time, or errors in the external sources (Office of Management and Budget, 2001). Estimates from large multipurpose surveys such as the American Community Survey and the Current Population Survey are often used in these comparisons, as are administrative record data sources.
- Using interviewer observation data, such as ratings of dwellings and neighborhood information as observed and recorded by the interviewers. These data are used as a proxy for individual household data in ascertaining whether respondents differ in important respects from nonrespondents and whether reluctant respondents differ from more cooperative respondents.
- Collecting information from nonrespondents in a follow-up survey to measure how they differ from respondents. There are a wide variety of techniques used in follow-up surveys, but most collect some key variables either from all or a randomly selected sample of nonrespondents. As might be expected, it is critical to have high response rates in follow-up surveys.

Follow-up surveys tend to be expensive because of the very intensive nonresponse conversion techniques that are employed to minimize nonresponse in the follow-up sample.

Recommendation 2-6: Research is needed to improve the modeling of response as well as to improve methods to determine whether data are missing at random.

NEW METRICS FOR UNDERSTANDING NONRESPONSE BIAS

The response rate is still useful for many reasons. It is reported at the survey level, and it documents differences between surveys. It has become ingrained, and survey sponsors and academic journals have established response rate requirements. It is also a useful tool for managing fieldwork. For these reasons, it is likely that response rates will remain an important indicator of data quality.

However, the response rate has serious shortcomings. It is not directly linked to bias. It is also not variable specific. Most important, its use in field operations may distort data collection practices—for example, by suggesting that interviewers should attempt to interview the remaining cases with the highest response propensity, which is not necessarily a strategy that will reduce bias.

Because the nonresponse rate can be such a poor predictor of bias, researchers have turned to developing new metrics for depicting the risk of nonresponse bias. This is not without its own dangers, and moving the focus from the response rate to more abstract measures may have unintended consequences. For example, interviewers hearing that "response rates don't matter" might mistakenly infer that their efforts to obtain relatively difficult cases did not matter. Moreover, because nonresponse bias is specific to a particular estimate (or model) and not to a survey in general, there could be confusion about the quality of a survey to support estimates other than those for which nonresponse bias has been studied.

Those interested in developing alternative indicators of bias have generally turned to two typologies: those at the survey level and those that pertain to estimate levels. Wagner (2008, 2011) proposed a typology that separates response rate from other survey-level indicators, identifies the data involved, and differentiates indicators based on implicit and explicit model assumptions. These indicators can be further differentiated as those using a response indicator (i.e., response rate); those using a response indicator and frames and paradata; and those using a response indicator, frames and paradata, and a survey variable or variables.

Researchers have increasingly recognized that indicators that also use paradata have some advantages. Examples of attempts to construct this class of indicators include a comparison of respondents and nonrespondents on auxiliary variables, variance functions of nonresponse or poststratification weights, and goodness-of-fit statistics of propensity models (Schouten et al., 2009). They utilize more data and are reported at the survey level. But, like response rates, they are mainly useful only in the context of a specific survey, whereas nonresponse bias is a statistics-level problem. As a result, they are not commonly used.

Representativity indicators (R-indicators) are an example of a new metric for assessing the effects of nonresponse.[1] R-indicators attempt to measure the variability of response propensities (van der Grijn et al., 2006; Cobben and Schouten, 2007). They do this by measuring "the similarity between the response to a survey and the sample or the population under investigation" (van der Grijn et al., 2006, p. 1). A survey that exhibits less variability in response propensities is likely to exhibit a better match between the characteristics of the respondents and the population they are meant to represent for the variables that the model uses to estimate the propensities.

R-indicators can be monitored during data collection to permit survey managers to direct effort to cases with lower response propensities and, in so doing, to reduce the variability among subgroup response rates. The indicator can be monitored during survey collection because the response propensities can be calculated with complete information available on the frame. In order for the R-indices to be completely comparable across surveys, they need to be estimated with the same variables. Wagner (2008) suggests that this would be facilitated if all surveys had a common set of frame data. However, there is no proof that such indices would apply across all types of surveys, or even across all relevant estimates within a given survey. More research that uses these indicators in survey settings is needed.

An example of the use of the R-index was provided by John Dixon of the Bureau of Labor Statistics in his presentation to the panel (Dixon, 2011). As shown in Figure 2-1, he charted the R-index for the Current Population Survey (top line) with the response rate for the survey (bottom line). He noted that, at 95 percent confidence intervals, the R-index is somewhat flatter than the response rate, which suggests that, for this survey, response propensities indicate a good match between the characteristics of the respondents and the population they are meant to represent.

Estimate-level indicators are indicators that use a response indicator, frame and paradata, and survey variables. They require an explicit model for each variable, and the model is usually estimated from the observed data and relies on the assumption that the missing data are missing at random (MAR). Wagner reported that there is very little research into non-MAR

[1]Additional content on R-indicators is provided by Wagner (2011).

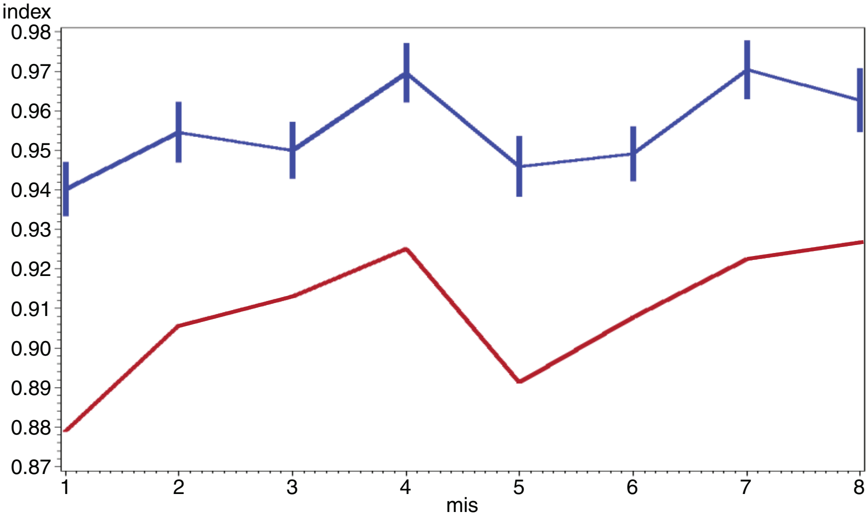

FIGURE 2-1 Plot of R-index (top line) and response rate (bottom line) for Current Population Survey cohorts by months in sample.
NOTE: Top line includes 95 percent confidence interval error bars around month-in-sample (mis) values for the R-index.
SOURCE: John Dixon presentation to the panel (Dixon, 2011).

assumptions, citing the work of Andridge and Little (2009) as an exception. This method also calls for the filling in of missing data. The results are valid within the context of the survey, but they are not necessarily comparable across surveys and, as a result, this method is not commonly used. Examples of these indicators are the correlations between post-survey weights and survey variables, variation of means across the deciles of survey weights, comparisons of early and late responders, and the fraction of missing information (FMI). The FMI is computed within a multiple imputation framework (Rubin, 1977) in which the model relates the complete frame data and paradata to the incomplete survey data. The FMI is the ratio of the imputation variance to total variance.

Balance indicators (B-indicators) were introduced by Särndal (2011) in a presentation at the annual Morris Hansen Lecture. They are an alternative indicator of bias—a measure of the lack of balance between the set of respondents and the population. The degree of balance is defined as the degree of fit between the respondent and population characteristics on a (presumably) rich set of frame variables. Särndal introduced a concept of a balanced response set, stating, "If means for measurable auxiliary variables are the same for respondents as for all those sampled, we call the response set perfectly balanced" (p. 4). The B-indicator can be used to measure how effective techniques for improving the balance of the sample, such as responsive design (see Chapter 4), have been.

In addition to the R-indicators and B-indicators, other indicators could be imagined, such as a measure based on the variance of the weights. Such indicators are promising, but, as Wagner reminded the committee, a research agenda on alternative indicators of bias should include research on the behavior of different measures in different settings, the bounds on nonresponse bias under different assumptions (especially non-MAR), how different indicators influence data collection strategies, and how to design or build better frames and paradata.

Recommendation 2-7: Research and development is needed on new indicators for the impact of nonresponse, including application of the alternative indicators to real surveys in order to determine how well the indicators work.

NEED FOR A THEORY OF NONRESPONSE BIAS

This chapter describes a large and growing body of research into the characteristics of nonresponse bias and its relationship (or lack of relationship) to response rates. While encouraging, the work has gone forward in piecemeal fashion and has not been conducted under an umbrella of a comprehensive statistical theory of nonresponse bias.

In his presentation to the panel, panel member Michael Brick suggested that a more comprehensive statistical theory would enhance the understanding of such bias and aid in the development of adjustment techniques to deal with bias under different circumstances (Brick, 2011). A unifying theory would help ensure that comparisons of nonresponse bias in different situations would lead to the development of standard measures and approaches to the problem. In the next chapter, the need for a comprehensive theory will again be discussed, this time in the context of refining overall adjustments for nonresponse.

3

Mitigating the Consequences of Nonresponse

Survey nonresponse has consequences, most notably the potential for nonresponse bias. The techniques and procedures for dealing with nonresponse bias depend on how one approaches the problem. Singer (2006) points out that statisticians have been concerned mainly with imputation and weighting as ways of adjusting for the bias introduced by nonresponse, while social scientists and survey methodologists have tended to focus on measuring, understanding, and reducing the nonresponse rates themselves.

Because of the negative effect that nonresponse can have on survey quality, in recent years survey researchers and managers have been responding very aggressively to the problem of growing nonresponse in surveys. However, as Couper (2011a) observed in the panel's workshop, the approaches have become more selective, and survey researchers are rejecting the older approach that aimed simply to maximize the overall response rate. The newer approaches target interventions at subgroups, at domains of interest, and at maximizing the response from specific cases based on their perceived special contribution to the quality of key statistics. Survey researchers are focusing less attention on increasing overall rates and are increasingly focusing on understanding the causes and correlates of nonresponse and making adjustments based on that understanding.

In this chapter, we explore some of the ways in which survey methodologists and managers are responding to the growing problem of survey nonresponse. We outline the results of some very good work that has gone into the development of weighting adjustments and adjustment models, and we document the increased use of paradata in nonresponse adjustment.

Some of this work is in early stages, and other work is more advanced. We make several recommendations for research to solidify and further advance these lines of development.

NONRESPONSE WEIGHTING ADJUSTMENT METHODS[1]

The need for nonresponse adjustment arises because probability samples, in which all units have a known, positive probability of selection, require complete responses. Without other non-sampling errors, estimators for probability samples are approximately design-unbiased, consistent, and measurable. Base weights, or inverse probability selection weights, can be used to implement standard estimators.

One possible simple estimator is the ratio mean, which is approximately unbiased and consistent for the population mean:

$$\hat{\bar{Y}}_\pi = \frac{\hat{Y}_\pi}{\hat{N}} = \frac{\sum d_i y_i}{\sum d_i},$$

with d_i equal to the inverse of the probability of selection. If some of the sample units do not respond (unit nonresponse), and the estimator is unchanged, then the estimator may be biased:

$$\hat{\bar{Y}}_0 = \frac{\sum_r d_i y_i}{\sum_r d_i}.$$

The bias can be expressed in two ways:

(1) A *deterministic* framework assumes the population contains a stratum of respondents and a stratum of nonrespondents. Let the population means in the two strata be $\bar{Y}_r$ and $\bar{Y}_{nr}$, respectively. The respondent stratum is R percent of N, and the bias of the unadjusted estimator is

$$\text{bias}\left(\hat{\bar{Y}}_0\right) \doteq (1-R)\left(\bar{Y}_r - \bar{Y}_{nr}\right).$$

In the deterministic view, the bias arises when the means of the respondents and of the nonrespondents differ.

(2) A *stochastic* framework assumes every unit in the population has some non-zero probability of responding (its response propensity). The bias of the unadjusted estimator is

[1]The discussion of nonresponse weighting adjustment methods is abstracted from the presentation by Michael Brick at the panel's workshop (Brick, 2011).

$$\text{bias}\left(\hat{\bar{Y}}_0\right) \approx \frac{1}{N\bar{\varphi}} \sum_{i=1}^{N} \left(Y_i - \bar{Y}\right)\left(\varphi_i - \bar{\varphi}\right),$$

where φ_i is an individual response propensity and $\bar{\varphi}$ is the mean of the response propensities. Thus, in the stochastic view, the bias arises when the characteristic and response propensity covary.

A natural adjusted estimator is then

$$\hat{\bar{Y}}' = \frac{\sum_r d_i \hat{\varphi}_i^{-1} y_i}{\sum_r d_i \hat{\varphi}_i^{-1}},$$

where $\hat{\varphi}_i$ is an estimate of the response propensity and $d_i \hat{\varphi}_i^{-1}$ represents the adjusted weight w_i'.

The selection of the weighting framework—deterministic or stochastic—depends then on the theoretical model of the response mechanism. In other words, the underlying model is the rationale for the selection of the adjustment scheme.

Most adjustments now in use assume that the missing data are missing completely at random (MCAR) or missing at random (MAR). The MCAR assumption holds if all of the units in the population have the same probability of responding, that is, if the respondents are just a smaller random sample. MCAR means that the distribution of the missingness (an indicator for whether the unit responds or not) is independent of the y-variable and all auxiliary (or x) variables.

MAR is a more realistic assumption than MCAR. MAR implies that the probability of response does not depend on the y-variable once we control for a vector of known x-variables. Weighting class adjustment schemes that define subgroups using the auxiliary data, assuming that the sample units within the subgroups ($h = 1,\ldots,H$) have the same response propensity, are consistent with the MAR assumption. These methods adjust the weights for respondents in the group with $\hat{\varphi}_i = \hat{\varphi}_h \ \forall\, i \in h$.

This type of estimator is either a weighting-class estimator or a post-stratified estimator, depending on the type of data available for computing the adjustment. If the data are at the sample level (known for sampled units but not for the entire population), it is a weighting-class estimator; if the data are at the population level, then it is a post-stratified estimator. The adjustment requires that sample members can be divided into cells using the vector of observable characteristics.

In the weighting-class approach, the adjusted weight is calculated in four stages:

1. A base weight is calculated that is the reciprocal of the probability of selection of the case under the sample design.
2. When there is nonresponse and the eligibility of the nonrespondents cannot be determined, the base weights for these nonrespondents are distributed into the eligible nonresponse category based on the proportion of the weights that are eligible in the respondent set.
3. These weights are adjusted to compensate for eligible nonrespondents.
4. A final weight for the eligible respondent cases is computed as the product of the base weight, the eligibility adjustment factor, and the nonresponse adjustment factor (Yang and Wang, 2008).

In choosing weighting classes for the adjustment in stage 3, bias is eliminated when the variables and classes are such that either:

(1) $\hat{\varphi}_i = \bar{\varphi}_h \ \forall\, i \in h$ or

(2) $Y_i = \bar{Y}_h \forall\, i \in h.$

That is, the bias from nonresponse is eliminated if, within the weighting cells, all cases have the same response propensity or the same value for the survey variable. As noted in the stochastic model of nonresponse, nonresponse bias only exists when the response propensities and the outcomes are correlated. However, since most surveys have a multitude of survey outcomes, the idea of using classes that are related only to the response propensities is commonly adopted.

Models constructed to meet stage 1 are called *response propensity stratification*, and those designed to meet stage 2 are referred to as *predicted mean stratification*. The classes themselves are sometimes formed by subject matter experts, based on information on the key survey outcomes.

An empirical method that is used often with categorical data is to form weighting classes by using classification software such as CART, CHAID, or SEARCH. Often the dependent variable is the response (respondent or not), and sometimes the survey outcomes are used as dependent variables, depending on the criteria being used. In either case, this approach may result in a very large number of weighting classes. Eltinge and Yansaneh (1997) suggested methods to test whether appropriate classes are formed.

Many alternative methods of making these adjustments are sometimes used. We describe several of these alternatives below.

Propensity Model Approach

The propensity model approach uses multiple logistic regression analysis (or some similar approach) to examine the nonresponse mechanism and calculate a nonresponse adjustment. In this method a response indicator is regressed on a set of independent variables, such as those used to define weighting class cells. A predicted value derived from the regression equation is called the propensity score, which is simply an estimated response probability (Rosenbaum and Rubin, 1983). Survey population members with the same observable characteristics are assigned the same propensity score. The response propensity can be used to adjust directly by using the inverse of the estimated response propensity to adjust the weights for the respondents. This is the response propensity stratification, although in many cases the propensity scores are used to divide the sample into propensity classes based on quintiles of the distribution, and the average propensity within the class is the adjustment.

Advantages of propensity-weighting methods over the traditional weighting-class methods are that continuous variables can be used to define cells, the models can accommodate a large number of variables, and the technique is simple to apply (Hazelwood et al., 2007). If large adjustments are avoided by using classes rather than the inverse of the estimated response propensities, then the methods can be as stable as other weighting-class methods. Like other response probability adjustments, this approach also implicitly assumes that one weighting adjustment is sufficient to address nonresponse bias in all estimates.

Selection Models

Heckman (1979) first proposed the sample selection model for regressions. The model is based on the observation that respondents self-select to participate in a survey, either explicitly by refusing to participate or implicitly through inability to answer or be contacted.

Selection models are the conventional method among empirical economists for modeling samples with nonresponse (or other types of selectivity). Sampling statisticians have viewed this approach with skepticism, mainly because most selection models make strong assumptions about the nonresponse mechanism that may not hold in practice. Selection models are also typically variable-specific solutions, in the sense that the model is constructed for one particular estimate and cannot be used for a wide variety of statistics. While this feature can be of benefit because selection models may improve the quality of individual estimates, survey users are typically interested in producing many statistics. The goal is most often a consistency among estimates (e.g., the sum of the estimates for males and females should equal the total) that the selection models do not confer.

The most popular form of selection model requires an explicit distributional assumption. In principle, different selectivity corrections could be made from a given set of data, depending on the model to be estimated.

Raking Ratio Adjustment Approach

Raking ratio adjustments are used to benchmark sampling weights to known control totals and can be considered a form of multidimensional post-stratification. This approach reduces sampling error through the use of auxiliary variables correlated to survey response and has been used to reduce nonresponse bias (Brick et al., 2008). The advantage of raking is that more variables that are correlated with response propensities and outcome variables can be included in the weighting process without creating large weight adjustments. Like post-stratification and weighting-class methods, careful review of the weights is required to make sure large weight adjustments are not introduced by the raking process.

Calibration

Post-stratification and raking are two specific methods of calibration, as described by Särndal (2007). Calibration is a method of computing weights in a manner that equates the sum of the calibrated totals to totals defined by auxiliary information. Calibrated weights can then be used to produce estimates of totals and other finite population parameters that are consistent internally, as discussed above in raking.

Calibration is used to correct for survey nonresponse (as well as for coverage error resulting from frame undercoverage or unit duplication). Kott and Chang (2010) showed that calibration weighting treats response as an additional phase of random sampling. This method is particularly valuable when many important auxiliary variables are related either to response propensity or to the key survey outcomes. As a result, it has been heavily studied for use in countries with population registers or when the sampling frame is rich in auxiliary data, such as in establishment surveys.

Mixture Models

Selection models can be thought of as expressing the joint distribution of the outcome and "missingness" as the product of the distribution of the missing data mechanism conditional on the outcome variable and on the marginal distribution of the outcome variable. An alternative approach is to write the joint distribution as the product of the distribution of the outcome conditional on the missingness mechanism and on the marginal distribution of the missingness mechanism. The two approaches do not result in the same estimates in some situations. Little (1993) described the difference in the two approaches and discussed when pattern mixture models might be preferred.

Observations About Weighting Adjustment Approaches

All of these weighting adjustment schemes depend very heavily on the availability of auxiliary data that are highly correlated with either the response propensities or the key outcomes. Without these types of data, the adjustments are ineffective in reducing nonresponse bias. As response rates decline, these weighting adjustments may become even more important tools for producing high-quality survey estimates.

Recommendation 3-1: More research is needed on the use of auxiliary data for weighting adjustments, including whether weighting can make estimates worse (i.e., increase bias) and whether traditional weighting approaches inflate the variance of the estimates.

In his summary, Brick (2011) makes the case for the development and refinement of survey theory, suggesting that empirical adjustment methods may work in many cases, but pointing out that they are unsatisfying in several ways. Some possible paths to a solution would be to develop a more comprehensive theory relating response mechanisms to nonresponse bias, and a more comprehensive statistical theory of adjustment to deal with different types of statistics.

Recommendation 3-2: Research is needed to assist in understanding the impact of adjustment procedures on estimates other than means, proportions, and totals.

USE OF PARADATA IN REDUCING NONRESPONSE AND NONRESPONSE BIAS

There is a growing interest in paradata—that is, data about the process by which the survey data were collected that are obtained in the process of conducting the survey. Paradata encompass information about the interviews (times of day interviews were conducted and how long the interviews took); about the contacts (how many times contact was made with each sample person or how many attempts to contact the sample person were made, the apparent reluctance of the sample person); as well as survey modes (such as phone, Web, e-mail, or in person). These data have many uses. They help in managing the survey operation (scheduling and evaluating interviewers) and assessing its costs. They are also important for understanding the findings of a survey and making inferences about nonrespondents. Indeed, there is a long history in the research literature of collecting additional data (what has become known as paradata) for nonresponse.

Hansen and Hurwitz (1946) suggested two-phase sampling, with the second phase of sampling looking at nonrespondents by using an intensive follow-up of units selected for the second phase. If data can be collected from all the sampled second-phase nonrespondents, then standard two-phase sampling weights can be developed to eliminate nonresponse bias. Even with an incomplete response at the second phase, the potential for bias can be reduced by using information from the additional second-phase sample.

Significant advances have been made in the state of the science for using paradata for reducing nonresponse bias (Olson, 2013). Today there are two main options for reducing such bias. One is to use paradata to introduce new design features to recruit uncontacted or uncooperative sample members—and, hopefully, respondents with different characteristics—into the respondent pool. The new design features rely on the use of paradata in a "responsive" design. The second option is to use paradata in adjusting base weights of the respondents. A third use of paradata is to use the data to better understand the survey participation phenomenon so that future surveys may reduce nonresponse, but this use does not result in reducing nonresponse bias for the survey at hand.

The initial focus of research on paradata was to explore nonresponse rates. The types of paradata that were considered as predictors of response were respondent-voiced concerns, the presence of a locked entrance or other safety measures, a multiunit building, and an urban setting (Campanelli et al., 1997; Groves and Couper, 1998).

More recently the work on paradata has taken a new direction and has focused more on the reduction of nonresponse bias by using paradata in responsive designs or in weighting adjustments. Adjustments that are effective in reducing nonresponse bias must be based on data that are predictive of the likelihood of participating in a survey or on the key survey outcomes or both (Little, 1986; Kalton and Flores-Cervantes, 2003; Little and Vartivarian, 2005; Groves, 2006; Kreuter et al., 2010).

The current challenges for paradata research are to enhance the underlying theory (e.g., what paradata are correlated with both response propensities and outcome measures); to better understand measurement error in the paradata and what effect these errors have on the utility of the paradata for reducing nonresponse bias; to operationally assign new tasks for interviewers that are feasible and do not detract from their ability to conduct the interviews; and to better understand the environment for the interview—such as doorstep interactions, reasons for non-participation, and the quality of interviewer observations of the neighborhoods and the housing unit—so that better paradata measures can be developed.

Generally, the quality of paradata is relatively good if the data are automatically generated. When interviewers are asked to collect additional data that are not a byproduct of the data collection process, there is often a drop in quality. Additional data collection requirements often lead to substantial missing data rates. Likewise, when interviewer judgment is required, the data are of varying quality (see Casas-Cordero, 2010; Kreuter and Casas-Cordero, 2010; McCulloch et al., 2010; and West and Olson, 2010).

Kreuter holds that paradata carry a compelling theoretical potential for nonresponse adjustment. With paradata, the development of proxy variables is possible, and it is also possible to identify large variations in correlations across outcome variables. However, research has shown that although interviewers are good at making observations that are relevant for the primary act of data collection, they can have difficulty in collecting the additional proxy variables (*Y*s).

According to Kreuter, a case can be made for further collaboration with subject-matter experts, statisticians, psychologists, and fieldwork staff in the refinement of paradata, so these are examples of areas in which further investigation may be fruitful. Such collaboration would be useful, for instance, with substantive researchers to develop interviewer observation measures for labor force surveys (at-home pattern), health surveys (too ill to participate), housing surveys (condition of the dwelling), crime surveys (bars on windows), and educational surveys (literacy). Collaboration with statisticians could help improve statistical models, providing answers to such questions as how to balance multiple predictors of response and *Y*s, how to handle large and messy data, how to model and cluster sequences of unequal length, and how to address issues of discrete times and mixed processes. Psychologists, in collaboration with survey methodologists, could aid in understanding the factors that drive errors in interviewer observation, how training could improve ratings, how much error can be tolerated, and what the quality is relative to other sources. Finally, collaboration with fieldwork staff could help identify the costs associated with paradata collection as well as cheaper alternatives, the risks in interviewer multitasking, the appropriate level of observation, and ethical and legal matters issues that need to be resolved.

CONCLUDING OBSERVATION

A focus on reducing nonresponse by using paradata in responsive designs, or through other means, should not lead to neglect of other sources of error in survey estimates. In particular, measurement error (overreporting, underreporting, or misreporting) should be addressed, particularly in terms of the possible interaction with nonresponse.

> **Recommendation 3-3: Research is needed on the impact that reduction of survey nonresponse would have on other error sources, such as measurement error.**

4

Approaches to Improving Survey Response

In previous chapters, we have summarized evidence that survey nonresponse is a growing problem. In a paper that has been cited often in this report, Brick and Williams (2013) raised the disturbing possibility, based on their analyses, that the intrinsic rate of increase in nonresponse in U.S. household surveys might be 0.5 percentage points or so per year. We have provided evidence that survey nonresponse is more prevalent with some modes of data collection than others, that it can produce errors in survey estimates, and that sophisticated adjustment techniques are required to ameliorate the impact it has on estimates.

In Chapter 1, we laid out many potential reasons for the growth in nonresponse, concluding that the decision of a person to respond or not respond to a survey involves several key factors. The elaboration of these factors provides a convenient conceptual point of departure for the review in this chapter of approaches to improving response.

Ultimately, responding or not responding to a survey is a decision made by a sample member (or a proxy). These decisions are informed by social factors (e.g., social disorganization, crime); membership in a social category or group (e.g., age, gender, political party); the survey setting (e.g., interviewer-mediated or self-administered); the social climate (e.g., time pressures, general concerns about privacy); the proliferation of surveys; and so on. Approaches to improving survey response must take these factors into account.

One possibility suggested by researchers is that the decline in response rates reflects a corresponding increase in the overall level of burden that surveys place on sample populations. Thus, this chapter begins with a discussion of respondent burden and the relationship of real and perceived burden with the willingness to take part in surveys. Several of the methods we discuss in detail, such as matrix sampling or greater reliance on administrative records, represent attempts to greatly reduce the burden on respondents.

We then discuss several approaches that are being taken or have been proposed to increase survey response rates. The first group of approaches involve sampling procedures—respondent-driven sampling (RDS), matrix sampling, and address-based sampling (ABS)—that may have implications for response rates. Other approaches are aimed at increasing our understanding of the conditions and motivations underlying nonresponse; changing the interaction of interviewer and respondent; making better use of information collected in the survey process to adjust the collection strategy in an attempt to achieve higher response rates, lower costs, or both; using other data sources (e.g., transaction data and administrative data) as strategies to reduce burden; and using mixed-mode methods of data collection.

UNDERSTANDING AND REDUCING RESPONDENT BURDEN

It is widely accepted that nonresponse is, at least in part, related to the perceived burden of taking part in a survey. It is less clear how to define and measure burden. Two flawed but widely used indicators of burden are the number of questions in the survey and the average time taken by respondents to complete those questions. The notion that the time used in responding is directly related to burden seems to be the working principle behind the federal government's Paperwork Reduction Act. This act requires the computation of burden hours for proposed federal data collections and has provisions that encourage limiting those burden hours. The use of a time-to-complete measure (in hours) for response burden is fairly widespread among the national statistical agencies (Hedlin et al., 2005, pp. 3–7).

The factors to be taken into account in the calculation of burden hours are important considerations. Burden could relate only to the actual time spent on completing the instrument, but it also could take into account the time respondents need to collect relevant information before the interviewer arrives (for example, in keeping diaries) and any time after the interview is completed. For example, the time incurred when respondents are recontacted to validate data could also be taken into account. Without these additions, a measure that uses administration time or total respondent time per interview as a metric for burden is clearly problematic.

Bradburn (1978) suggested that the definition of respondent burden should include four elements: interview length, required respondent effort, respondent stress, and the frequency of being interviewed. The effort required of respondents could refer to the cognitive challenges of a task (e.g., remembering the number of doctor visits in a year) or the irritation of answering poorly written questions. The frequency of being interviewed could refer either to the multiple interviews required by longitudinal surveys or to the increased likelihood of being selected for a study as society's demands for information increase. In addition, some complex studies may involve requests for biomarkers, record linkages, multiple modes of response, and more. Some of these requests (e.g., for waist measurement) may be perceived as intrusive; if so, this may increase the sense of burden. Furthermore, multiple requests require many decisions using different criteria (e.g., the decision to allow waist measurement may use criteria different from those used to decide about providing a DNA sample), and these decisions may add to burden. Difficult or upsetting questions are, in this view, more burdensome than easy or enjoyable ones, and any measure of burden should reflect the cognitive and emotional costs associated with the questions as well as the time spent answering them.

Presser and McCulloch (2011) documented a sharp increase in the number of federal surveys. Although the probabilities of being interviewed for a survey are likely still relatively small, members of the general population are at somewhat greater risk of being interviewed because of that proliferation. Presser and McCulloch argued that the increased number of survey requests people are subjected to may be one reason for the decline in response rates.

It is clear that "burden" has many possible aspects. Progress in understanding burden and its impact on survey response must begin with an analysis of the concept, its dimensions, and how it is operationalized. Unfortunately, there is little research to show conclusively that there is a causal relationship between measured burden and propensity to respond. The research design needed to examine such a relationship would be influenced by the type and extent of burden that is imposed, and in many cases, sample members cannot know the extent of the burden of a specific request until they complete the survey—or at least until they are contacted. An increasing number of sample members are not contacted; consequently, measuring the number of survey requests that households or individuals receive and relating this to their overall response propensity is problematic. To fully understand the impact of burden on response, more testing of the so-called burden–participation hypothesis is needed.

The literature includes a few studies of the factors affecting perceptions of burden, which usually focus on survey instrument length and difficulty. In a 1983 paper, Sharp and Frankel examined the length of the survey instrument, the effort required to answer some of the questions, and the impact of a request for a second interview approximately one year after the first. Behavioral indicators and responses to a follow-up questionnaire were used to measure the perception of burden. The study found that the instrument length produced a statistically significant (although generally small) effect on perceived burden. The perception of burden was more strongly influenced by attitudinal factors than by the survey length. Respondents who see surveys as useful rated the survey as less burdensome than those who did not. Similarly, those who saw the survey questions as an invasion of privacy rated the survey as more burdensome.

A literature review by Bogen (1996) found mixed results concerning the relationship between questionnaire length and response rate. Bogen reviewed both non-experimental and experimental studies that were available in the mid-1990s. She concluded that "the non-experimental literature paints a picture about the relationship between interview length and response rates that is not uniform" (p. 1021). Likewise, the experimental literature produced some studies that found that shorter interviews yielded higher response, others that found longer interviews to yield higher response, and still others that suggested that the length of the interview did not matter. She concluded that the experimental studies could have been affected by logistical and scheduling considerations and interviewer expectations. Clearly, reasons other than interview length are at play in the decision of an individual to respond or not respond.

Very little is known about where in the process of receiving and responding to the request to participate in a survey the sample member's perception of burden is formed or how well-formed or fluid this perception is. Attitudes toward burden may precede any request and may insulate the sample member from processing new information about a specific survey, or attitudes may be quickly formed based on an impression of a specific request. The survey topic and other information are often communicated in the advance letters used in many surveys, but whether the letters are received, read, and understood is not known.

Without a very basic understanding of the dimensions of burden and the factors that generate the perception of burden, it is difficult to take the next step and determine the relationship between perception of burden and the propensity to respond.

Recommendation 4-1: Research is needed on the overall level of burden from survey requests and on the role of that burden in the decision to participate in a specific survey.

The questions to be addressed in the recommended research program include: What are the dimensions of response burden? How should they be operationalized? What factors (e.g., time, cognitive difficulty, or invasiveness, such as with the collection of biomarkers) determine how potential respondents assess the burden involved in taking part in a survey? How much can interviewers, advance letters, or other explanatory or motivational material alter perceptions about the likely burden of a survey?

IMPROVING RESPONSE IN TELEPHONE AND MAIL SURVEYS

Telephone Surveys

This report has documented that some of the most troublesome declines in response rates in social science survey operations have taken place in telephone surveys. This is particularly vexing because of the extensive reliance on this mode for sample member recruitment and data collection. This reliance was summarized during a panel workshop by Paul Lavrakas, who chaired an American Association for Public Opinion Research (AAPOR) task force on including cell phones in telephone surveys (American Association for Public Opinion Research, 2010a).

Drawing on examples from six ongoing cell phone collections (see American Association for Public Opinion Research, 2010a), he described the current environment for telephone surveying as one in which only 67 percent to 72 percent of households have a landline and just 8 percent to 12 percent of households have *only* a landline. On the other hand, 86 percent to 91 percent of households have a cell phone, and 27 percent to 31 percent of households have *only* a cell phone. Only a very few (1 percent to 2 percent) of households have neither a landline nor a cell phone.

The growth in cell phone usage poses a severe challenge to telephone surveys. Lavrakas noted that federal regulations that limit calling options for cell phones and the telephony environment in the United States create special challenges for researchers trying to conduct surveys that include cell phone numbers.

Despite these obstacles, many random digit dialing (RDD) surveys now include cell phones. The Centers for Disease Control and Prevention (CDC) has been in the forefront of testing and implementing cell phone data collection. In 2006, the CDC Behavioral Risk Factor Surveillance System (BRFSS) responded to the growing percentage of cell phone–only households by testing changes in BRFSS survey methods to accommodate cell phone data collection. The tests included pilot studies in 18 states in 2008, and in 2010 the test was expanded to 48 states. These pilot studies gathered data from test samples including landline and cell phone–only households. This extension to cell phone collection has increased the complexity of the survey operation and data processing, including the need for different weighting techniques by mode. In 2012, the proportion of all completed BRFSS interviews conducted by cellular telephone was approximately 20 percent (Centers for Disease Control and Prevention, 2012).

In terms of response rates, the AAPOR panel found that landline RDD surveys rarely had response rates higher than 40 percent; they are mostly in the 10 percent to 25 percent range, and often less than 10 percent. Response rates for cell phone RDD surveys are even lower: rarely above 30 percent, and mostly in the 10 percent to 15 percent range.

The AAPOR panel concluded that, as with other surveys, the main reasons for telephone survey nonresponse are noncontact, refusals, and language barriers. (Language barriers involve a failure to communicate, which often results in a nonresponse if an interpreter is not available to translate the questions and answers.)

Noncontacts are higher with shorter periods of field collection and are affected by the increased availability of caller ID, which allows households to screen incoming calls. Calling rules that are imposed by survey management may limit the number and timing of callbacks and thus may raise the noncontact rates. On the other hand, messages left on voice mail and answering machines may reduce noncontacts.

There are many reasons for refusals. Among the main reasons are the failure to contact sample members ahead of time; negative attitudes toward the sponsor and the survey organization; the survey topic; the timing of the request; confidentiality and privacy concerns; and a belief that responding will be burdensome. In some cases, the interviewers use poor introductory scripts, and they may not be able to offer incentives, or they may use incentives poorly (Lynn, 2008).

Mail Surveys

Low response rates have long been considered a major problem for mail surveys, so much so that much of the early research on improving response rates focused on mail surveys. In 1978, Heberlein and Baumgartner carried out a meta-analysis to test the effects of a large number of survey characteristics on mail response rates. Their final model predicted about two-thirds of the variation in the final response rate. Variables that had a positive effect on response rates were (a) more contacts with the sample household via advance letters, reminder postcards, sending replacement questionnaires, and telephone prompts; (b) a topic of interest to members of the target group; (c) government sponsorship of the survey; (d) target populations, such as students and military personnel, that were more likely to take part in surveys than the general population as a whole; (e) the use of special follow-up procedures, such as more expensive mailing procedures (e.g., certified mail) or personal contacts; and (f) incentives included with the first mailing. However, three factors had a negative effect on response rate: (1) the collection of marketing research information to benefit a specific firm; (2) a general population sample; and (3) long questionnaires.

Goyder (1982) replicated this study with similar results, except that the negative effect of market research sponsorship disappeared. Other studies have elaborated on these basic findings, paying particular attention to the effects of respondent incentives (Fox et al., 1988; Church, 1993).

The lessons of this early research were codified in the development of a comprehensive system designed to achieve higher response rates for mail surveys. The total design method (TDM), developed by Dillman (1978), was guided primarily by social exchange theory, which posits that questionnaire recipients are most likely to respond if they expect that the perceived benefits of responding will outweigh the costs of responding. TDM emphasizes how the elements fit together more than the effectiveness of any individual technique.

Specific well-known TDM recommendations that have been shown to be likely to help improve responses include the following:

- Use graphics and various question-writing techniques to ease the task of reading and answering the questions.
- Put some interesting questions first.
- Make the questions user-friendly.
- Print the questionnaire in a booklet format with an interesting cover.
- Use bold letters.
- Reduce the size of the booklet or use photos to make the survey seem smaller and easier to complete.
- Conduct four carefully spaced mailings beginning with the questionnaire and a cover letter and ending with a replacement questionnaire and cover letter to nonrespondents seven weeks after the original mailing.
- Include an individually printed, addressed, and signed letter.
- Print the address on the envelopes rather than use address labels.
- Explain that an ID number is used and that the respondent's confidentiality is protected.
- Fold the materials in a way that differs from an advertisement.

To adapt the original TDM to different survey situations, such as those used in mixed-mode surveys, Dillman developed the tailored design method (Dillman et al., 2009), in which the basic elements of survey design and implementation are shaped further for particular populations, sponsorship, and content. Despite these advances in understanding the determinants of high response rates in mail surveys, which are grounded in research covering more than a quarter of a century, the challenges continue.

NEW FRAMES AND METHODS OF SAMPLING

It is appropriate to begin consideration of approaches to improving survey response or lowering survey costs with the design of the survey. This section discusses several options, ranging from adopting a whole new approach to survey design (ABS) to more traditional methods for improving sample design.

Address-Based Sampling

Survey researchers have recently begun to explore the use of lists of mailing addresses as sampling frames. There are several reasons for this development, including the potential for cost savings (for surveys relying on area probability samples) and the potential for better response rates (for surveys relying on RDD sampling). Iannacchione et al. (2003) were the first to publish results on the use of the U.S. Postal Service Delivery Sequence File (DSF) as a potential sampling frame, a method that has come to be known as ABS. Link et al. (2008) were the first to publish work on switching from telephone to mail data collection and from RDD to ABS sampling. They also coined the term "address-based sampling."

Link and his colleagues (2008) compared mail surveys based on ABS with telephone surveys based on RDD using the BRFSS questionnaire in six low-response rate states (California, Illinois, New Jersey, North Carolina, Texas, and Washington). The BRFSS covers all 50 states plus the District of Columbia. The pilot survey was conducted in parallel with the March, April, and May 2005 regular RDD data collection process. In five of the six states, the mail/ABS BRFSS achieved a higher response rate than the regular telephone/RDD BRFSS. However, after this testing, the ABS design was not implemented in the BRFSS for reasons that have not been documented.

The National Household Education Survey (NHES) program has also undertaken to transition from RDD to an ABS methodology. This new methodology was used recently in a very large NHES field test. The field test included several experiments to discover the best methods for a mail ABS approach. The experiments compared different questionnaires and survey materials, levels of incentives and mailing services, and the effects of including a pre-notice letter. Preliminary results from the field test indicate that ABS response rates were substantially higher than those attained in the last round of RDD surveys (Montaquila and Brick, 2012).

In addition to the testing and experimentation conducted with the BRFSS and NHES surveys, several other surveys have adopted an ABS design. The Health Information National Trends Survey (HINTS) of the National Cancer Institute (which used an ABS component in addition to an RDD component in 2007), the Nielsen TV Ratings Diary (which moved from a landline RDD frame to ABS), and Knowledge Networks (which switched from RDD to ABS recruitment for its online panel surveys) will yield additional information on the ability of this design to increase response over time.

In summary, research has so far indicated that ABS provides good coverage and is also cost-effective. In conjunction with mail data collection, it appears to produce higher response rates than telephone interviewing and RDD sampling produce. However, it has been pointed out that when eligibility rates fall below a certain point, it is no longer cost-effective (Amaya and Ward, 2011). There are still major issues to be researched concerning ABS, including within-household selection of a single respondent (Montaquila et al., 2009).

Recommendation 4-2: Research is needed on how to best make a switch from the telephone survey mode (and frame) to mail, including how to ensure that the right person completes a mail survey.

Respondent-Driven Sampling

Some populations are hard to include in surveys because they are very rare, difficult to identify, or elusive. When these groups are the target population for a survey, they have very high non-interview and nonresponse rates. According to a presentation to the panel by Heckathorn (2011), many hard-to-reach populations cannot be sampled using standard methods because they lack a sampling frame (list of population members), represent small proportions of the general population, have privacy concerns (e.g., stigmatized groups), or are part of networks that are hard for outsiders to penetrate (e.g., jazz musicians).

The traditional methods for sampling such hard-to-reach populations all have problems. One traditional method is to sample population members through location sampling (e.g., selecting a sample of homeless persons by selecting persons who sleep at a homeless shelter). However, such samples would exclude members who avoid those contacts; as a result, those with contacts at sample locations may differ systematically from those without them. Another approach is to draw a probability sample of population members who are accessible in public venues, but the coverage of those samples is limited because it excludes those who shun public settings.

Snowball samples (or chain-referral methods) may offer better coverage because respondents are reached through their social networks, but they produce convenience samples rather than probability samples. Hence, there is a dilemma. There is a trade-off between maximizing coverage of hard-to-reach populations and realizing the statistical advantages offered by probability sampling. Heckathorn (2011) argued that RDS resolves this dilemma by turning chain referral into a probability sampling method.

RDS starts with eligible "seeds" to gain entry into the network. Then the seeds recruit other members of the population. There are often incentives both for participation and for recruiting. Advocates claim that there is a lower cost per case than with traditional designs; that it reduces time and demands on interviewers; that it can reach populations that traditional methods cannot; and that it eliminates travel and personal safety issues. However, the method relies on a number of critical "assumptions that must be met to determine if it is an appropriate sampling method to be used with a particular group" (Lansky et al., 2012, p. 77). Included among the assumptions is that the recruited population must know one another as members of the group, and that the members are adequately linked so that the whole population is covered.

There are several approaches for measuring nonresponse in network samples. One approach is to compare the reported network composition with the yield of actual recruits. For example, in Bridgeport, Connecticut, a sample of drug injectors yielded only blacks, although respondents reported knowing many Hispanic injectors. In this case, recruitment excluded an important group. The interview site was in a black neighborhood, where Hispanics did not feel comfortable. The solution was to move the interview site to neutral ground in downtown Bridgeport. Subsequently, recruitment of both blacks and Hispanics was more successful, and the reported network converged with the composition of the recruited sample. Comparing self-reported network composition and peer recruitment patterns provided a qualitative measure of representativeness even though it could not be expressed in a traditional response rate.[1]

Another approach is to ask those who are responsible for recruiting respondents about those who refused to be recruited. This technique was used in a CDC study of young drug injectors in Meriden, Connecticut. The most common reason for refusing recruitment was being "too busy" (see also Iguchi et al., 2009).

Experience to date suggests that the operational aspects of reducing nonresponse in RDS are challenging, to say the least, and the ability of the method to yield results much like probability samples is not yet proven. In her presentation to the panel's workshop, Sandra Berry (2011) suggested that it is important for the future of this survey technique that research be conducted on the following operational aspects of this still-in-development method:

[1]This study was conducted as part of a research grant from CDC to the Institute for Community Research; see http://www.incommunityresearch.org/research/nhbsidu.htm [March 2013].

- How well does RDS perform in community survey contexts? How do we judge this?
- How can we get better measures of network size from individuals?
- What features of RDS can be altered and at what cost to response rates, overall bias, or the variance of the estimates?
- In what situations (populations or modes of contact and data collection) does RDS work well?
- Which of RDS's assumptions are likely to be met in practice, and which are likely to be violated?
- How can RDS enhance and integrate with traditional data collection?

Matrix Sampling

While the goal of RDS is to identify and maximize responses from a hard-to-reach population at a reasonable cost, the goal of matrix sampling is to reduce any particular respondent's burden and thereby improve survey response rates. Matrix sampling is a procedure in which a questionnaire is split into sections of questions, and each section is then administered to subsamples of the main sample. Even though individual survey respondents answer only a part of the questionnaire, estimates can be obtained for all the variables derived from survey items (Shoemaker, 1973). Partitioning a long questionnaire into smaller, bite-sized pieces is a way to encourage people to respond more readily.

There are several examples from the fields of educational assessment, federal statistics, and public health (Gonzalez and Eltinge, 2007) in which matrix sampling has been applied:

- The largest ongoing example of matrix sampling is the National Assessment of Educational Progress (NAEP), which surveys the educational accomplishments of students in the United States. Because NAEP assesses a large number of subject-matter areas, it uses a matrix sampling design to assess students in each subject. Blocks of items drawn from each content domain are administered to groups of students, thereby making it possible to administer a large number and range of items while keeping individual testing time to an hour. Because of its design, NAEP reports only group-level results.[2]
- One of the major U.S. surveys to have investigated matrix sampling as a way to reduce burden and improve response is the Consumer Expenditure Quarterly Interview Survey (CEQ). Gonzalez and Eltinge (2009) conducted a simulation study using CEQ data from April 2007 to March 2008

[2]See http://nces.ed.gov/nationsreportcard [March 2013] for information about NAEP.

for the full questionnaire. They then split the dataset into six subsamples, each containing a subset of items and explored different ways of imputing a full dataset for each subsample.

- Munger and Loyd (1988) looked at the viability of matrix sampling in a survey of 307 randomly selected school principals in the state of Virginia. The principals were randomly assigned to four separate groups. The first group, which consisted of 100 principals, was assigned the full questionnaire containing 61 items. The remaining three groups, each consisting of 69 principals, were each assigned a shortened questionnaire containing 27 items. The study found that the survey sample members were more likely to respond to a shortened questionnaire than to the lengthy version, even though a larger percentage of those assigned the long questionnaire said they always responded to surveys, while a larger percentage of those assigned to one of the short questionnaires said they seldom responded to surveys. The matrix sampling design required a larger overall sample to achieve the same reliability.
- Thompson et al. (2009) used matrix sampling for a survey on library services assessment to explore burden reduction and response rates. The long form of the questionnaire consisted of 22 survey items. Randomly selected participants were asked to complete a short version that contained 8 of the 22 items. The completion rates were higher for short-form survey participants relative to long-form survey participants. Moreover, the long form elicited participation from respondents who were more positive about library services, thereby exaggerating the positive assessment of library services.

Available research indicates that for lengthy surveys, matrix sampling methodology may improve cooperation rates and reduce break-offs, straight-line responses, and nonresponse to "filter" questions in order to avoid answering subsequent and more specific questions. The matrix sampling procedure is also said to have the advantage of reducing costs because a short questionnaire requires less interviewing time (Gonzalez and Eltinge, 2007). To the extent that this advantage holds, survey administrators should be able to use matrix sampling to achieve higher response rates with lower costs.

The method poses challenges to those who analyze the data. Joint analysis of data that are not included in all versions of the questionnaire requires strong assumptions about the distribution of the unobserved correlations; for some types of data, this is a severe limitation. In addition, the sampling variance of estimates may be increased without an increase in the overall sample size.

NEW AND EMERGING DATA COLLECTION MODES

Another promising avenue for investigation is tailoring the mode of data collection to the target population. This tailoring of modes has long been a key consideration in the survey design stage. Today, with technological advances and new communications options, survey managers have new and exciting options of employing targeted modes to maximize response and minimize cost in real time through the intelligent use of paradata. In this section, we discuss cell phone options, the use of the Internet, and self-administered modes.

Cell Phone Surveys

The explosive growth in cell phone usage has created challenges for survey managers even as it has opened new possibilities for survey operations. A recent AAPOR task force report on cell phone survey techniques (American Association for Public Opinion Research, 2010a) suggested several strategies for improving response rates for this mode.

Among the report's suggestions are using longer field periods, making advance contact (which is not possible with cell RDD numbers), tailoring the caller ID display, leaving voice messages to encourage cooperation, and preparing well-written introductory scripts that allow for easy tailoring to individual respondents. The introductory contact is especially important in calling cell telephones, for which an advance letter is not usually possible. Offering remuneration for cell phone costs and contingent incentives to try to stimulate cooperation among sample members who might otherwise refuse are two strategies that are often effective, provided that interviewers are well trained on when and how to offer the incentives. In addition, offering a short version of the questionnaire, thus lowering respondent burden, may help, as may offering multiple modes to respond.

In discussing the AAPOR report, its chair, Paul Lavrakas (2011), said that there is a need for research on countering nonresponse. Traditionally all sample members have been approached with a "one-size-fits-all" recruitment method. Although this approach makes practical and operational sense, it fails to take advantage of the computer-assisted environments that support surveys today.

One line of inquiry would be to test matching interviewer and respondent characteristics, including language and dialect, and to examine the impact of those characteristics on participation. Even on the telephone, an interviewer's voice may convey information to a respondent about that interviewer's characteristics. In theory, a respondent will have a greater affinity for a stranger (the interviewer) who is thought to be similar to the respondent. A recent review found no experimental studies on matching interviewers and respondents on social characteristics (Schaeffer et al., 2010). One social characteristic for which the interaction between the sample member's characteristic and that of the interviewer has been examined is race; the available non-experimental studies found no significant effect of race of interviewer on participation (Singer et al., 1983; Merkle and Edelman, 2002).

As researchers pursue means of increasing response, it should be recognized that there are limits as to what efforts can be effective. For example, Brick and Williams (2009) speculated that the increased number of callbacks in telephone surveys may actually increase households' inclination to refuse.

Internet Panel Surveys

Many survey researchers see increased use of the Web as the key to controlling escalating data collection costs in surveys. In the committee's workshop, Reg Baker, chair of the AAPOR panel on online surveys, summarized the results of the AAPOR panel's study of these surveys (American Association for Public Opinion Research, 2010b). The AAPOR task force concluded that probability-based online panels can provide good coverage of the general population (since they provide Internet access to those lacking it), but overall response rates tend to be very low (5 to 15 percent). Nonprobability designs, involving pre-selected respondents ("access panels"), generally ignore coverage error, and they report participation rates to specific survey requests anywhere from less than 1 to 20 percent.

Thus, the reduced costs from the use of online panels come at a price. Most panels use non-probability samples, provide poor coverage, and obtain low rates of participation. These issues with Internet panels have led to the development and publication of international quality standards for access panels, which are becoming a key tool of market, opinion, and social research (ISO 26362). The standard lays out criteria for assessing the quality of access panels and applies to all types of access panels, whether Internet or not. The ISO standard aims to provide international criteria to help compare the results of access panels worldwide (International Standards Organization, 2009).

Self-Administered Modes

Online surveys are one type of self-administered survey, but there are other types as well. Couper (2011a, 2011b) categorizes self-administered modes as fully self-administered or as involving interviewers. Those that are fully self-administered include surveys conducted by mail, Web, and inbound or automated outbound interactive voice response (IVR). Those that are self-administered with interviewer involvement include computer-assisted self-interviewing (CASI), audio computer-assisted self-interviewing (ACASI), recruit-and-switch IVR or telephone audio computer-assisted self-interviewing (T-ACASI), and paper-and-pencil self-administered questionnaire (SAQ).

Couper (2011a) makes the argument that self-administered modes have some measurement advantages and are generally cheaper but do not solve the inferential issues facing surveys (especially coverage and nonresponse). In Couper's view, self-administered modes will increasingly supplement rather than replace interviewer administration. He outlined the opportunities and challenges facing each of the modes.

Fully self-administered modes are less expensive than interviewer-administered modes, and they reduce social desirability effects (that is, respondents providing answers that they believe are more socially acceptable). With mail surveys, the respondent can take time to consider answers, look up records, and consult other household members. Mail surveys have the potential to allow a respondent to reread complex questions, thus reducing the load on working memory. The Web has all the advantages of mail, plus those of computerization.

The mode that is selected for the self-administered questionnaire makes a difference in eliciting survey responses. Kim et al. (2010) examined the nonresponse correlates for self-administered questionnaires using paper-and-pencil personal interview (PAPI) versus those conducted in a computer-assisted personal interview (CAPI) and CASI format. The authors found that CASI not only was associated with lower response rates compared with the other modes but also affected response dynamics. Those age 45 to 64 and blacks and other ethnic groups were more likely to be nonrespondents with CASI.

Fully self-administered modes have disadvantages that can affect the quality of the responses. There is no interviewer available to motivate the sample member or to provide clarifications. IVR surveys likely experience more nonresponse break-offs than other modes (see, for example, Kreuter et al., 2008).

Researchers wanting to use the Web as a principal data collection mode face sampling and coverage issues. There is no general population frame of Internet users nor is there an RDD-like mechanism to generate one. That means, for probability samples, that the frame must come from elsewhere (e.g., RDD, ABS, or traditional area-probability samples). Although Internet penetration is more than 70 percent, considerable disparities exist between those who have Internet access and those who do not and that may bias the estimates for the general population.

Some approaches to Internet surveys restrict inference to the population with access to the Internet (which may be a poor substitute for a general population) or dispense with probability sampling altogether. Non-probability online samples may be based on such techniques as "river sampling" (in which participants are recruited using banner ads, pop-up ads, or similar methods, screened for their demographic characteristics, and assigned to an appropriate survey) and RDS as described above. While these techniques may yield a willing population, they do not result in a representative population and thus cannot yield generalizable inferences. In some panels, the survey researchers have provided equipment for those without Internet access (e.g., the Knowledge Network's KnowledgePanel, the Face-to-Face Recruited Internet Survey Platform, the Measurement and Experimentation in the Social Sciences and Longitudinal Internet Studies for the Social Sciences panels in the Netherlands, and the RAND American Life Panel).

Some surveys address the coverage problem by using a mixed-mode design with mail for non-Internet cases. The Gallup panel is one example of this approach. A research experiment in 2007 tested the effects of various approaches and incentives for improving response in this multimode panel with Internet and mail components (Rao et al., 2010).

MULTIPLE MODES

There has been increased interest over the last two decades in mixed-mode alternatives. The thinking is that if surveys that rely on a single mode have unacceptably low response rates, then combining modes may take advantage of different modes to increase response rates and potentially reduce nonresponse bias. In a presentation to the panel, Mick Couper (2011a) suggested that the research evidence to date is quite mixed and that success may depend on how the modes are mixed and on the evaluation criteria used (e.g., cost, coverage, nonresponse, or measurement error).

Some mixed-mode methods have proven more productive than others, while some may actually increase nonresponse. Research findings have determined that mail-plus-phone designs produce higher response rates than Web-plus-phone designs and that giving respondents a choice of mode is less effective than offering each mode in sequence (Cantor et al., 2009). Couper observed that while mixed modes may reduce errors of nonobservation by improving coverage versus Internet or telephone-only modes or may reduce nonresponse bias relating to literacy relative to mail-only methods, mixing modes may add complications in terms of measurement error (Couper, 2011a). Nonetheless, the mixed-mode approach has gained in popularity over time, particularly for large government-sponsored social science surveys.

Growth of Multiple Mode Surveys

The use of mixed survey modes for conducting surveys has been growing fairly extensively and over a long period.[3] With the continued growth in the use of mixed modes, the methodology has advanced from buzzword (see Dillman and Tarnai, 1988) to widespread usage.

The research interest in various modes has changed over time. Telephone and personal interview modes have played a dominant role in the mix for some time. Mail has increasingly become a part of the mix. The resurgence in the use of mail as a mode has probably been due to the large drop in response rates in telephone studies and the development of near-comprehensive ABS frames, such as the U.S. Postal Service's DSF. More recently, research has focused on the use of mail to induce respondents to use the Internet, which has significant cost savings over interviewer-administered modes (and perhaps over mail self-administered questionnaires) and the additional benefit of more complex instruments being made possible by the Web.

In research focusing on one statewide general public household survey, the 2008 and 2009 Washington Community Survey, Messer and Dillman (2010, 2011) sampled from the DSF and asked respondents in nine and six treatment groups, respectively, to respond by Internet or mail or both. The treatment groups varied the procedures and incentives for the Web–mail implementations. The mail-only groups responded at higher rates than the Web panels, but both achieved higher response rates than might be expected with only an RDD telephone survey. Yet despite this and other research on Web-mail surveys, Messer and Dillman conclude that it "remains unclear as to what procedures are most effective in using the DSF with mail and the Internet survey modes to obtain acceptable levels of non-response" (2010, p. v). (See also Messer and Dillman, 2011.)

Shih and Fan (2008) conducted a meta-analysis of experiments comparing Web and mail response rates in some 39 studies. They observed "a preference of mail survey mode over the Web survey mode, with the mail survey mode response rate being 14 percent higher than the Web-survey mode response rate" (p. 269). However, when sample members were offered both mode options at the same time, there was no significant difference in response rates. This suggested to the researchers that it would be advantageous to offer nonrespondents in one mode a different mode in the follow-up.

Their meta-analysis considered several study features that might have affected response rate differences between modes, including "(1) whether

[3] The literature uses the terms "mixed mode," "multimode," and "multiple modes" interchangeably. In this report, we simply refer to mixed mode.

potential respondents in a comparative study were randomly assigned to receive Web or mail surveys; (2) what type of population was involved; (3) what incentive was provided; (4) whether there was a follow-up reminder for initial nonrespondents; and (5) the year a study was published" (p. 255). They found that two of the study features (population types and follow-up reminders) contributed to the response rate differences between Web and paper surveys. College sample members appeared to be more responsive to Web surveys, while some other sample member types (e.g., medical doctors, school teachers, and general consumers) appeared to prefer traditional mail surveys. Follow-up reminders appeared to be less effective for Web surveys than for mail surveys.

American Community Survey: A Sequential Mixed-Mode Case Study[4]

The most ambitious use of a mixed-mode approach to improve survey response rates is the approach in the American Community Survey (ACS). The ACS is an ongoing survey designed to provide information about small areas. It was developed to replace the long-form survey that was part of the decennial census for many decades. The ACS is conducted on a continuous basis. The data from a given year are released in the fall of the following year. Each month, the ACS questionnaire—similar in content to the census long form—is mailed to 250,000 housing units that have been sampled from the Census Bureau's Master Address File.[5]

The ACS adopted a mixed-mode approach based on extensive research. Three sequential modes were selected for monthly data collection: mail, telephone, and personal visit. For the mail option, the residential housing unit with usable mailing addresses—about 95 percent of each month's sample—are sent a pre-notification letter, followed four days later by a questionnaire booklet. A reminder postcard is sent three days after the questionnaire mailing. Whenever a questionnaire is not returned by mail within three weeks, a second questionnaire is mailed to the address. If there is still no response and if the Census Bureau is able to obtain a telephone number for the address, trained interviewers will conduct telephone follow-up surveys using computer-assisted telephone interviewing (CATI) equipment. Interviewers also follow up on a sample of the following: households at addresses for which no mail or CATI responses are received after two months, households for which the postal service returned the questionnaire because it could not be delivered as addressed, and households for which a questionnaire could not be sent because the address was not in the proper street name and number format. The interviewers visit housing units in

[4]This discussion is based on a presentation by Deborah Griffin (2011).

[5]The monthly sample size was increased in June 2011 to almost 300,000 housing units.

person and collect the ACS data through CAPI (or, in 20 percent of the cases, the follow-up is conducted by telephone).

The pattern of response rates across three modes shows that self-selection does take place. For example, sample members from households in less economically advantaged areas and ethnic enclaves are less likely to respond to the mail surveys than are other households. ACS 2006 data also showed that individuals not in the labor force were more likely than those who were employed to respond to the mail mode, while those with no high school education had low response to the initial mail questionnaire and were more likely to participate by telephone or personal interview.

ACS data show how the sequential mode design improves not only participation across different social groups but also overall response rates. The weighted mail response rate has stayed between 55 and 57 percent in the first five survey rounds. For the same period, by contrast, the weighted telephone response rate dropped from 60.4 to 50 percent, while the weighted personal visit response rate increased from 94.3 to 95.6 percent. The weighted combined-mode response rate was around 98 percent from 2005 to 2009.

Recently, the Census Bureau conducted research on using the Internet as a response mode for the ACS with the goal of reducing costs.[6] Based on favorable results in response rates and data quality, an Internet response option was implemented in mid-December 2012. Most households are sent a letter urging them to respond via the Internet and providing secure sign-on information. Only if they do not respond within two weeks are they sent a paper questionnaire.

Mixed Modes in Panel Studies

Panel studies provide a rich source of data for understanding mode effects because interview modes may change between waves, and the effects of changes in modes can be examined for individual sample members as well as in the aggregate. Longitudinal studies, such as the National Longitudinal Survey of Youth (NLSY), the Panel Study of Income Dynamics (PSID), and the Health and Retirement Study (HRS), commonly show variations in aggregate response rates from wave to wave, but these changes may reflect not only changes in mode but also other changes in field procedures, the aging of the sample or increasing fatigue with participation, and secular changes affecting sample members' propensity to respond.

---

[6]See http://www.census.gov/acs/www/Downloads/library/2012/2012_Matthews_01.pdf; http://www.census.gov/acs/www/Downloads/library/2012/2012_Matthews_01.pdf; and http://www.census.gov/newsroom/releases/archives/american_community_survey_acs/cb12-247.html [January 2013].

Longitudinal studies have increased their use of telephone interviewing in order to contain costs. They have also replaced PAPI interviews with CAPI interviewing. These mode changes do not seem to have affected response rates, likely because respondents in longitudinal studies have already made a commitment to the survey and have already had some experience with the interview process. (Refer back to Tables 1-6 through 1-9 in Chapter 1 for response rate information for the NLSY79, NLSY97, PSID, and HRS.)

One in-depth study of mode effects in a longitudinal study used the Round 11 CAPI experiment data of the NLSY97 to compare the differences in CAPI and PAPI interviews (Baker et al., 1995). The introduction of CAPI reduced branching and skipping errors made by interviewers because the computer program acted as a checking and editing mechanism. The average difference in interview length between the two modes was only 0.9 minutes. A few measures were affected by the switchover. For example, the proportion of respondents who reported that they were paid by the hour was higher in CAPI. On a separate questionnaire, many CAPI survey respondents reported that they were more willing to be forthright in their responses to sensitive questions than they had been in their previous NLSY interviews using paper-and-pencil questionnaires. This result, which has been widely replicated, occurred presumably because there was a greater perception of anonymity when the interviewer entered the answers on the computer screen instead of a form that had the respondent's identifiable information on it.

A question is whether longitudinal surveys can maintain high response rates in the Internet age. One study that evaluated the impact of shifting to the Internet for the HRS concluded that changes in the wording of questions (which often accompany a mode shift) had more of an effect than the change in the interview mode (Van Soest and Kapteyn, 2009a).

The same study investigated the issue of selection effects. Van Soest and Kapteyn (2009a) used a random sample of HRS 2002 and HRS 2004 respondents to investigate the mode effect of Internet surveys on measurement of household assets: checking accounts, savings accounts, stocks, and stock mutual funds. The 2002 and 2004 questionnaires contained questions on Internet access and willingness to participate in an Internet survey in between the biannual surveys. Those who were willing were administered the Internet questionnaires in 2003 and 2006. The authors analyzed these responses along with overlapping items from the 2002 and 2004 core survey questionnaires and found large selection effects. Respondents who used the Internet mode were likely to own more stocks and stock mutual funds than other respondents. HRS Internet 2003 survey respondents not only owned larger amounts of stocks and stock mutual funds; they also had more money in checking and saving accounts relative to respondents in HRS Internet 2006, HRS 2002, and HRS 2004.

More such experiments are required before panel surveys can move to the Internet mode with confidence, as each new mode shift brings its own set of challenges. Such experiments should also look into how quickly respondents of different types are able to learn to respond on the Web, particularly as technological innovations (and the type of data requested) make Web instruments more complex and demanding.

Needed Research

As more researchers turn to mixed-mode designs in an effort to maintain response rates, it is increasingly important to conduct research on mode effects, not only on response rates but also on measurement errors. Furthermore, mode research needs to go beyond simple comparisons that document differences between modes to use of stronger measurement criteria, such as the impact of mode on reliability and validity.

Recommendation 4-3: Research is needed on understanding mode effects, including ways in which mixed-mode designs affect both nonresponse and measurement error and the impact of modes on reliability and validity.

INTERVIEWER EFFECTS

In the personal interview setting, whether face to face or over the telephone, the role of the interviewer in securing participation needs to be considered. The model of survey cooperation proposed by Groves and Couper (1998) is useful in this regard. Their model distinguishes among various factors that could influence survey participation on the basis of whether the factors are under the researcher's control. For example, the social environment and behavior characteristics of household members are not under the researcher's control. In addition, the sample member brings his or her underlying propensity to participate to the encounter with the interviewer (or to avoiding an encounter with the interviewer).

The researcher chooses the design of the survey and selects and trains the interviewers. Ultimately, the interaction between the interviewer and the household member or members often determines the sample member's decision to participate in the survey. Because researchers have some control over the behavior of interviewers through training and monitoring, and because personal encounters are believed to have some inherent persuasive capacity, substantial responsibility for generating high response rates rests with the interviewer. However, interviewer training, previous experience at interviewing, the characteristics of the assignment area, features of the survey design, and socioeconomic characteristics of respondents that affect interviewer expectations all influence the interviewer's behavior during the recruitment process.

To understand how strongly such factors influence survey response and cooperation, researchers have looked into assignment characteristics, observable and unobservable personal characteristics of the interviewer, and the behavior of the interviewer. The following section summarizes the findings and conclusions from the research done in these three areas. (These and other factors are discussed in the review by Schaeffer et al., 2010.)

Assignment Characteristics

Differences in interviewers' success may be due to differences in their assignments. Assignments may not be random in centralized phone facilities, particularly after an initial contact in which some information about the household is obtained, as better interviewers may be assigned to more difficult cases (those with lower probability of success in refusal conversions). In studies that have used random assignment, there is evidence that variability in survey participation rates is influenced by the characteristics both of the assigned cases and of the interviewer.

In the few face-to-face studies with interpenetrated designs that allow the influence of the interviewer on participation to be separated from the influence of the assignment area, it appears that the interviewer contributes at least as much or more to variance in response rates than area does (O'Muircheartaigh and Campanelli, 1998; Schnell and Kreuter, 2005). In addition, some of what appears to be interviewer variance in survey responses may be due to the effects of the interviewer on participation rather than on measurement (West and Olson, 2010).

Observable Personal Characteristics

The impact on participation of an interviewer's personal characteristics, such as gender, race, age, education, and voice, has been analyzed by various researchers. In research on gender, Fowler and Mangione (1990) found that respondents described female interviewers as "friendly," while Morton-Williams (1993) found that respondents perceived female interviewers to be "approachable." Do friendliness and approachability result in higher response rates for female interviewers? Campanelli and O'Muircheartaigh (1999) and Hox and De Leeuw (2002) found a positive effect, with female interviewers obtaining higher response rates. Campanelli and O'Muircheartaigh (1999) used data from an experiment implemented during the second wave of the British Household Panel Survey and found that female interviewers had higher response rates than male interviewers.

Hox and De Leeuw (2002), in a comparison of 32 surveys from nine countries, found female interviewers to have response rates that were, on average, 0.8 percentage point higher than those of male interviewers. Similarly, Groves et al. (2008) found that "less masculine" voices tended to generate higher response rates, although Durrant et al. (2010) found the effect to be restricted to female respondents.

Very few studies have investigated the association between an interviewer's race and participation. In an early study, Singer et al. (1983) looked at the effects of an interviewer's personal characteristics and expectations on response rates in a telephone and personal interview survey. The race of the interviewer did not significantly explain variation in response rates. Merkle and Edelman (2002) found a similar result when examining nonresponse rates in exit polls.

Does the age of the interviewer make a difference? The four studies mentioned above (Singer et al., 1983; Campanelli and O'Muircheartaigh, 1999; Hox and de Leeuw, 2002; Merkle and Edelman, 2002) also included age of the interviewer as one of the controls or independent variables in their regression models. These four studies found that older interviewers were able to elicit higher response rates.

The educational background of the interviewer may play a role in participation rates. Durrant et al. (2010) investigated the effects of characteristics of both the interviewer and household members using a study in which U.K. census data were matched to survey data. They found that when the education levels of interviewers closely matched those of sample persons, higher cooperation rates were observed.

Vocal characteristics of the interviewer may also play a role. Lower refusal rates were found among interviewers rated as speaking more quickly, loudly, distinctly, and in a higher pitch (Oksenberg et al., 1986). Participation may be better predicted by how the voice of the interviewer is perceived than by actual acoustical vocal qualities (Van der Vaart et al., 2006). In telephone interviews, a moderate level of speech disfluency, such as false starts and non-lexical utterances in the flow of otherwise fluent speech, may actually result in higher response rates (Conrad et al., 2010).

Although it is clear from this discussion that interviewer and sample person characteristics do play a role in survey response, it is not apparent that matching those characteristics necessarily results in improved response rates. Davis et al. (2010) observed that there is surprisingly little evidence to indicate whether sociodemographic interviewer–respondent matching improves survey response rates. A recent study that tested whether local or outside interviewers had better response rates suggested that outside interviewers had a better chance of obtaining sensitive information (Sana et al., 2012).

Unobservable Personal Characteristics

Some characteristics of interviewers that are not directly observed by the respondent, such as the interviewer's attitudes and expectations, may play a role in securing a response. Researchers have used such measures as an interviewer's confidence, attitudes about persuasion, belief in confidentiality and the importance of refusal conversions, and an expression of willingness to proceed in the face of obstacles to determine if unobservable characteristics could play a role in securing a response.

In the work cited above, Durrant et al. (2010) found that interviewer confidence and attitudes toward persuading reluctant respondents play an important role in reducing refusal rates. Groves and Couper (1998), however, found that a measure of tailoring derived from contact forms (i.e., a measure of how well the interviewer adapted to household characteristics), was not a significant explanatory variable. Sinibaldi et al. (2009) looked at the interviewer's personality traits and interpersonal skills. They found that extroverted interviewers and more conscientious interviewers were more likely to achieve cooperation from respondents. They also found that interpersonal skills were not predictive of cooperation rates. Therefore, the available literature does not offer a clear picture of the mechanism connecting an interviewer's unobservable characteristics and the survey participation that he or she achieves.

Experience, however, is important. Work done by Campanelli et al. (1997), Groves and Couper (1998), and Sinibaldi et al. (2009) found that experienced interviewers were more successful, as measured by the likelihood of obtaining a completed survey. In a range of designs and across modes, experience was found to relate positively to cooperation, as interviewers with five or more years of experience were better able to overcome negative responses (Durbin and Stuart, 1951; Groves and Fultz, 1985; Couper and Groves, 1992; Groves and Couper, 1998; Pickery and Looseveldt, 1998; Hox and De Leeuw, 2002; Sinibaldi et al., 2009; Durrant et al., 2010). Experience is also related to lower non-contact rates in face-to-face and telephone interviews; two studies suggest that interviewers who succeed at one succeed at the other (O'Muircheartaigh and Campanelli, 1999; Pickery and Looseveldt, 2002).

These results, however, may reflect self-selection of interviewers—experienced interviewers are more likely to have been successful from the outset and therefore more likely to stay in the survey business compared with less experienced interviewers. Finally, it matters how experience is defined. Studies that measured experience as "number of organizations" and "number of surveys" found no relationship or a negative relationship with cooperation rates.

Interactions Between Sample Person and Interviewer

Studies on the factors influencing participation have begun looking more closely at the interaction between the interviewer and the sample person, which has three main phases. The first phase is the interaction during the survey introduction, which takes place for less than a minute on the phone (Oksenberg et al., 1986) and up to five minutes in the case of a face-to-face interview (Groves and Couper, 1994). The second phase is the persuasion attempt by the interviewer if he or she faces reluctance from the householder to participate in a survey. If the householder agrees to participate in the survey, then the third phase of interaction takes place, in which the interviewer elicits responses to survey questions. Research investigating the interaction of the interviewer and householders has looked at all three phases; only the first two phases are relevant for survey nonresponse decisions.

The theory proposed by Groves and Couper (1998) provided a description of two techniques that should be employed by interviewers during the three phases: tailoring and maintaining interaction. *Tailoring* is the technique employed by expert interviewers who customize their interactions with sample persons based on a variety of cues. *Maintaining interaction* is a technique in which interviewers continue engaging respondents in conversation to obtain more information for tailoring and to reduce the likelihood that sample members will refuse to participate in a given turn of talk. The authors stressed the fact that interaction must be maintained for tailoring to occur.

Groves and McGonagle (2001) developed nonresponse aversion training based on these two concepts. They broke the task of the interviewer into four steps: (1) identifying the concern, (2) classifying it, (3) providing an appropriate response, and (4) performing those tasks as quickly as possible. The training improved the response rates of interviewers, especially for those who had lower response rates before the training. Relatedly, Dijkstra and Smit (2002) recorded and analyzed spontaneously occurring persuasion techniques and found that such techniques increased participation.

Survey Introduction

Survey introductions can vary in content (sponsor's name, confidentiality concerns), amount of information (level of detail about topic), and scriptedness. O'Neil et al. (1979) experimentally varied what the interviewer said after a short introduction and found marginal differences in response rates between groups who were administered different sets of introductions. In a telephone survey, Singer et al. (1983) varied the information provided to sampled households on survey content and on the purpose of the interview. This variation did not affect the overall response rate. In another study, scripted introductions were found to generate lower response rates (Morton-Williams, 1993).

Houtkoop-Steenstra and van den Bergh (2000) hypothesized that if interviewers varied their survey introduction style, without altering the content, they could achieve greater cooperation. They looked at response rates in a telephone survey in the Netherlands. Four types of introductions were given. The first was an agenda-based introduction, in which the interviewers formulated their own introductions on the basis of a limited number of catchwords. The other three were standardized introductions of varying length—short, medium, and long. The short version included a greeting and request for participation which were not part of the agenda-based introduction. The medium version included the elements of the short version and the reason for calling. The long version included elements that in theory and sometimes in research increase response rates, such as (a) the information about the length of interview ("The interview will not take long."); (b) the nature of questions ("The questions are simple."); (c) an authority statement mentioning the name of the company ("You may know about us from television."); (d) a statement about the importance of the information ("Your opinion is important."); and (e) a confidentiality statement. The authors did not find any differences among the respondent groups assigned to the standardized introductions, but the agenda-based introduction induced higher response rates.

Saliency

A recent study by Maynard et al. (2010) discusses the leverage–saliency framework outlined in Chapter 1. Interviewers may increase the probability of obtaining a response by emphasizing features of the study or participation with "positive leverage and neutralizing the salience of those with negative leverage" (p. 792). The authors point out that the theory accords with actual practice—interviewers tend to emphasize positive aspects of participating or downplay negative aspects. For example, an interviewer might acknowledge that an interview takes a long time but note that it can be broken into parts. By emphasizing that the leverage a survey attribute has differs across sample persons, leverage–saliency theory calls attention to the importance that interviewers tailor requests to individual sample persons. Interviewers can encourage participation by "observ[ing] idiosyncratic concerns of the householder and customiz[ing] their remarks to those concerns" (Groves et al., 2000, p. 299; see also Couper and Groves, 1992; Groves et al., 1992; Maynard and Schaeffer, 2002).

In a presentation to the panel, Schaeffer pointed out that approaches such as leverage–saliency theory draw attention to the predispositions of the sample member (Schaeffer, 2011). However, the response propensity that the sample member brings to the contact with the interviewer might be modified over the course of the encounter and may affect the leverage that a feature of the survey design has with a respondent. These propensities, and their fluctuations, are difficult to incorporate into practical study designs. However, using conversation analytic techniques, Schaeffer et al. (2013) found that the interactional environment provided by the sample member (encouraging, discouraging, or ambiguous) is a very strong predictor of subsequent participation.

Questions by sample members may provide evidence of their predispositions. Previous studies have identified questions by sample members as predictive of whether the sample member is likely to accept the request to participate. Drawing on interviewers' descriptions of their interaction with sample members, Groves and Couper (1996), for example, concluded that questions indicate cognitive engagement by the sample member and are associated with an increased likelihood of participation in future contacts. A more recent study that selected pairs of sample members matched on propensity to participate refined this finding. Schaeffer et al. (2013) found that wh-type questions (i.e., questions beginning with "wh," such as what, why, when) before the request to participate were associated with decreased odds of participating. On the other hand, questions about the length of interview or wh-type questions after the request to participate were associated with increased odds of participating. The predictive value of sample members' questions is of practical significance because interviewers could be alerted to interpret such questions as a sign that the sample member is positively engaged.

Interviewer Training

In a new survey, all interviewers begin equal, with no knowledge about the survey; even in existing surveys, experienced interviewers may not clearly recall all the relevant facts. Training provides an opportunity for the survey designers to make the factual information sufficiently salient to interviewers that it can shape their engagement with respondents. If interviewers do not understand a given study, it seems likely that they will be less effective in motivating respondents to participate.

Besides making information mentally accessible, training can help in developing interviewers' strategies for interactions with respondents. Demonstrations of various approaches can provide models of effective recruitment behavior. Role playing, particularly when accompanied by coaching, can assist in developing confident introductions and delivery of particular arguments. Role playing can be effective both in helping interviewers cope with the stresses of rejection and in learning how to back off gracefully without completely forestalling future attempts. When a variety of survey materials are available for providing to respondents, training can be effective in determining when different pieces of information might be most appropriate.

Training can also be seen as ongoing throughout the field period. As interviewers learn more about the respondent population, they can interact with more senior staff for coaching. To the extent that information is available about actual performance—through recordings, direct observation, or notes recorded by interviewers as part of their record keeping—such feedback can be more focused.

Recommendation 4-4: Research is needed on the structure and content of interviewer training as well as on the value of continued coaching of interviewers. Where possible, experiments should be done to identify the most effective techniques.

Concluding Remarks on the Role of Interviewers

In summary, interviewers play a valuable role in obtaining survey responses. The survey participation literature summarized above has scrutinized various aspects of that role. But it is important to acknowledge that an interviewer's actions are very much dependent on sampling frame, survey design, survey mode, and interviewer training. Future research studies investigating an interviewer's role in survey participation should provide insights into how to integrate interviewers' efforts with design features. Interviewers can be provided material that will contain information on respondents and records of prior contacts. Interviewer training can explain the importance of participation, how to assure the respondents of confidentiality, how to approach previous refusals, how to diagnose reluctance and respond appropriately, how to make a graceful exit, and various strategies to handle high-priority but low-propensity cases. As for respondents, they can be persuaded through advance letters, survey materials (explaining reasons for conducting the survey or addressing respondents' fears or reservations directly), and (monetary) incentives. Further empirical research into survey participation requires collection of more information on interviewers and behavior of respondents.

INCENTIVES

Singer (2011) spoke to the panel on the use of monetary incentives to counter the trend toward increasing nonresponse in national household surveys. She noted that monetary incentives, especially prepaid incentives, are being employed more often (Singer and Ye, 2013). Her talk summarized research on incentives focusing on findings largely drawn from randomized experiments. She examined the effects of incentives on response quality, sample composition, and response distributions. She noted that incentives have been found to reduce nonresponse rates, primarily by reducing refusals, but that little is known about their effect on nonresponse bias.

There are a number of answers to the question of why people respond to surveys. All theories of survey response emphasize the role of incentives in motivating behavior, though these need not be monetary incentives. Singer noted that results from responses to open-ended questions suggest that there are three main reasons for responding to surveys: altruistic reasons (e.g., wanting to be helpful); egoistic reasons (e.g., monetary incentives); and reasons associated with aspects of the survey (e.g., topic interest, trust in the sponsor). Both theory and observation confirm the importance of incentives for participation in surveys.

Effects on Cooperation

Incentives improve cooperation (Church, 1993; Singer and Ye, 2013). For example, Mann et al. (2008) reported that in a longitudinal study of young adults, parents receiving incentives of $1 or $2 were more likely than those receiving no incentives to provide addresses for their adult children, and children of parents receiving incentives responded more quickly to the survey.

In another study, Holbrook et al. (2008) analyzed 114 RDD surveys between 1996 and 2005 and found, after controlling for other variables, that incentives were significantly associated with higher response rates, with the effect due mainly to a reduction in refusals (with no change in contact rates).

Beydoun et al. (2006) compared the results of unconditional (prepaid) and conditional (promised) incentives on tracing and contact rates in a sample of Iowa postpartum women. The unconditional incentive rates were slightly higher than were the conditional rates, and the highest rates were attained when the incentives were combined.

Effects in Mail Surveys

One meta-analysis by Church (1993) found that prepaid incentives yielded significantly higher response rates to mail surveys than promised or no incentives, that they yielded higher response rates than gifts, and that response rates increased with increasing amounts of money. In another meta-analysis, Edwards et al. (2002) reported similar results.

Effects in Interviewer-Mediated Surveys

A meta-analysis of 39 experiments by Singer et al. (1999) found results for surveys using interviewers that were similar to those in mail surveys, although the effects of incentives were generally smaller. The analysis of 114 RDD surveys by Holbrook et al. (2008) found that surveys offering incentives had significantly higher response rates than those offering no incentives; the effect came mainly from a reduction in refusals. The 2008 analysis by Cantor et al. of 23 RDD experiments found that:

- a prepayment of $1 to $5 increased response rates from 2 to 12 percentage points;
- larger incentives led to higher response rates;
- the effect of incentives has not declined over time, but baseline response rates have dropped substantially;
- incentives at refusal conversion had about the same effect as those sent at initial contact; and
- promised incentives of $5 and $25 did not increase response rates compared to no incentives, but promising larger incentives sometimes did.

These findings are generally consistent with other experiments involving interviewer-mediated surveys, including face-to-face surveys.

Effects in Cell Phone Surveys

Brick et al. (2007) conducted a dual-frame survey including both cell phones and landlines that include an incentive experiment. This study used two promised incentive conditions ($10, $5) and two message conditions (sample members notified of the survey and incentive by text messaging, or not notified). They found that the $10 group had a higher screener response rate than the $5 group as well as a higher cooperation rate. The message had no effect on either screener or survey response rates, and there were no interaction effects.

Incentives in Longitudinal Studies

Longitudinal surveys have special issues, because incentives are usually part of a larger motivational package designed to retain respondents. As in cross-sectional studies, initial survey round incentives have been found to increase response rates, usually by reducing refusals but sometimes by reducing non-contacts (e.g., McGrath, 2006). Some studies suggest that an initial payment may continue to motivate participation in subsequent waves (Singer and Kulka, 2001; McGrath, 2006; Creighton et al., 2007;

Goldenberg et al., 2009). Singer concluded that incentives appear to increase response among those who have previously refused, but not among those who have previously cooperated (Zagorsky and Rhoton, 2008).

In a study by Jaeckle and Lynn (2008) of incentive payments in a U.K. longitudinal study, the researchers found that (1) attrition was significantly reduced by incentives in all waves, (2) the attrition was reduced proportionately among subgroups and so did not reduce attrition bias, (3) the effect of the incentive decreased across waves, (4) incentives increased item nonresponse, but (5) there was a net gain in information.

The NLSY97 has been a rich source of analysis of the effects of incentive payments on participation in a longitudinal survey because, from the beginning, NLSY management has had discretion over the level of incentives to be offered to participants. The amount of the incentive has also been adjusted on an experimental basis. In an early study conducted in NLSY97 Round 4 and extended into Round 5, Datta et al. (2001) found that sample members who were paid $20 had higher participation rates than those paid $10 or $15. However, there were no measurable effects on data quality from the higher level of incentives.

Subsequent NLSY97 experiments found that higher incentives had a particular effect on bringing those who dropped from prior rounds back into participation in later rounds. Pierret et al. (2007) studied the results of the incentive experiments and concluded that incentives moderately increased response rates and had a greater impact on those respondents who did not participate in the previous round relative to those who did participate.

An incentive experiment was conducted as part of the 2000 wave of HRS, in which the incentive amount was increased from $20 to $30 or $50. Rodgers (2011) found an improvement in response rates as the incentive increased. A lowered incentive amount of $40 in subsequent rounds or waves did not result in lowered response rates. He also found a statistically significant decrease in item nonresponse among respondents receiving larger incentives.

Other Findings on Incentive Effects

Two experiments failed to find a role for interviewers in mediating incentive effects. Singer et al. (2000) kept interviewers blind to households' receipt of incentives in one condition but not in another and found that there were no differences in incentive effects between the two conditions. Lynn (2001) randomly offered promised incentives to half of each interviewer's assigned households and then asked interviewers how useful they thought the incentives had been. Interviewers' judgments were almost uniformly negative, but incentives had significant positive effects on completion of a household interview, completion of interviews with individual members, and completion of time use diaries. Nevertheless, as Singer (2011) pointed out to the panel, interviewer expectations may have an independent effect on respondent behavior—for example, Lynn's effects might have been larger had interviewers had a more positive view of incentives. Also, the possibility of contamination in Singer's experiment cannot be entirely ruled out since the same interviewers administered both conditions. It may also be that incentives vary in their effect at different points over the field period of a survey.

An important consideration is the effect of incentives on response quality. Singer and Kulka (2001) found no decline in quality of response to incentives in terms of differences in nonresponse or length of open-ended answers. Since then, the small number of studies (mail, RDD, and face to face) that have examined incentive effects on data quality have, with one exception, found no effects. The exception is Jaeckle and Lynn (2008), who found that incentives increased item nonresponse. Cantor et al. (2008) argued for additional tests that would control for such factors as survey topic, size and type of incentive (e.g., prepaid, promised, refusal conversion), and whether studies are cross-sectional or longitudinal.

Do incentives affect sample composition? Cantor et al. (2008), in their review of 23 RDD studies, concluded that incentives, whether prepaid or promised, have little effect on measures of sample composition. Nevertheless, a number of studies have demonstrated such effects on specific characteristics (see Singer, 2013, pp. 128–129). But specific attempts to use incentives to bring groups into the sample that are less disposed to respond because of lower topic interest have received only qualified support (Groves et al., 2004, 2006). Singer points out that very few studies have considered the sample composition effect of Web survey incentives, and she concluded that more research is clearly needed.

A key question concerns the effect of incentives on the responses that respondents provide. The research findings are mixed. James and Bolstein (1990), Brehm (1994), and Schwarz and Clore (1996) reported results consistent with the mood hypothesis—that incentives boost mood and therefore affect responses—and Curtin et al. (2007) found an interaction between race and receipt of incentives (nonwhites receiving an incentive gave more optimistic answers on the Index of Consumer Confidence). Groves et al. (2004, 2006) reduced nonresponse due to lack of topic interest by offering incentives, and the change in bias due to increased participation of those with less interest was not statistically significant. The possibility that incentives bias responses directly through an effect on attitudes has found no support in the experimental literature, although Dirmaier et al. (2007) specifically tested such a hypothesis. There is no evidence that randomly administered incentives increase bias, but not enough is known about the effect to use them effectively to reduce bias.

Whether incentives have an effect on Internet surveys is still unknown. Singer (2011) pointed out that research in this area is limited. Much of the published experimental research has been done by Göritz (2006), who finds that incentives increase the proportion of invitees starting a survey and the proportion completing it over a no-incentive group. Lotteries are the incentives most often used. Her literature review concluded that specific tests of lotteries against other types of incentives or against no incentives show that lotteries are no more effective in Web surveys than in other kinds of surveys. In most tests, lotteries did not significantly increase response rates over a no-incentive or alternative incentive group.

Incentives are sometimes differential: different amounts are offered primarily to convert refusals, often on the basis that differential incentives are more economical than prepaid incentives and that they are more effective in reducing bias. But there is also a question of fairness; many sample members are likely to consider differential incentives to be unfair. Nonetheless, Singer et al. (1999) found that respondents would be willing to participate in a new survey by the same organization even when told that differential incentives would be paid. When differential incentives are used to reduce bias, they are commonly paired with a small prepaid incentive to all possible participants, which serves to increase the sample size and helps to satisfy the fairness criterion.

Whether or not there are long-term incentive effects is not yet known. Singer (2011) said that there is no evidence of long-term effects thus far, but studies have been done only over short intervals.

Conclusions on Incentives

Singer (2011) drew the following conclusions regarding incentives:

- Incentives increase response rates to surveys in all modes and in cross-sectional as well as panel studies.
- Monetary incentives increase response rates more than gifts do, and prepaid incentives increase them more than promised incentives or lotteries do.
- There is no good evidence for how large an incentive should be. In general, although response rates increase as the size of the incentive increases, they do so at a declining rate. Also, there may be ceiling effects to the extent that people come to expect incentives in all surveys.
- Few studies have evaluated the effect of incentives on the quality of response; most of these have found no effects.

- Few studies have examined the effect of incentives on sample composition and response distributions; most of these have found no effects.
- Effects on sample composition and response distributions have been demonstrated in a small number of studies in which incentives have brought into the sample a larger-than-expected number of members of particular demographic categories or interest groups.
- Incentives have the potential for both increasing and reducing nonresponse bias. They may reduce bias if they can be targeted to groups that might otherwise fail to respond. They may increase bias if they bring into the sample more of those who are already overrepresented. And if they affect all subgroups equally, they will leave the nonresponse bias unchanged.

Recommendation 4-5: Research is needed on the differential effects of incentives offered to respondents (and interviewers) and the extent to which incentives affect nonresponse bias.

PARADATA AND AUXILIARY DATA

Paradata are data about the survey process itself and are collected as part of the survey operation (Couper, 1998). These data may include records of calls, reasons for refusals, responses to incentive offers, and characteristics of the interviewers (Couper and Lyberg, 2005; Bates et al., 2008; Laflamme et al., 2008; Lynn and Nicolaas, 2010; Stoop et al., 2010; Olson, 2013). As discussed in Chapter 3, paradata are used for many purposes: to monitor the status of field collection, to confirm that fieldwork has been carried out according to instructions, to compute response rates, to identify reasons for nonresponse, to implement responsive design strategies, and to adjust for nonresponse bias. They can be in the form of macro paradata, sometimes also termed metadata, which are aggregate statistics about the survey (e.g., response rates, coverage rates, editing rates). Studies have found such aggregate data to be useful for coming to an understanding of the survey information and in the weighting process (Dippo and Sundgren, 2000). Paradata can also be in micro form and include information carried on individual records, such as imputation flags, together with call records and the like.[7]

Sometimes paradata comprise auxiliary information external to the information collected on the survey questionnaire. Auxiliary data "can be thought of as encompassing all information not obtained directly from the interview of a respondent or case-specific informant" (Smith, 2011, p. 389).

[7]However, care should be taken in using call record data because the process that causes such records to be created is often non-neutral, in that their presence or absence reflects a decision to pursue or not pursue a case.

Auxiliary data can be information about the sample frame taken from external sources such as census data by block group, census tract, and other geographic areas (Smith and Kim, 2009). Auxiliary data can also be in the form of observational data about the interview environment. For example, the European Social Survey collects data about the type of dwellings and about neighborhood characteristics such as accumulated litter and signs of vandalism (Stoop et al., 2010). The researchers found that refusals and non-contacts were more likely in areas in which buildings were poorly maintained and where litter abounded. However, these data have not been found useful for nonresponse adjustment purposes (Stoop et al., 2010).

The power of employing paradata and auxiliary data for improving response rates is now becoming recognized. In her presentation to the panel, panel member Kristen Olson observed that, in response to declining response rates, survey practitioners can often introduce design features on the basis of paradata to recruit previously uncontacted or uncooperative sample members into the respondent pool (Olson, 2013). Paradata may also be helpful in tailoring a survey to increase its saliency to sample members. Paradata can be used as well to create nonresponse adjustments, reflecting each respondent's probability of being selected and observed in the respondent pool (see Chapter 2).

There is an increasing recognition that auxiliary data can be used in sample design (i.e., to achieve large samples of rare or hard-to-find groups) or to improve approaches to the interview (Smith, 2011). In this regard, Smith (2011, p. 392) concludes that research is needed on methods for linking databases in order to augment "sample-frame data with information from other databases and sources." Auxiliary data may also be useful to improve imputation techniques (Smith and Kim, 2009).

Olson (2013) observed that few auxiliary variables are available on both respondents and nonrespondents and that there is a dearth of research on predicting survey variables. She concluded that paradata have traditionally been developed as measures of participation, not as survey variables. Paradata have the potential to assist in nonresponse adjustment and may be useful in developing responsive designs. However, the use of paradata requires upfront investment and additional research to demonstrate when the application of paradata is effective and when it is not. The use of paradata also requires the development of standards for its collection.

Recommendation 4-6: Research leading to the development of minimal standards for call records and similar data is needed in order to improve the management of data collection, increase response rates, and reduce nonresponse errors.

RESPONSIVE DESIGN

Paradata make it possible for survey managers to monitor the survey process in real time and to make decisions and alterations in order to improve response rates. This approach to survey design has been termed responsive design (Groves and Heeringa, 2006). As envisioned by Groves and Heeringa (2006, p. 1), responsive designs "pre-identify a set of design features potentially affecting costs and errors of survey statistics, identify a set of indicators of the cost and error properties of those features, monitor those indicators in initial phases of data collection, alter the active features of the survey in subsequent phases based on cost/error tradeoff decision rules, and combine data from separate design phases into a single estimator." Responsive design is a flexible menu of design approaches that can be employed in real time to ameliorate the damage caused by reduced response rates to surveys. The effectiveness of this approach depends critically on the ability to pre-identify variables that can provide basic data on costs and error sources so that survey managers can make rational decisions about the trade-offs between costs and errors.

One particular application of responsive design was described by Laflamme (2011) at the panel's workshop. This application, called responsive collection design (RCD), uses the information available prior to and during data collection to adjust the collection strategy for the remaining in-progress cases. Statistics Canada has conducted two experimental surveys with RCD and control groups for two major CATI surveys: the 2009 Households and Environment Survey, using the 2009 Canadian Community Health Survey sampling frame, and the 2010 Survey of Labour and Income Dynamics (SLID). The SLID 2011 was designed with full RDC techniques in which there was an embedded experiment for the first call. The paradata consisted of the information on the Blaise transaction file (i.e., calls and contact information), interviewer payroll hours, budget and target figures, previous and current collection cycle information, and response propensity model results. These produced indicators that were used to identify when to start RCD.

The results of the Statistics Canada tests indicated that there was a higher overall response rate when RCD was used compared to the previous survey cycle. The responsive design group achieved the same response rate with less effort. On the basis of the evidence, it was concluded that RCD is technically feasible. However, Laflamme (2011) asserted that RCD is not a "magic" solution.

Recommendation 4-7: Research is needed on the theory and practice of responsive design, including its effect on nonresponse bias, informa-

tion requirements for its implementation, types of surveys for which it is most appropriate, and variance implications.

ADMINISTRATIVE RECORDS

Administrative records may be helpful in reducing potential bias due to nonresponse, and they may be helpful in correcting for bias. John Czajka's remarks (2009) addressed the use of administrative records to reduce nonresponse bias. He gave the example of the use of Internal Revenue Service (IRS) tax records by the Census Bureau. The IRS conducts an annual enumeration, and Social Security numbers (SSNs) provide a link to age, sex, race, and Spanish origin data stored in other files, such as the Social Security files. These administrative records yield population coverage estimated at 95 percent of U.S. residents.

The potential of these administrative data led to the development of the Census Bureau's Statistical Administrative Records System (StARS), which combines data from six large federal files: IRS 1040 and 1099, Selective Service, Medicare, Indian Health Service, and Housing and Urban Development's Tenant Rental Assistance Certification System (TRACS). However, administrative records have limitations. For example, IRS records do not include undocumented residents, dependents of non-filers, and non-filers with no reported income. In addition, the unit of observation is not the household, and the reported address may not be residential. The records may also be incomplete. Race is often missing from Social Security files and, when present, may not reflect current definitions.

Although administrative records hold promise for helping to improve survey operations, in Czajka's judgment, it is unlikely that they can be substituted for survey reports. Reasons include that the set of survey items for which there is a high quality administrative records alternative is small and largely limited to federal records; the concepts underlying administrative records may differ from survey concepts (e.g., tax versus survey income); the records may be outdated; and there are severe limitations on the ability to use administrative records because of confidentiality concerns. Little work has been done on informed consent procedures to enable the use of the confidential administrative records.

Yet work has moved forward on ways to use administrative records to make data collection programs more cost-effective. Thus, as part of planning for the 2020 census, the Census Bureau has continued and expanded its acquisition of administrative records and its research on ways to use administrative records with census and survey data (see http://www.census.gov/2010census/pdf/2010_Census_Match_Study_Report.pdf [April 2013]). Other statistical agencies have explored the uses of administrative records to augment survey data, such as matching health survey data with health expenditure claims data (see, e.g., http://www.cdc.gov/nchs/data_access/data_linkage/cms_medicare.htm [April 2013]).

Recommendation 4-8: Research is needed on the availability, quality, and application of administrative records to augment (or replace) survey data collections.

OTHER MEANS OF COLLECTING SOCIAL SCIENCE DATA

The problems with obtaining cooperation with social science surveys do not mean that probability-based sampling should be abandoned. However, it would be useful for the survey community to continue to prepare for a time when current modes are no longer tenable, due to excessive costs and burden. Several emerging methods for gathering social science data are briefly discussed here because they warrant consideration in the development of a research agenda for dealing with the problem of declining response rates.

Non-probability Samples

Non-probability samples are a troubling alternative to traditional probability samples, but, largely owing to their cost and timeliness advantages, they are rapidly growing as a means of gathering data. Much of that growth is associated with the growth of online survey methods. In fact, non-probability samples now account for the largest share of online research, according to the AAPOR Report on Online Panels (American Association for Public Opinion Research, 2010b).

Nonresponse, according to the AAPOR report, is an issue for non-probability samples, just as it is for probability-based samples. Nonresponse in the respondent recruitment phase is likely to be considerable, but since the target population is not known as it is with probability-based samples, it is not easily measured. Relatively little is known about nonresponse in non-probability samples, but nonresponse is not likely to be random and is likely to include the effects of self-selection. For example, the AAPOR report notes that self-selected, non-probability-based online panels are more likely to include white, younger, more active Internet users and those with higher levels of educational attainment than the general population. The report concludes that these surveys offer no foundation on which to draw inferences to the general population.

Internet Scraping Technologies

Another possible approach to the growing problem of survey nonresponse is not to survey at all, but instead to gather information via data mining techniques, essentially mining the Internet to gather information. There are several examples of the use of this technique for the generation of economic and social statistics.

- *MIT Billion Prices Project*. The Billion Prices Project (BPP) is an initiative by economists Roberto Rigobon and Alberto Cavallo to collect prices from hundreds of online retailers around the world on a daily basis (see http://bpp.mit.edu [April 2013]). The project monitors daily price fluctuations of about 5 million items sold by approximately 300 online retailers in more than 70 countries. For the United States, the project collects about 500,000 prices. It has been collecting prices since 2007. The BPP is said to have closely tracked the Consumer Price Index.
- *Google Price Index*. The Google Price Index is a project of the company's chief economist, Hal Varian. Varian uses Google's vast database of Web prices to construct the constantly updated measure of price changes and inflation. Google has not yet decided whether it will publish the price index, and it has not released its methodology, but it has reported that the preliminary index tracks the Consumer Price Index closely.
- *Flu Epidemic Prediction*. For several years, studies have been conducted to detect the onset of U.S. seasonal flu epidemics by extracting patterns of flu-related search terms from the billions of queries stored by Google and Yahoo! Inc. (Butler, 2008). These studies have been taken to provide real-time indicators to complement the CDC reports that are compiled using a combination of data about hospital admissions, laboratory test results, and clinical symptoms. These reports are often weeks old by the time hospitals get them, and so they do not allow frontline health-care workers enough time to prepare for a surge in flu cases. The studies have found that patterns of searches matched official flu surveillance data almost perfectly—and often weeks in advance of these official data.
- *Predicting the Stock Market Using Twitter Feeds*. Recent research has tested whether measurements of collective mood states derived from large-scale Twitter feeds are correlated with the value of the Dow Jones Industrial Average (DJIA) over time. Bollen et al. (2011) analyzed the text content of daily Twitter feeds with two mood-tracking tools and were able to predict the daily up and down changes in the closing values of the DJIA with an accuracy of 87.6 percent.

Admittedly, these examples of the use of Internet data mining to produce socioeconomic indicators are currently developmental, but their apparent success in replicating data collected administratively or through costly surveys suggests that further development and testing could be warranted. However, caution is in order. Availability of many types of information on the Web is a matter of choice; thus, the data available on a given topic may not fully reflect the underlying range of information, raising the possibility of unknown biases. Often, there is no built-in constraint to make measures conceptually compatible. In addition, there is no guarantee that biases would necessarily be stationary in a meaningful sense over time, which would compromise the validity of trend analysis.

Recommendation 4-9: Research is needed to determine the capability of information gathered by mining the Internet to augment (or replace) official survey statistics.

5

Research Agenda

This report documents the rich set of extant findings about the causes, consequences, and remedies for the general decline in survey response throughout the developed world. This decline represents a growing threat to the quality of social science data in the United States and elsewhere. This report also identifies a number of gaps in that knowledge and promising paths to advance the state of the science and develop more effective remedies.

In the various sections of this report, the panel has recommended research on a long list of topics. These topics fall into three broad categories: (1) research that would deepen our understanding of the phenomenon of nonresponse and the causes of the decline in response rates over the past few decades; (2) research aimed at clarifying the consequences of nonresponse; and (3) research designed to improve our tools for boosting response rates or more effectively compensating for the effects of nonresponse statistically. The panel thus supports a series of research programs and projects that are brought together here. We believe that, together, these topics constitute a comprehensive and multifaceted research agenda.

The panel is aware that, in these times of increasingly limited human and financial resources for the social science survey community, a research agenda must reflect both costs and benefits. Where possible, priorities have been suggested in the report itself.

The panel does not attempt to assign responsibility for these research items among the various players that make up the social science survey community and the research and academic institutions that support that community. Much of the path-breaking basic research, which often does not require significant investment of resources for testing and development, will come from academic and other research centers. Large data collection and analytical organizations in the private sector and in government would be responsible for conducting the research that requires new data collection, such as research on interviewer and mode effects. Organizations that provide a platform for integrating the research work generated in these various venues, such as the AAPOR, the American Statistical Association, the International Statistical Institute, and, within the federal government, the Federal Committee on Statistical Methodology, would also have important roles to play.

RESEARCH ON THE PROBLEM

The first set of research topics would help further define the problem, develop appropriate measures, and deepen understanding of the scope, causes, and extent of the problem:

- Research to identify the person-level and societal variables responsible for the downward trend in response rates. These variables could include changes in technology, in communication patterns, and in methods of collecting survey data.
- Research on people's general attitudes toward surveys and on whether these have changed over time.
- Research about why people take part in surveys and the factors that motivate them to participate.
- Research on the factors affecting contact and cooperation rates.
- Research on the nature (mode of contact, content) of the contacts that people receive over the course of a survey based on data captured in the survey process.
- Research on the overall level of burden from survey requests and on the role of that burden in the decision to participate in a specific survey.

In considering burden, it is important to conduct basic research on the dimensions of response burden and how they should be operationalized. It would be useful to consider factors (e.g., time, cognitive difficulty, or invasiveness) that may determine how potential respondents assess the burden involved in taking part in a survey. These research paths should lead to more practical consideration of how interviewers, advance letters, or other explanatory or motivational material could effectively alter perceptions about the likely burden of a survey.

RESEARCH ON CONSEQUENCES

The second set of topics concerns statistical and other tools for understanding and mitigating the consequences of nonresponse:

- Research on the cost implications of nonresponse and on how to capture cost data in a standardized way.
- Research on the relationship between nonresponse rates and nonresponse bias and on the variables that determine when such a relationship is likely.
- Research to examine unit and item nonresponse bias and to develop models of the relationship between nonresponse rates and bias.
- Research on the theoretical limits of what nonresponse adjustments can achieve, given low correlations with survey variables, and on the effects of measurement errors, missing data, and other problems with the covariates.
- Research on the impact of nonresponse reduction on other error sources, such as measurement error.
- Research to quantify the role that nonresponse error plays as a component of total survey error.
- Research on the differential effects of incentives offered to respondents (and interviewers) and the extent to which incentives affect nonresponse bias.

The panel notes that research activities designed to expand knowledge of the relationship between response rates and nonresponse bias are likely to assist in the development of a much needed theory of nonresponse bias. A more comprehensive theory not only would further our understanding of the relationship between response rates and nonresponse bias but also would aid in the development of adjustment techniques to deal with bias under different circumstances.

RESEARCH ON COPING

The third set of research topics concerns methods for coping with nonresponse:

- Research to establish empirically the cost–error trade-offs in the use of incentives and other tools to reduce nonresponse.
- Research on and development of new indicators for the impact of nonresponse, including application of the alternative indicators to real surveys to determine how well the indicators work.

- Research on understanding mode effects, including the impact of mode on reliability and validity.
- Research leading to the development of minimal standards for call records and similar data in order to improve the management of data collection, increase response rates, and reduce nonresponse errors.
- Research on the structure and content of interviewer training as well as on the value of continued coaching of interviewers. Where possible, experiments should be supported to identify the most effective techniques.
- Research to improve the modeling of responses as well as to improve the methods to determine whether data are missing at random.
- Research on the use of auxiliary data for weighting adjustments, including whether weighting can make estimates worse (i.e., increase bias) and whether traditional weighting approaches inflate the variance of the estimates.
- Research to assist in understanding the impact of adjustment procedures on estimates other than means, proportions, and totals.
- Research on the impact that reduction of survey nonresponse would have on other error sources, such as measurement error.
- Research on how to best make a switch from the telephone survey mode (and frame) to mail, including how to ensure that the right person completes a mail survey.
- Research on the theory and practice of responsive design, including its effects on nonresponse bias, information requirements for its implementation, types of surveys for which it is most appropriate, and variance implications.

RESEARCH ON ALTERNATIVES

Finally, the panel recognizes the need to explore alternatives to traditional survey data collection. There are increasing suggestions that administrative data and Internet "scraping" can produce data that could substitute for surveys. The panel suggests that further research is needed to ascertain the quality of data gleaned from these sources, and makes two final recommendations:

- Research into the availability, quality, and application of administrative records to augment (or replace) survey data collections.
- Research to determine the capability of information gathered by mining the Internet to augment (or replace) official survey statistics.

References and Selected Bibliography

Abraham, K., Helms, S., and Presser, S.
2009 *How Social Processes Distort Measurement: The Impact of Survey Nonresponse on Estimates of Volunteer Work in the United States*. NBER Working Paper 14076. Available: http://www.nber.org/papers/w14076 [January 2013].

Amaya, A., and Ward, C.
2011 Cost Efficiency: Which Design is Cheapest? Unpublished paper presented at the Joint Statistical Meetings, July 31, Miami, FL.

American Academy of Political and Social Science
2013 The nonresponse challenge to surveys and statistics. D. Massey and R. Tourangeau, Eds., Special issue of *The ANNALS of the American Academy of Political and Social Science* 645(1).

American Association for Public Opinion Research
2010a New Considerations for Survey Researchers When Planning and Conducting RDD Telephone Surveys in the U.S. with Respondents Reached via Cell Phone Numbers. AAPOR Cell Phone Task Force. Available: http://www.aapor.org/AM/Template.cfm?Section=Cell_Phone_Task_Force_Report&Template=/CM/ContentDisplay.cfm&ContentID=3189 [January 2013].
2010b AAPOR Report on Online Panels. Available: http://www.aapor.org/AM/Template.cfm?Section=AAPOR_Committee_and_Task_Force_Reports&Template=/CM/ContentDisplay.cfm&ContentID=2223 [January 2013].
2011 *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys*. Seventh Edition. Available: http://www.aapor.org/AM/Template.cfm?Section=Standard_Definitions2&Template=/CM/ContentDisplay.cfm&ContentID=3156 [January 2013].

Andridge, R.H., and Little, R.J.
2009 The use of sample weights in hot deck imputation. *Journal of Official Statistics* 25(1):21–36.

Baker, R., Bradburn, N.M., and Johnson, R.A.
1995 Computer-assisted personal interviewing: An experimental evaluation of data quality and cost. *Journal of Official Statistics* 11(4):413–431.

Ballard-LeFauve, K., Barro, M., and Battaglia, M.P.
2003 Provider Characteristics and Response Rates in the National Immunization Survey. Presentation at the 37th Annual National Immunization Conference, March 17–20, Chicago, IL.

Bates, N., Doyle, P., and Winters, F.
2000 Survey Nonresponse: New Definitions and Measurement Methods. Paper presented at the Federal Committee on Statistical Methodology Statistical Policy Seminar hosted by the Council of Professional Associations on Federal Statistics, November 8–9.

Bates, N., Dahlhamer, J., and Singer, E.
2008 Privacy concerns, too busy or just not interested: Using doorstep concerns to predict survey nonresponse. *Journal of Official Statistics* 24(4):591–612.

Bates, N., Dahlhamer, J., Phipps, P., Safir, A., and Tan, L.
2010 Assessing contact history paradata quality across several federal surveys. Pp. 91–105 in *Proceedings of the Section on Survey Research Methods*. Alexandria, VA: American Statistical Association.

Battaglia, M.P., Ezzati-Rice, T.M., Hoaglin, D.C., Loft, J.D., and Maes, E.
1997 *Response Rates in a Survey That Collects Childhood Vaccination Information from Households and Providers*. Alexandria, VA: American Statistical Association.

Battaglia, M.P., Kahare, M., Frankel, M.R., Muray, M.C., Buckley, P., and Peritz, S.
2008 Response rates: How have they changed and where are they headed? Pp. 529–560 in J. Lepkowski, C. Tucker, M. Brick, E.D. De Leeuw, L. Japec, P.J. Lavarakas, M.W. Link, and R.L. Sangster, Eds., *Advances in Telephone Survey Methodology*. New York: Wiley.

Berry, S.H.
2011 Non-Response in Social Network Surveys. Presentation to the Workshop of the Panel on a Research Agenda for the Future of Social Science Data Collection, April 29, National Research Council, Washington, DC.

Beydoun, H., Saftlas, A., Harland, K., and Triche, E.
2006 Combining conditional and unconditional recruitment incentives could facilitate telephone tracing in surveys of postpartum women. *Journal of Clinical Epidemiology* 59(7):732–738.

Billiet, J., Philippens, M., Fitzgerald, R., and Stoop, I.
2007 Estimation of nonresponse bias in the European Social Survey: Using information from reluctant respondents. *Journal of Official Statistics* 23(2):135–162.

Blom, A., and Kreuter, F.
2011 Preface. *Journal of Official Statistics* 27(2):151–152.

Blumberg, S., Luke, J., Ganesh, N., Davern, M., and Boudreaux, M.
2012 Wireless substitution: State-level estimates from the National Health Interview Survey, 2010–2011. *National Health Statistics Reports* 61. Available: http://www.cdc.gov/nchs/data/nhsr/nhsr061.pdf [January 2013].

Bogen, K.
1996 The effect of questionnaire length on response rates—A review of the literature. Pp. 1020–1025 in *Proceedings of the Section on Survey Research Methods*. Alexandria, VA: American Statistical Association.

Bollen, J., Mao, H., and Zeng, X.
2011 Twitter mood predicts the stock market. *Journal of Computational Science* 2(1):1–8.

Bose, J.
2001 Nonresponse bias analysis at the National Center for Education Statistics. In *Proceedings of Statistics Canada Symposium, Achieving Data Quality in a Statistical Agency: A Methodological Perspective*. Available: http://www.fcsm.gov/committees/ihsng/StatsCan2_JB.pdf [January 2013].

Bradburn, N.
1978 Respondent burden. Pp. 35–40 in *Proceedings of the Section on Survey Research Methods*. Alexandria, VA: American Statistical Association. Available: http://www.amstat.org/sections/srms/proceedings/papers/1978_007.pdf [January 2013].
1992 A response to the nonresponse problem. 1992 AAPOR Presidential Address. *Public Opinion Quarterly* 56(3):391–397.
Bradburn, N.M., Frankel, M.R., Baker, R.P., and Pergamit, M.R.
1991 *A Comparison of Computer-Assisted Personal Interviews (CAPI) with Paper-and-Pencil Interviews (PAPI) in the National Longitudinal Survey of Youth*. NLS Discussion Paper No. 92-02. Washington, DC: U.S. Department of Labor, Bureau of Labor Statistics.
Branden, L., Gritz, R.M., and Pergamit, M.R.
1995 *The Effect of Interview Length on Attrition in the National Longitudinal Survey of Youth*. Working Paper NLS 95-28. Washington, DC: U.S. Department of Labor, Bureau of Labor Statistics.
Brehm, J.
1993 *The Phantom Respondents: Opinion Surveys and Political Representation*. Ann Arbor: University of Michigan Press.
1994 Stubbing our toes for a foot in the door? Prior contact, incentives and survey response improving response to surveys. *International Public Opinion Research* 6(1):45–64.
Brick, J.M., and Jones, M.E.
2008 Propensity to respond and nonresponse bias. *Metron* LXVI:51–73.
Brick, J.M., Edwards, W.S., Cervantes, I.F., and Lee, S.
2008 Sampling and Weighting Cell Phone Surveys to Supplement RDD Surveys. Paper presented at the 63rd annual conference of the American Association for Public Opinion Research, New Orleans.
Brick, J.M., and Williams, D.
2013 Explaining rising nonresponse rates in cross-sectional surveys. *The ANNALS of the American Academy of Political and Social Science* 645:36–59.
Brick, M.
2011 Nonresponse Weighting Adjustments. Presentation to the Workshop of the Panel on a Research Agenda for the Future of Social Science Data Collection, February 18, National Research Council, Washington, DC.
Brick, M., Collins, M., and Chandler, K.
1997 *An Overview of Response Rates in the National Household Education Survey: 1991, 1993, 1995, and 1996*. NCES 97-948. Washington, DC: U.S. Department of Education, National Center for Education Statistics.
2000 *NHES: 99 Methodology Report*. NCES 2000-078. Washington, DC: U.S. Department of Education, National Center for Education Statistics.
Brick, M., Brick, P., Dipko, S., Presser, S., Tucker, C., and Yuan, Y.
2007 Cell phone survey feasibility in the U.S.: Sampling and calling cell numbers versus landline numbers. *Public Opinion Quarterly* 71(1):23–39.
Bureau of Labor Statistics
2001a *NLS Mature Women Cohort Users Guide 2001*. Available: http://www.bls.gov/nls/mwguide/nlsmwusg.htm [February 2012].
2001b *Young Women User's Guide*. Available: http://www.bls.gov/nls/ywguide/nlsywusg.htm [February 2012].
2002 *Overview of the NLSY79 Child/Young Adult Sample and NLSY79 User's Guide*. Available: http://www.bls.gov/nls/y79cyaguide/nlsy79cusg.htm [February 2012].

2005 *NLS Handbook, 2005*. Available: http://www.bls.gov/nls/handbook/nlshndbk.htm [May 2013].
2006 *NLS Original Cohorts: Older and Young Men*. Available: http://www.bls.gov/nls/oldyoungmen.htm [February 2012].

Butler, D.
2008 Web data predict flu. *Nature* 456:287–288. Available: http://nature.com/news/2008/081119/full/456287a.html [January 2013].

Callegaro, M., and DiSogra, C.
2008 Computing response metrics for online panel. *Public Opinion Quarterly* 72(5): 1008–1032.

Campanelli, P., Sturgis, P., and Purdon, S.
1997 *Can You Hear Me Knocking?: An Investigation into the Impact of Interviewers on Survey Response Rates*. London, England: The Survey Methods Centre at SCPR.

Campanelli, P., and O'Muircheartaigh, C.
1999 Interviewers, interviewer continuity, and panel survey nonresponse. *Quality and Quantity* 33:59–76.

Cantor, D., O'Hare, B., and O'Connor, K.
2008 The use of monetary incentives to reduce non-response in random digit dial telephone surveys. Pp. 471–498 in J. Lepkowski, C. Tucker, J.M. Brick, E.D. De Leeuw, L. Japec, P.J. Lavrakas, M.W. Link, and R.L. Sangster, Eds., *Advances in Telephone Survey Methodology*. New York: Wiley.

Cantor, D., Coa, K., Crystal-Mansour, T., Davis, T., Dipko, S., and Sigman, R.
2009 *Health Information National Trends Survey (HINTS) 2007: Final Report*. Westat. Available: http://hints.cancer.gov/docs/HINTS2007FinalReport.pdf [January 2013].

Casas-Cordero, C.
2010 Disentangling Neighborhood Effects on Survey Participation. Paper presented at the 2010 Annual Meeting of the American Association for Public Opinion Research, May 14, Chicago, IL.

Cellini, S.R., Mckernan, S.-M., and Ratcliffe, C.
2008 The dynamics of poverty in the united states: A review of data, methods and findings. *Journal of Policy Analysis and Management* 27(3):577–605.

Centers for Disease Control and Prevention
2011 *Behavioral Risk Factor Surveillance System 2010 Summary Data Quality Report*. Version 1. Available: ftp://ftp.cdc.gov/pub/data/brfss/2010_summary_data_quality_report.pdf [January 2013].
2012 Methodologic changes in the Behavioral Risk Factor Surveillance System in 2011 and potential effects on prevalence estimates. *Morbidity and Mortality Weekly Report* 61(22):410–413. Available: http://www.cdc.gov/mmwr/preview/mmwrhtml/mm6122a3.htm?s_cid=mm6122a3_w [January 2013].

Centers for Disease Control and Prevention, National Center for Immunization and Respiratory Diseases, and National Center for Health Statistics
2011 *A User's Guide for the 2010 Public-Use Data File*. Available: ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Dataset_Documentation/NIS/NISPUF10_DUG.pdf [May 2013].

Cheshire, H., Ofstedal, M.B., Scholes, S., and Schroder, M.
2011 A comparison of response rates in the English Longitudinal Study of Aging and the Health and Retirement Study. *Longitudinal and Life Course Studies* 2(2). Available: http://www.llcsjournal.org/index.php/llcs/article/view/118 [April 2013].

Cho, M.J., Eltinge, J.L., and Steinberg, B.P.
2004 *Temporal Patterns of Survey Response Rates and Reporting Rates in the U.S. Consumer Expenditure Interview and Other Panel Surveys*. BLS Survey Methodology Paper. Washington, DC: U.S. Department of Labor, Bureau of Labor Statistics.

Church, A.H.
1993 Estimating effect of incentives on mail survey response rates: A meta-analysis. *Public Opinion Quarterly* 57(1):62–79.

Cobben, F., and Schouten, B.
2007 *An Empirical Evaluation of R-Indicators*. Discussion Paper 08006. Voorburg/Heerlen: Statistics Netherlands. Available: http://www.risq-project.eu/papers/cobben-schouten-2008-a.pdf [January 2013].

Conrad, F., Broome, J., Benki, J., Groves, R.M., Kreuter, F., and Vannette, D.
2010 To agree or not to agree: Impact of interviewer speech on survey participation decisions. *Proceedings of the Section on Survey Research Methods*. Alexandria, VA: American Statistical Association. Available: http://www.amstat.org/sections/srms/proceedings/y2010/Files/400119.pdf [January 2013].

Couper, M.P.
1998 Measuring survey quality in a CASIC environment. Pp. 41–49 in *Proceedings of the Section on Survey Research Methods*. Alexandria, VA: American Statistical Association.
2011a Self-Administered Modes. Presentation at the Workshop of the Panel on a Research Agenda for the Future of Social Science Data Collection, February 17, National Research Council, Washington, DC.
2011b The future of modes of data collection. *Public Opinion Quarterly* 75(5):889–908. Available: http://poq.oxfordjournals.org/content/75/5/889.full.pdf+html [January 2013].

Couper, M.P., and Groves, R.M.
1992 The role of the interviewer in survey participation. *Survey Methodology* 18(2):263–277.

Creighton, K.P., King, K.E., and Martin, E.A.
2007 The use of monetary incentives in Census Bureau Longitudinal Surveys. Working Paper 32 in *Survey Methodology Research Report Series No. 2007-2*. Washington, DC: U.S. Census Bureau.

Curtin, R., Presser, S., and Singer, E.
2000 The effects of response rate changes on the index of consumer sentiment. *Public Opinion Quarterly* 64:413–428.
2005 Changes in telephone survey nonresponse over the past quarter century. *Public Opinion Quarterly* 69(1):87–98.
2007 Incentives in random digit dial telephone surveys: A replication and extension. *Journal of Official Statistics* 23(1):91–105.

Czajka, J.L.
2009 Administrative Records: Can They Be Used to Reduce Response Bias? Presentation at Planning Meeting on a Research Agenda for the Future of Social Science Data Collection, December 14, Committee on National Statistics, National Research Council, Washington, DC.

Datta, A.R., Horrigan, M.W., and Walker, J.R.
2001 Evaluation of a Monetary Incentive Payment Experiment in the National Longitudinal Survey of Youth. Paper presented at the 2001 Federal Committee on Statistical Methodology Conference. Available: http://www.fcsm.gov/01papers/Horrigan.pdf [January 2013].

Davis, R.E., Couper, M.P., Janz, N.K., Caldwell, C.H., and Resnkicow, K.
2010 Interviewer effects in public health surveys. *Health Education Research* 25(1):14–26. Available: http://her.oxfordjournals.org/content/25/1/14.full [January 2013].
De Leeuw, E.D.
2005 To mix or not to mix data collection modes in surveys. *Journal of Official Statistics* 21(2):233–255.
De Leeuw, E.D., and De Heer, W.
2002 Trends in household survey nonresponse: A longitudinal and international comparison. Pp. 41–54 in R.M. Groves, D.A. Dillman, J.L. Eltinge, and R.J.A. Little, Eds., *Survey Nonresponse*. New York: Wiley.
Dijkstra, W., and Smit, J.H.
2002 Persuading reluctant recipients in telephone surveys. Pp. 121–134 in R.M. Groves, D.A. Dillman, J.L. Eltinge, and R.J.A. Little, Eds., *Survey Nonresponse*. New York: Wiley.
Dillman, D.A.
1978 *Mail and Telephone Surveys: The Total Design Method*. New York: Wiley.
1991 The design and administration of a mail survey. *Annual Review of Sociology* 17:225–249.
1999 *Mail and Internet Surveys—The Tailored Design Method*. New York: Wiley.
Dillman, D.A., and Tarnai, J.
1988 Administrative issues in mixed-mode surveys. Pp. 509–528 in R.M. Groves, P. Biemer, L. Lyberg, J. Massey, W.L. Nicholls, and J. Waksberg, Eds., *Telephone Survey Methodology*. New York: Wiley.
Dillman, D.A., Smythe, J.D., and Christian, L.M.
2007 *Internet, Mail, and Mixed-Mode Surveys: The Tailored Design Method*. Third Edition. New York: Wiley.
Dillman, D.A., Phelps, G., Tortora, R., Swift, K., Kohrell, J., Berck, J., and Messer, B.L.
2008 Response rate and measurement differences in mixed-mode surveys using mail, telephone, interactive voice response (IVR), and the Internet. *Social Science Research* 38:1–18.
Dippo, C., and Sundgren, B.
2000 *The Role of Metadata in Statistics*. Second International Conference Establishment Surveys. Available: http://www.amstat.org/meetings/ices/2000/proceedings/S57.pdf [January 2013].
Dirmaier, J., Harfst, T., Koch, U., and Schulz, H.
2007 Incentives increased return sales but did not influence partial nonresponse or treatment outcome in a randomized trial. *Journal of Clinical Epidemiology* 60(12):1263–1270.
Dixon, J.
2011 Nonresponse in BLS Surveys. Presentation to the Workshop of the Panel on a Research Agenda for the Future of Social Science Data Collection, April 28, National Research Council, Washington, DC.
2012 *Total Survey Error in the American Time Use Survey.* Available: http://www.fcsm.gov/12papers/Dixon_2012FCSM_X-C.pdf [January 2013].
Durbin, J., and Stuart, A.
1951 Inversions of rank correlation coefficients. *Journal of the Royal Statistical Society Series B* 13:303–309.
Durrant, G.B., Groves, R.M., Staetsky, L., and Steele, F.
2010 Interviewer attitudes and behaviors on refusal in household surveys. *Public Opinion Quarterly* 74(1):1–36.

Edwards, P., Roberts, I., Clarke, M., DiGuisepi, C., Paratap, S., Wentz, R., and Kwan, I.
2002 Increasing response rates to postal questionnaires: Systematic review. *British Medical Journal* 324:1183.

Eltinge, J.L., and Yansaneh, I.S.
1997 Diagnostics for formation of nonresponse adjustment cells, with an application to income nonresponse in the U.S. Consumer Expenditure Survey. *Survey Methodology* 23:33–40.

Fay, R., Bates N., and Moore, J.
1991 *Lower Mail Response in the 1990 Census: A Preliminary Interpretation.* U.S. Bureau of the Census. Research Report Series. Survey Methodology #2010-13. Available: http://www.census.gov/srd/papers/pdf/rsm2010-13.pdf [January 2013].

Fowler, F.J., and Mangione, T.W.
1990 *Standardized Survey Interviewing: Minimizing Interviewer-Related Error*. Newbury Park, CA: Sage.

Fox, R.J., Crask, M.R., and Kim, J.
1988 Mail survey response rate: A meta-analysis of selected techniques for inducing response. *Public Opinion Quarterly* 52:467–491.

Fricker, S., and Tourangeau, R.
2010 Examining the relationship between nonresponse propensity and data quality in two national household surveys. *Public Opinion Quarterly* 74(5)935–955.

Goldenberg, K., McGrath, D., and Tan, L.
2009 The effects of incentives on the Consumer Expenditure Survey. Pp. 5985–5999 in *Proceedings of the Section on Survey Research Methods*. Alexandria, VA: American Statistical Association.

Gonzalez, J.M., and Eltinge, J.L.
2007 *Multiple Matrix Sampling: A Review.* Washington, DC: Bureau of Labor Statistics, Office of Survey Methods Research.
2009 *Adaptive Matrix Sampling for the Consumer Expenditure Quarterly Interview Survey.* Washington, DC: Bureau of Labor Statistics, Office of Survey Methods Research.

Göritz, A.
2006 Incentives in web surveys: Methodological issues and a review. *International Journal of Internet Science* 1(1):58–70.

Goyder, J.
1982 Further evidence on factors affecting response rates to mailed questionnaires. *American Sociological Review* 47:550–553.
1987 *The Silent Minority: Nonrespondents on Sample Surveys*. Cambridge, MA: Polity Press.

Goyder, J., Boyer, L., and Martinelli, G.
2006 Integrating exchange and heuristic theories of survey nonresponse. *Bulletin of Sociological Methodology* 92:28–44. Available: http://bms.revues.org/index554.html [January 2013].

Griffin, D.
2011 Mixed Mode Surveys: The American Community Survey Experience. Presentation to the Workshop of the Panel on a Research Agenda for the Future of Social Science Data Collection, April 29, National Research Council, Washington, DC.

Groves, R.M.
1989 *Survey Errors and Survey Costs*. Hoboken, NJ: Wiley Interscience.
2004 *Survey Errors and Survey Costs*. New York: Wiley.

2006 Nonresponse rates and nonresponse error in household surveys. *Public Opinion Quarterly* 70:646–675.

Groves, R.M., and Fultz, N.H.
1985 Gender effects among telephone interviews in a survey of economic attitudes. *Sociological Methods and Research* 14:31–52.

Groves, R.M., Cialdini, R.B., and Couper, M.P.
1992 Understanding the decision to participate in a survey. *Public Opinion Quarterly* 56:475–495.

Groves, R.M., and Couper, M.P.
1994 *Householders and Interviewers: The Anatomy of Pre-interview Interactions*. U.S. Census Bureau Survey Methodology Program Working Paper Series, N. 11. Washington, DC: U.S. Census Bureau.
1996 Contact-level influences on cooperation in face-to-face surveys. *Journal of Official Statistics* 12(1):63–83.
1998 *Nonresponse in Household Interview Surveys*. New York: Wiley.

Groves, R.M., Singer, E., and Corning, E.
2000 Leverage-saliency theory of survey participation: Description and an illustration. *Public Opinion Quarterly* 64:299–308.

Groves, R.M., and McGonagle, K.A.
2001 A theory-guided interviewer training protocol regarding survey participation. *Journal of Official Statistics* 17(2):249–265.

Groves, R.M., Dillman, D.A., Eltinge, J.L., and Little, R.J.A.
2002 *Survey Nonresponse*. New York: Wiley.

Groves, R.M., Presser, S., and Dipko, S.
2004 The role of topic interest in survey participation decisions. *Public Opinion Quarterly* 68(1):2–31. Available: http://poq.oxfordjournals.org/content/68/1/2.full [January 2013].

Groves, R.M., and Heeringa, S.
2006 Responsive design for household surveys: Tools for actively controlling survey costs and errors. *Journal of the Royal Statistical Society, Series A* 169(3):439–457.

Groves, R.M., Couper, M., Presser, S., Singer, E., Tourangeau, R., Acosta, G., and Nelson, L.
2006 Experiments in producing nonresponse bias. *Public Opinion Quarterly* 70(5): 720–736.

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

Groves, R.M., O'Hare, B., Gould-Smith, D., and Benki, J.
2008 Telephone interviewer voice characteristics and the survey participation decision. Pp. 385–400 in J. Lepkowski, C. Tucker, M. Brick, E.D. De Leeuw, L. Japec, P.J. Lavarakas, M.W. Link, and R.L. Sangster, Eds., *Advances in Telephone Survey Methodology*. New York: Wiley.

Hagedorn, M., Montaquila, J., Vaden-Kiernan, N., Kim, K., and Chapman, C.
2004 *National Household Education Surveys of 2003: Data File User's Manual, Volume III, Adult Education for Work-Related Reasons Survey*. Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Hagedorn, M., Montaquila, J., Carver, P., O'Donnell, K., and Chapman, C.
2006 *National Household Education Surveys Program of 2005: Public-Use Data File User's Manual, Volume I, Study Overview and Methodology*. NCES 2006-078. Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Hagedorn, M., Roth, S. B., Carver, P., Van de Kerckhove, W., and Smith, S.
2009 *National Household Education Surveys Program of 2007: Methodology Report*. NCES 2009-029. Washington, DC: U.S. Department of Education, National Center for Education Statistics. Available: http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2009047 [January 2013].

Hansen, M.H., and Hurwitz, W.N.
1946 The problem of non-response in sample surveys. *Journal of the American Statistical Association* 41:517–529.

Harris-Kojetin, B., and Tucker, C.
1999 Exploring the relation of economic and political conditions with refusal rates to a government survey. *Journal of Official Statistics* 15(2):167–184.

Hazelwood, L., Mach, T., and Wolken, J.
2007 *Alternative Methods of Unit Nonresponse Weighting Adjustments: An Application from the 2003 Survey of Small Business Finances*. Finance and Economics Discussion Series 2001-10. Washington, DC: Federal Reserve Board.

Heberlein, T.A., and Baumgartner, R.
1978 Factors affecting response rates of mailed questionnaires: A quantitative analysis of the published literature. *American Sociological Review* 43:446–462.

Heckathorn, D.D.
2011 New Metrics for Non-Response. Presentation to the Workshop of the Panel on a Research Agenda for the Future of Social Science Data Collection, February 17, National Research Council, Washington, DC.

Heckman, J.
1979 Sample selection bias as a specification error. *Econometrica* 47(1):153–161.

Hedlin, D., Dale, T., Haraldsen, G., and Jones, J.
2005 Developing Methods for Assessing Perceived Response Burden. Statistical Office of the European Union. Available: http://epp.eurostat.ec.europa.eu/portal/page/portal/quality/documents/DEVELOPING%20METHODS%20FOR%20ASSESSING%20PERCEIVED%20RESPONSE%20BURD.pdf [January 2013].

Heyneman, S.P.
2000 From the party/state to multi-ethnic democracy: Education and social cohesion in the Europe and Central Asia region. *Educational Evaluation and Policy Analysis* 22(2):173–191.

Hill, M.S.
1999 *The Panel Study of Income Dynamics. A User's Guide*. Volume 2. Newbury Park, CA: Sage.

Holbrook, A., Krosnick, J., and Pfent, A.
2008 The causes and consequences of response rates in surveys by the news media and government contractor survey research firms. Pp. 499–528 in J. Lepkowski, C. Tucker, M. Brick, E.D. De Leeuw, L. Japec, P.J. Lavarakas, M.W. Link, and R.L. Sangster, Eds., *Advances in Telephone Survey Methodology*. New York: Wiley.

Homans, G.C.
1958 Social behavior as exchange. *American Journal of Sociology* 63:597–606.

Houtkoop-Steenstra, H., and van den Bergh, H.
2000 Effects of introductions in large-scale telephone survey interviews. *Sociological Methods and Research* 28:281–300.

Hox, J., and De Leeuw, E.
2002 The influence of interviewers' attitude and behavior on household survey nonresponse: an international comparison. Pp. 103–120 in R.M. Groves, D.A. Dillman, J.L. Eltinge, and R.J.A. Little, Eds., *Survey Nonresponse*. New York: Wiley.

Iannacchione, V.G., Staab, J.M., and Redden, D.T.
2003 Evaluating the use of residential mailing addresses in a metropolitan household survey. *Public Opinion Quarterly* 67:202–210.
Iguchi, M., Ober, A., Berry, S., Fain, T., Heckathorn, D., Gorbach, P., Heimer, R., Kozlov, A., Ouellete, L., Shoptaw, S., and Zule, W.
2009 Simultaneous recruitment of drug users and men who have sex with men in the United States and Russia using respondent-driven sampling: Sampling methods and implications. *Journal of Urban Health* 86(Suppl 1):5–31.
International Standards Organization
2009 ISO Standard for Access Panels Aims to Increase Effectiveness of Market, Opinion, and Social Research. Available: http://www.iso.org/iso/home/news_index/news_archive/news.htm?refid=Ref1211 [January 2013].
Jaeckle, A., and Lynn, P.
2008 Respondent incentives in a multi-mode panel survey: Cumulative effects on nonresponse and bias. *Survey Methodology* 34(1):105–117.
James, J.M., and Bolstein, R.
1990 The effect of monetary incentives and follow-up mailings on the response rate and response quality in mail surveys. *Public Opinion Quarterly* 5(3):346–361.
Johnson, A.E., Botman, S.L., and Basiotis, P.
1994 Nonresponse in federal demographic surveys: 1981–1993. Pp. 983–988 in *Proceedings of the Section on Survey Research Methods*. Alexandria, VA: American Statistical Association.
Kalton, G., and Flores-Cervantes, I.
2003 Weighting methods. *Journal of Official Statistics* 19(2):81–97.
Keeter, S., Miller, C., Kohut, A., Groves, R.M., and Presser, S.
2000 Consequences of reducing nonresponse in a telephone survey. *Public Opinion Quarterly* 64:125–48.
Keeter, S., Kennedy, C., Dimock, M., Best, J., and Craighill, P.
2006 Gauging the impact of growing nonresponse on estimates from a national RDD telephone survey. *Public Opinion Quarterly* 70(5):759–779.
Kennickell, A.
2007 Look and Listen, But Don't Stop: Interviewers and Data Quality in the 2007 SCF. Paper prepared for the 2007 Joint Statistical Meetings, May 11, Salt Lake City, UT.
2010 Try, Try Again: Response and Nonresponse in the 2009 SCF Panel. Paper prepared for the 2010 Joint Statistical Meetings, Vancouver, Canada.
Khare, M., Mohadjer, L.K., Ezzati-Rice, T.M., and Waksberg, J.
1994 An evaluation of nonresponse bias in NHANES III (1988–1991). In *Proceedings of the Section on Survey Research Methods*. Alexandria, VA: American Statistical Association. Available: http://www.amstat.org/sections/srms/proceedings/papers/1994_165.pdf [January 2013].
Kim, J., Kang, J., Kim, S., Smith, T.W., Son, J., and Berktold, J.
2010 Comparison between self-administered questionnaire and computer-assisted self-interview for supplemental survey response. *Field Methods* 22(1):57–69.
King, J.
1978 Nonresponse Bias in Household Expenditure Surveys. Paper presented at the 9th Annual Workshop on Nonresponse, Bled, Slovenia.
Kott, P., and Chang, T.
2010 Using calibration weighting to adjust for non-ignorable unit nonresponse. *Journal of the American Statistical Association* 105(491):1265–1275.

Kreuter, F., Presser, S., and Tourangeau, R.
2008 Social desirability bias in CATI, IVR, and Web surveys: The effects of mode and question sensitivity. *Public Opinion Quarterly* 72:847–865.

Kreuter, F., and Kohler, U.
2009 Analyzing contact sequences in call record data: Potential and limitations of sequence indicators for nonresponse adjustments in the European Social Survey. *Journal of Official Statistics* 25:203–226.

Kreuter, F., and Casas-Cordero, C.
2010 *Paradata*. Working Paper Series des Rates für Sozial-und Wirtschaftsdaten 136. Available: https://www.econstor.eu/dspace/bitstream/10419/43631/1/62386357X.pdf [January 2013].

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

Laflamme, F.
2011 Experiences in Implementing and Monitoring Responsive Collection Design (RCD) for CATI Surveys. Presentation to the Workshop of the Panel on a Research Agenda for the Future of Social Science Data Collection, April 28, National Research Council, Washington, DC.

Landman, C.
2009 Internal U.S. Census Bureau Memorandum: Noninterview Rates for Selected Major Demographic Household Surveys, October 20.

Lansky, A., Drake, A., Wejnert, C., Pham, H., Cribbin, M., and Heckathorn, D.
2012 Assessing the assumptions of respondent-driven sampling in the National HIV Behavioral Surveillance System among injecting drug users. *The Open AIDS Journal* 6(Suppl 1:M3):77–82.

Lavrakas, P.J.
2011 Nonresponse Issues in U.S. Cell Phone and Landline Telephone Surveys. Presentation to the Workshop of the Panel on a Research Agenda for the Future of Social Science Data Collection, February 17, National Research Council, Washington, DC.

Lessler, J.T., and Kalsbeek, W.D.
1992 *Nonsampling Error in Surveys*. New York: Wiley.

Letki, N.
2006 Investigating the roots of civic morality: Trust, social capital, and institutional performance. *Political Behavior* 28(4):305–325.

Lin, I.F., and Schaeffer, N.C.
1995 Using survey participants to estimate the impact of nonparticipation. *Public Opinion Quarterly* 59:236–258.

Link, M.W.
2010 Address-Based Sampling: What Do We Know So Far? Presentation slides from webinar sponsored by the American Statistical Association (Section on Survey Research Methods) and the American Association of Public Opinion Research, November 30. Available: http://www.amstat.org/sections/SRMS/AddressBasedSampling11-29-2010.pdf [January 2013].

Link, M., Battaglia, M., Frankel, M., Osborn, I., and Mokdad, A.
2008 Comparison of address-based sampling (ABS) versus random-digit dialing (RDD) for general population surveys. *Public Opinion Quarterly* 72:6–27.

Little, R.J.A.
1986 Survey nonresponse adjustments for estimates of means. *International Statistical Review* 54(2):139–157.

1993 Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association* 86(1):125–134.

Little, R.J.A., and Vartivarian, S.
2005 Does weighting for nonresponse increase the variance of survey means? *Survey Methodology* 31(2):161–168.

Lorant, V., Demarest, S., Miermans, P.-J., and Von Oven, H.
2007 Survey error in measuring socio-economic risk factors of health status: A comparison of a survey and a census. *International Journal of Epidemiology* 36(6):1292–1299.

Lynn, P.
2001 The impact of incentives on response rates to personal interview surveys: Role and perceptions of interviewers. *International Journal of Public Opinion Research* 13(3):326–336.
2008 The problem of nonresponse. Pp. 35–55 in E. De Leeuw, J. Hox, and D. Dillman, Eds., *International Handbook of Survey Methodology*. New York: Taylor and Francis Group.
2009 *Methodology of Longitudinal Surveys*. Wiley Series on Survey Methodology. West Sussex, England: Wiley.

Lynn, P., and Nicolaas, G.
2010 Making good use of survey paradata. *Survey Practice* 3(April). Available: http://surveypractice.wordpress.com/2010/04/14/making-good-use-of-survey-paradata [January 2013].

Mann, S., Lynn, D., and Peterson, A.
2008 The "downstream" effect of token prepaid cash incentives to parents on their young adult children's survey participation. *Public Opinion Quarterly* 72(3):487–501.

Maynard, D.W., and Schaeffer, N.C.
2002 Refusal conversion and tailoring. Pp. 219–239 in D.W. Maynard, H. Houtkoop-Steenstra, N.C. Schaeffer, and J. van der Zouwen, Eds., *Standardization and Tacit Knowledge: Interaction and Practice in the Survey Interview.* New York: Wiley.

Maynard, D.W., Freese, J., and Schaeffer, N.C.
2010 Calling for participation: Requests, blocking moves, and rational (inter)action in survey introductions. *American Sociological Review* 75(5):791–814.

McCulloch, S.K., Kreuter, F., and Calvano, S.
2010 Interviewer Observed vs. Reported Respondent Gender: Implications for Measurement Error. Paper presented at the Annual Meeting of the American Association for Public Opinion Research, May 14, Chicago, IL.

McGonagle, K., and Schoeni, R.
2006 *The Panel Study of Income Dynamics: Overview and Summary of Scientific Contributions after Nearly 40 Years*. PSID Technical Series Paper #06–01. Ann Arbor: University of Michigan.

McGrath, D.E.
2006 An incentive experiment in the U.S. Consumer Expenditure Quarterly Survey. Pp. 3411–3418 in *Proceedings of the Section on Survey Research Methods*. Alexandria, VA: American Statistical Association.

Mensch, B., and Kandel, D.
1988 Underreporting of substance use in a national longitudinal youth cohort: Individual and interviewer effect. *Public Opinion Quarterly* 52(1):100–124.

Merkle, D., Edelman, M., Dykeman, K., and Brogan, C.
1998 An Experimental Study of Ways to Increase Exit Poll Response Rates and Reduce Survey Error. Paper presented at the Annual Conference of the American Association for Public Opinion Research, May 14–17, St. Louis, MO.

Merkle, D., and Edelman, M.
2002 Nonresponse in exit polls: A comprehensive analysis. Pp. 243–258 in R.M. Groves, D.A. Dillman, J.L. Eltinge, and R.J.A. Little, Eds., *Survey Nonresponse*. New York: Wiley.

Messer, B.L., and Dillman, D.A.
2010 Using Address Based Sampling to Survey the General Public by Mail vs. "Web plus Mail." Social and Economic Sciences Research Center Technical Report, No. 10-13. Available: http://www.sesrc.wsu.edu/dillman/papers/2010/Messer%20Dillman%20WCSTechReport.pdf [January 2013].
2011 Surveying the general public over the Internet using address-based sampling and mail contact procedures. *Public Opinion Quarterly* 75(3):429–457.

Meyers, S., and Oliver, J.
1978 Privacy and hostility toward government as reasons for nonresponse in the National Medical Care Expenditure Survey. Pp. 509–513 in *Proceedings of the Section on Social Statistics*. Alexandria, VA: American Statistical Association.

Mokdad, A.H.
2009 The Behavioral Risk Factors Surveillance System: Past, present and future. *Annual Review of Public Health* 30:43–54.

Montaquila, J.M., Hsu, V., Brick, J.M., English, N., and O'Muircheartaigh, C.
2009 A comparative evaluation of traditional listing vs. address-based sampling frames: Matching with field investigation of discrepancies. In *Proceedings of the Section on Survey Research Methods*. Alexandria, VA: American Statistical Association.

Montaquila, J., and Brick, M.
2012 Transitioning from RDD to ABS with Mail as the Primary Mode. Paper prepared for the Section on Survey Research Methods, Joint Statistical Meetings, August, American Statistical Association, Alexandria, VA.

Morton-Williams, J.
1993 *Interviewer Approaches*. Aldershot, England: Dartmouth.

Munger, G.F., and Loyd, B.H.
1988 The use of multiple matrix sampling for survey research. *Journal of Experimental Education* 56(4):187–191.

National Center for Health Statistics
2012 2011 National Health Interview Survey Public Data Release, NHIS Survey Description, Appendix I, Table 1B. Unconditional Response Rates, NHIS 1979–2011. Available: ftp://ftp.cdc.gov/pub/Health_statistics/NCHS/Dataset_Documentation/NHIS/2011/srvydesc.pdf [January 2013].

National Institute on Aging
2011a Health and Retirement Study Mailout Experiment. Health and Retirement Study Project Staff, Washington DC.
2011b Sample Sizes and Response Rates, Health and Retirement Study Project Staff, Washington, DC.

National Research Council
1983 *Incomplete Data in Sample Surveys*. Volumes 1–3. Panel on Incomplete Data, Committee on National Statistics. Commission on Behavioral and Social Sciences and Education. New York: Academic Press.
2011 *The Future of Federal Household Surveys: Summary of a Workshop*. K. Marton and J.C. Korberg, Rapporteurs. Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

Nolin, M.J., Montaquila, J., Nicchitta, P., Collins Hagedorn, M., and Chapman, C.
2004 *National Household Education Surveys Program: 2001: Methodology Report.* NCES 2005-071. Washington, DC: U.S. Department of Education, National Center for Education Statistics. Available: http://nces.ed.gov/pubs2005/2005071.pdf [January 2013].

NORC
2011 General Social Surveys, 1972–2010. Cumulative Codebook. Appendix A. Available: http://publicdata.norc.org/GSS/DOCUMENTS/BOOK/GSS_Codebook.pdf [October 2012].

Office of Management and Budget
2001 *Measuring and Reporting Sources of Error in Surveys.* Subcommittee on Measuring and Reporting the Quality of Survey Data. Statistical Policy Working Paper 31. Available: http://www.fcsm.gov/01papers/SPWP31_final.pdf [January 2013].
2006 Guidance on Agency Survey and Statistical Information Collections—Questions and Answers When Designing Surveys for Information Collections. Available: http://www.whitehouse.gov/omb/inforeg/pmc_survey_guidance_2006.pdf [January 2013].

Ofstedal, M. B., Fisher, G.G., and Herzog, A.R.
2005 Documentation of Cognitive Functioning Measures in the Health and Retirement Study. HRS Documentation Report DR-006. Available: http://hrsonline.isr.umich.edu/sitedocs/userg/dr-006.pdf [January 2013].

Oksenberg, L., Coleman, L., and Cannell, C.
1986 Interviewers' voices and refusal rates in telephone surveys. *Public Opinion Quarterly* 50(1):97–111.

Olson, K.
2013 Paradata for nonresponse adjustment. *The ANNALS of the American Academy of Political and Social Science* 645:142–170.

O'Muircheartaigh, C., and Campanelli, P.
1998 The relative impact of interviewer effects and sample design effects on survey precision, *Journal of the Royal Statistical Society*, *Series A* 161(1):63–77.

O'Neil, M., Groves, R.M., and Cannell, C.C.
1979 Telephone interview introductions and refusal rates: Experiments in increasing respondent cooperation. Pp. 252–255 in *Proceedings of the Section on Survey Research Methods*. Alexandria, VA: American Statistical Association.

Peytchev, A.
2009 Consequences of Survey Nonresponse. Presentation at the planning meeting on the Future of Social Science Data Collection, December 14, National Research Council, Washington, DC.
2013 Consequences of survey nonresponse. *The ANNALS of the American Academy of Political and Social Science* 645:88–111.

Peytchev, A., Riley, S., Rosen, J., Murphy, J., and Lindblad, M.
2010 Reduction of nonresponse bias in surveys through case prioritization. *Survey Research Methods* 4(1):21–29.

Pickery, J., and Looseveldt, G.
1998 The impact of interviewer and respondent characteristics on number of "no opinion" answers. *Quality and Quantity* 32:31–45.

Pierret, C., Aughinbaugh, A., Datta, A., and Gladden, T.
2007 Event History Data: Lessons from the NSLY97. Presented at the Conference on Event History Calendar Method (organized by University of Michigan Panel Study on Income Dynamics and U.S. Census Bureau), December 5–6, Washington, DC.

Piliavin, J., and Callero, P.
1991 *Giving Blood: The Development of an Altruistic Identity*. Baltimore, MD: Johns Hopkins University Press.

Presser, S., and McCulloch, S.
2011 The growth of survey research in the United States: Government-sponsored surveys, 1984–2004. *Social Science Research* 40:1019–1024.

Putnam, R.
1995 Bowling alone: America's declining social capital. *Journal of Democracy* 6(1):65–78.
2001 *Bowling Alone: The Collapse and Revival of American Community*. New York: Simon and Schuster.

Rao, K., Kaminska, O., and McCutcheon, A.
2010 Recruiting probability samples for a multi-mode research panel with Internet and mail components. *Public Opinion Quarterly* 74(1):68–84.

Roberts, C.
2005 *Expert Workshop on Mixed Mode Data Collection in Comparative Social Surveys, 16th September 2005* [Summary of a Workshop]. Swindon, England: Economic and Social Research Council. Available: http://eprints.ncrm.ac.uk/406/1/FinalReport.pdf [June 2013].

Rodgers, W.L.
2011 Effects of increasing the incentive size in a longitudinal study. *Journal of Official Statistics* 27(2):279–299.

Rosenbaum, P.R., and Rubin, D.B.
1983 The central role of the propensity score in the observational studies for causal effects. *Biometrika* 70:41–55.

Roth, S.B., Montaquila, J.M., and Chapman, C.
2006 *Nonresponse Bias in the 2005 National Household Education Surveys Program.* Washington, DC: U.S. Department of Education, National Center for Education Statistics. Available: http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2007016 [January 2013].

Rubin, D.B.
1977 Formalizing subjective notions about the effect of non-respondents in sample surveys. *Journal of the American Statistical Association* 72:538–543.

Russell Sage Foundation
2010 *Research Agenda for the Future of Social Science Research*. Available: http://www.russellsage.org/awarded-project/research-agenda-future-social-science-data-collection [January 2013].

Sana, M., Strecklov, G., and Weinreb, A.
2012 Local or Outsider Interviewer? An Experimental Evaluation. Paper prepared for the Annual Meeting of the Population Association of America, San Francisco, CA. Available: http://paa2012.princeton.edu/papers/122313 [January 2013].

Särndal, C.
2007 The calibration approach in survey theory and practice. *Survey Methodology* 33(2):99–109.
2011 The 2010 Morris Hansen Lecture dealing with survey nonresponse in data collection, in estimation. *Journal of Official Statistics* 27(1):1–21.

Schaeffer, N.C.
2011 The Effects of Interviewers on Survey Participation. Presentation to the Workshop of the Panel on a Research Agenda for the Future of Social Science Data Collection, April 28, National Research Council, Washington, DC.

Schaeffer, N.C., Dykema, J., and Maynard, D.
2010 Interviewers and interviewing. Pp. 437–470 in P.V. Marsden and J.D. Wright, Eds., *Handbook of Survey Research, Second Edition*. Bingley, England: Emerald Group.

Schaeffer, N.C., Garbarski, D., Freese, J., and Maynard, D.W.
2013 An interactional model of the call for participation in the survey interview: Actions and reactions in the survey recruitment call. *Public Opinion Quarterly* 77(1):323–351.
Schnell, R., and Kreuter, F.
2005 Separating interviewer and sampling-point effects. *Journal of Official Statistics* 21(3):389–410.
Schouten, B., Cobben, F., and Bethlehem, J.
2009 Indicators for the representativeness of survey response. *Survey Methodology* 35(1):101–113. Available: http://www.risq-project.eu/papers/schouten-cobben-bethlehem-2009.pdf [January 2013].
Schwarz, N., and Clore, G.L.
1996 Feelings and phenomenal experiences. Pp. 385–407 in T. Higgins and A. Kruglanski, Eds., *Social Psychology: Handbook of Basic Principles*. New York: Guilford Press.
Sharp, L., and Frankel, J.
1983 Respondent burden: A test of some common assumptions. *Public Opinion Quarterly* 47(1):36–53.
Shettle, C., Guenther, P., Kaspryzk, D., and Gonzalez, M.
1994 Investigating nonresponse in federal surveys. Pp. 972–976 in *Proceedings of the Section on Survey Research Methods*. Alexandria, VA: American Statistical Association.
Shih, T.H., and Fan, X.
2008 Comparing response rates from web and mail surveys: A meta-analysis. *Field Methods* 20:249–271.
Shoemaker, D.M.
1973 *Principles and Procedures of Multiple Matrix Sampling*. Cambridge, MA: Ballinger.
Singer, E.
2000 The Use of Incentives to Reduce Nonresponse in Household Surveys. Survey Methodology Program. Working Paper #51, Institute for Social Research, University of Michigan.
2006 Introduction: Nonresponse bias in household surveys. *Public Opinion Quarterly* 70(5):637–645.
2011 The Use and Effects of Incentives in Surveys. Presentation to the Workshop of the Panel on a Research Agenda for the Future of Social Science Data Collection, February 18, National Research Council, Washington, DC.
Singer, E., Frankel, M., and Glassman, M.
1983 The effect of interviewer characteristics and expectations on response. *Public Opinion Quarterly* 47:68–83.
Singer, E., Gebler, N., Raghunathan, T., Van Hoewyk, J., and McGonagle, K.
1999 The effect of incentives on interviewer-mediated surveys. *Journal of Official Statistics* 15(2):217–230.
Singer, E., Van Hoewyk, J., and Maher, M.P.
2000 Experiments with incentives in telephone surveys. *Public Opinion Quarterly* 64:171–188.
Singer, E., and Kulka, R.A.
2001 Paying respondents for survey participation. Pp. 10–28 in National Research Council, *Studies of Welfare Populations: Data Collection and Research Issues*. Panel on Data and Methods for Measuring the Effects of Changes in Social Welfare Programs, M. Ver Ploeg, R.A. Moffitt, and C.F. Citro, Eds. Division of Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.

Singer, E., and Ye, C.
2013 The use and effects of incentives in surveys. *The ANNALS of the American Academy of Political and Social Science* 645(1):112–141.

Sinibaldi, J., Jäckle, A., Tipping, S., and Lynn, P.
2009 Interviewer characteristics, their doorstep behaviour, and survey co-operation. Pp. 5955–5969 in *Proceedings of the Section on Survey Research Methods*. Alexandria, VA: American Statistical Association.

Smith, T.
1995 Trends in nonresponse rates. *International Journal of Public Opinion Research* 7(2):157–171.
2011 The report of the International Workshop on Using Multi-level Data from Sample Frames, Auxiliary Databases, Paradata and Related Sources to Detect and Adjust for Nonresponse Bias in Surveys. *International Journal of Public Opinion Research* 23(3):389–402.

Smith, T., and Kim, A.
2009 *An Assessment of the Multi-level Integrated Database Approach*. GSS Methodological Report No. 116. Available: http://publicdata.norc.org:41000/gss/documents/mtrt/mr%20116.pdf [April 2013].

Steeh, C.G.
1981 Trends in nonresponse rates, 1952–1979. *Public Opinion Quarterly* 45:40–57.

Steeh, C., Kirgis, N., Cannon, B., and DeWitt, J.
2001 Are they really as bad as they seem? Nonresponse rates at the end of the twentieth century. *Journal of Official Statistics* 17:227–247.

Stoop, I.A.L.
2005 *The Hunt for the Last Respondent: Nonresponse in Sample Surveys*. The Hague: Social and Cultural Planning Office of the Netherlands.

Stoop, I.A.L., Billiet, J., Koch, A., and Fitzgerald, R.
2010 *Improving Survey Response: Lessons Learned from the European Social Survey*. New York: Wiley.

Thompson, B., Kyrillidou, M., and Cook, C.
2009 Item sampling in service quality assessment surveys to improve response rates and reduce respondent burden. *Performance Measurement and Metrics* 10(1):6–16.

Tourangeau, R.
2003 Recurring Surveys: Issues and Opportunities. A Report to the National Science Foundation Based on a Workshop Held on March 28–29. Available: http://www.nsf.gov/sbe/ses/mms/nsf04_211a.pdf [January 2013].

Van de Kerckhove, W., Montaquilla, J.M., Carver, P.R., and Brick, M.
2009 *An Evaluation of Bias in the 2007 National Household Education Survey Program*. Washington, DC: U.S. Department of Education, National Center for Education Statistics. Available: http://nces.ed.gov/pubs2009/2009029.pdf [January 2013].

van der Grijn, F., Schouten, B., and Cobben, F.
2006 Balancing Representativity, Costs and Response Rates in a Call Scheduling Strategy. Paper presented at the 17th International Workshop on Household Survey Nonresponse, August 28–30, Omaha, NE.

Van der Vaart, W., Ongena, Y., Hoogendoorn, A., and Dijkstra, W.
2006 Do interviewers' voice characteristics influence cooperation rates in telephone surveys? *International Journal of Public Opinion Research* 18(4):488–499.

Van Soest, A., and Kapteyn, A.
2009 *Mode and Context Effects in Measuring Household Assets*. University College of Dublin Geary Institute Discussion Paper Series, Working Paper 200949. Available: http://www.ucd.ie/geary/static/publications/workingpapers/gearywp200949.pdf [January 2013].

Wagner, J.R.
2008 Adaptive Survey Design to Reduce Nonresponse Bias. Ph.D. dissertation in survey methodology, Ann Arbor, University of Michigan.
2011 Alternative Measures for the Risk of Nonresponse Bias. Presentation to the Workshop of the Panel on a Research Agenda for the Future of Social Science Data Collection, February 18, National Research Council, Washington, DC.

West, B.T., and Olson, K.
2010 How much of interviewer variance is really nonresponse error variance? *Public Opinion Quarterly* 74(5):1004–1026.

Yang, U.M., and Wang, Y.
2008 Weighting class versus propensity model approaches to nonresponse adjustment: The SDR experience. In *Proceedings of the Section on Survey Research Methods*. Alexandria, VA: American Statistical Association. Available: http://www.amstat.org/sections/srms/proceedings/y2008/Files/301544.pdf [January 2013].

Zagorsky, J.L., and Rhoton, P.
2008 Effects of promised monetary incentives on attrition in a long-term panel survey. *Public Opinion Quarterly* 72(3):502–513.

Zell, E.R., Ezzati-Rice, T.M., Hoaglin, D.C., and Massey, J.T.
1995 Adjusting for response bias on vaccination status in a telephone survey. Pp. 684–689 in *Proceedings of the Section on Survey Research Methods*. Alexandria, VA: American Statistical Association.

Zuckerberg, A.
2010 *Redesigning the National Household Education Survey*. Washington, DC: U.S. Department of Education, National Center for Education Statistics. Available: http://www.bls.gov/cex/aaporsrvyredesign2010zuckerb1.pdf [January 2013].

Acronyms and Abbreviations

AAPOR	American Association for Public Opinion Research
ABS	address-based sampling
ACASI	audio computer-assisted self-interviewing
ACS	American Community Survey
AHEAD	Asset and Health Dynamics Among the Oldest Old Survey
AHS	American Housing Survey
ATUS	American Time Use Survey
BLS	Bureau of Labor Statistics
BRFSS	Behavioral Risk Factor Surveillance System
CAPI	computer-assisted personal interviewing
CARI	computer-assisted recorded interviewing
CASI	computer-assisted self-interviewing
CASRO	Council of American Survey Research Organizations
CDC	Centers for Disease Control and Prevention
CE	Consumer Expenditure Survey
CED	Consumer Expenditure Diary
CEQ	Consumer Expenditure Quarterly Interview Survey
CHI	contact history instrument
CNSTAT	Committee on National Statistics
CODA	Children of the Depression Age cohort
CPS	Current Population Survey
DSF	U.S. Postal Service Delivery Sequence File

EBB	Early Baby Boomer Survey
FCSM	Federal Committee on Statistical Methodology
FMI	fraction of missing information
GSS	General Social Survey
HINTS	Health Information National Trends Survey
HRS	Health and Retirement Study
IRS	Internal Revenue Service
IVR	interactive voice response
LST	leverage–saliency theory
MAR	missing at random
MCAR	missing completely at random
MSA	metropolitan statistical area
NCVS	National Crime Victimization Survey
NHES	National Household Education Survey
NHIS	National Health Interview Survey
NIS	National Immunization Survey
NLS	National Longitudinal Survey
NLSY	National Longitudinal Survey of Youth
NMCES	National Medical Care Expenditure Survey
NRC	National Research Council
OMB	Office of Management and Budget
PAPI	paper-and-pencil interviewing
PSID	Panel Study of Income Dynamics
PSU	primary sampling unit
RCD	responsive collection design
RDD	random digit dialing
RDS	respondent-driven sampling
SAQ	self-administered questionnaire
SCA	Survey of Consumer Attitudes
SCF	Survey of Consumer Finances
SIPP	Survey of Income and Program Participation
SLID	Survey of Labour and Income Dynamics

SSN	Social Security number
StARS	Statistical Administrative Records System
T-ACASI	telephone audio computer-assisted self-interviewing
TDM	total design method
TPOPS	Telephone Point of Purchase Survey
WB	War Baby cohort

Appendix A

Nonresponse Research in Federal Statistical Agencies

Although the panel considered the issue of nonresponse in surveys in both the private and the public sector and in both the United States and abroad, we placed more emphasis on U.S.-public-sector–sponsored surveys primarily because, with a few important exceptions, the largest, most consistent, and most costly survey operations in social science fields are conducted by and for the U.S. federal government.

In its two workshops, the panel heard from survey methodologists from five U.S. federal statistical agencies who summarized the state of nonresponse research in their agencies. These presentations are summarized in this appendix.

BUREAU OF LABOR STATISTICS

In a presentation to the panel, John Dixon of the Bureau of Labor Statistics (BLS) stated that the response rates in surveys sponsored by BLS range from a high of about 92 percent in the Current Population Survey (CPS) (labor force and demographics) to about 55 percent in the Telephone Point of Purchase Survey (TPOPS) (commodity and services purchasing behavior). The response trends for most BLS surveys are stable. The Consumer Price Index Housing Survey had a problem at the end of 2009 due to budgetary constraints, but has recovered. TPOPS had a decline in the last decade, but has stabilized. The American Time Use Survey (ATUS) has been low, but stable. TPOPS is a random digit dialing (RDD) survey, and ATUS is a telephone survey of specific members of CPS households. Reporting on bias studies, Dixon said that a CPS-Census match yielded propensity scores that indicated little bias in labor force statistics; the time-use survey studies have also found little bias except for "volunteering" (see Dixon, 2012). The Consumer Expenditure Survey studies have found very little bias in expenditures (Goldenberg et al., 2009).

In conducting these surveys, BLS tends to use six methods to evaluate nonresponse: linkage to administrative data; propensity scores and process data; the results of experiments with alternative practices and designs; comparisons to other surveys; benchmark data; and the R-index. When linking survey to administrative data, BLS has found that the estimate of bias due to refusals based on the last 5 percent is similar to the estimate based on linkage to the Census 2000 long-form sample. However, these studies have shortcomings in that rarely are all the records linked successfully. Consequently, the linked measure may be defined differently from the survey estimate, and it may have error.

The R-index uses a propensity score model for nonresponse and relates that to other variables (usually frame variables, such as urbanicity, poverty, etc.). The BLS studies used 95 percent confidence intervals for the R-index, somewhat flatter than the response rate. Since one of the major flaws in nonresponse studies lies in what is not known, the use of confidence intervals that account for the estimation of both the measure of interest and the model of nonresponse would be helpful.

CENSUS BUREAU

Panel member Nancy Bates from the Census Bureau reported that Census Bureau nonresponse research studies have covered the gamut. Topics have included causes of nonresponse, techniques for reducing nonresponse, nonresponse adjustments, nonresponse metrics and measurement, consequences of nonresponse (bias, costs), nonresponse bias studies, responsive designs and survey operations, the use of administrative records and auxiliary data and paradata, level of effort studies, and panel or longitudinal survey nonresponse. During her presentation, Bates offered different examples of research, including mid-decade decennial census tests to target bilingual Spanish language questionnaires; a test adding a response "message deadline" to mail materials; the addition of an Internet response option; and varying the timing of the mail implementation strategy (e.g., the timing of advance letters, replacement questionnaires, and reminder postcards). Nonresponse research in conjunction with the 2010 Census included an experiment that tested different confidentiality and privacy messages and another that increased the amount of media spending in matched-pair geographic areas. Additionally, the Census Bureau sponsored three ethnographic studies to better understand nonresponse among hard-to-count populations.

Bates also discussed nonresponse research associated with the American Community Survey (ACS), including a questionnaire format test (grid versus sequential layout), a test of sending additional mailing pieces to households without a phone number, and a test of adding an Internet option as a response mode. For other Census Bureau demographic surveys, Bates mentioned nonresponse tests involving incentives (debit cards) offered to refusals in the Survey of Income and Program Participation and in the National Survey of College Graduates. Other examples included nonresponse bias studies, including studies considering the use of propensity models in lieu of traditional post-adjustment nonresponse weights. She concluded with a discussion of administrative records and how they hold great potential for understanding non-ignorable nonresponse. Currently, most Census Bureau studies using administrative records are more focused on assessing survey data quality, such as underreporting or misreporting, and less focused on nonresponse.

Many Census Bureau nonresponse research projects are tied to a particular mode, namely mail, since both the decennial census and the ACS use this mode. Bates observed that many Census Bureau research projects are big tests with large samples and several test panels. The majority of tests try out techniques designed to reduce nonresponse, while only a few are focused on understanding the causes of nonresponse.

Bates concluded with the following recommendations:

- Leverage the survey-to-administrative-record match data housed in the new Center for Administrative Records Research and Applications. This could have great potential for studying nonresponse bias in current surveys.
- Make use of the ACS methods panel for future nonresponse studies. Its multimode design makes it highly desirable.
- Leverage decennial listing operations to collect paradata that could be used across surveys to examine nonresponse and bias.
- Select a current survey that produces leading economic indicators and do a "360-degree" nonresponse bias case study. (This ties into a recent Office of Management and Budget request on federal agency applications of bias studies.)
- Going forward, think about small-scale nonresponse projects that fill research gaps and can be quickly implemented (as opposed to the traditionally large-scale ones undertaken by the Census Bureau).
- Expand the collection and application of paradata to move current surveys toward responsive design (including multimode data collection across surveys).

NATIONAL AGRICULTURAL STATISTICS SERVICE

The National Agricultural Statistics Service (NASS) surveys farms, which are both establishments and, in surveys such as the Agricultural Resource Management Survey, households. Jaki McCarthy of NASS reported at the panel's workshop that NASS has conducted studies of its respondents and nonrespondents in an effort to test whether knowledge of and attitudes toward NASS as a survey sponsor had an effect on response. The agency found that cooperators have more knowledge and better opinions of NASS statistics. Other studies of the relationship between burden and response found no consistent relationship between nonresponse and burden as measured by the number and complexity of questions. In fact, the highest burden sample units tend to be more cooperative than low-burden units.

Other NASS studies looking at the impact of incentives on survey response have found that $20 ATM cards increased mail response, although not in-person interview responses, and that they were cost-effective and did not increase bias. Calibration-weighting studies found that calibration weighting decreased bias in many key survey statistics.

NASS is currently exploring use of data mining to help predict survey nonrespondents and determine if current patterns can be used to help provide explanatory power or if, instead, they are most useful for non-theoretical predictive power. Preliminary findings suggest that in large datasets many variables are significantly different among cooperators, refusals, and non-contacts, but although the differences are significant, they are usually small in practical terms. Many variables are correlated, and using these variables alone is not useful in predicting individual nonresponse or managing data collection.

A breakthrough procedure is to use classification trees in which the dataset is split using simple rules and all variables and all possible breakpoints are examined. In this procedure the variable maximizing the difference between subgroups is selected, and a rule is generated that splits the dataset at the optimum breakpoint. This process is repeated for each resulting subgroup. The classification trees are used to manage data collection and, in the process, allow an indication of nonresponse bias. By this means it is possible to identify likely nonrespondent groups that will bias estimates.

Despite this research, there are still a number of important and foundational "unknowns," which she summarized as follows: Is nonresponse affecting estimates? Is there bias after nonresponse adjustment? What are the important predictors of nonresponse? Can these be used to increase response? Who are the "important" nonrespondents?

NATIONAL CENTER FOR HEALTH STATISTICS

National Center for Health Statistics (NCHS) research supports a very active survey management activity designed to reduce nonresponse. As reported by Jennifer Madans of NCHS at the panel's workshop, the National Health Interview Survey (NHIS) research focuses on issues of nonresponse, with much of the research making use of paradata collected as part of the survey. NCHS uses a so-called contact history instrument, audit trails of items and interview times using the Blaise survey management platform, and analysis of the front and back sections of the survey instrument. The issues NCHS has been investigating include differences arising from reducing the length of the field period and the effort that the interviewer makes and the trade-offs between response rates and data quality. The research has found that the loss of high-effort households had minor impacts on estimates. The research also found that respondent reluctance at the first contact negatively impacts data quality. Interviewer studies have found that pressure to obtain high response rates can be counterproductive in that the pressure often leads to shortcuts and violations of procedures. These investigations have helped to develop new indicators to track interview performance in terms of time, item nonresponse, and mode.

The National Survey of Family Growth has focused on paradata-driven survey management. The survey collects paradata on what is happening with each individual case. These paradata are transmitted every night, analyzed the following day, and used to manage the survey. The paradata measures include interviewer productivity, costs, and response rates by subgroup. They emphasize sample nonrespondents, the use of different procedures (including increased incentives), and identification of cases to work for the remainder of field period.

To measure content effects the National Immunization Survey (NIS) has run several controlled experiments, along several lines of inquiry. In one experiment, NIS used such tools as an advance letter, screener introduction, answering machine messages, and caller ID (known name versus 800 number). Other experiments involved scheduling of call attempts by type of respondent and nonrespondent; incentives (prepay plus promised) to refusals and partials; propensity modeling for weighting adjustments; dual frame sampling (landline plus cell phone RDD samples) and oversampling using targeted lists; and benchmarking results against the NHIS. Findings thus far include that the response rate showed differences when the content and wording of the screener introduction were varied; advance letters, which were improved for content, readability, contact and callback information, and Website information, improved participation; a legitimate institutional caller ID improved callbacks and participation versus an 800 number; optimized call scheduling improved participation; an optimized number of call attempts by disposition type reduced costs and improved participation; and having call centers in different time zones led to improved contact and call scheduling.

NATIONAL CENTER FOR SCIENCE AND ENGINEERING STATISTICS

Work by the National Center for Science and Engineering Statistics (NCSES) centers on research to minimize nonresponse, handle nonresponse statistically, and evaluate nonresponse bias. Future research, according to Steven Cohen of the NCSES at the panel's workshop, will focus on responsive designs, increased use of paradata, and nonresponse bias analysis on the National Survey of College Graduates by making comparisons to the American Community survey.

Appendix B

Research Agenda Topics Suggested by the Literature

RESEARCH RECOMMENDATIONS FROM NATIONAL RESEARCH COUNCIL (1983)

As part of its three-volume report, the National Research Council Panel on Incomplete Data in Sample Surveys prepared separate sets of recommendations for improving survey operations and for structuring future research on nonresponse and other issues. The following text excerpts the 11 recommendations offered on future research (National Research Council, 1983, pp. 11–14).

> The recommendations on research have three objectives: to provide a capital investment in computer programs and data sets that will make nonresponse methodology cheaper to implement and evaluate; to encourage research on and evaluation of theoretical response mechanisms; and to urge that long-term programs be undertaken by individual or groups of survey organizations and sponsors to provide for and accomplish cumulative survey research, including research on nonresponse.
>
> *Recommendation 1.* General-purpose computer programs or modules should be developed for dealing with nonresponse. These programs and modules should include editing, imputing (single and multiple), and the calculation of estimators, variances, and mean square errors that, at least, reflect contributions due to nonresponse.
>
> *Recommendation 2.* Current methods of improving estimates that take account of nonresponse, such as poststratification, weighting methods, and hot-deck imputation, especially hot-deck methods of multiple imputation, require further study and evaluation.

Recommendation 3. Theoretical and applied research on response mechanisms should be undertaken so that the properties and applicability of the models become known for estimates of both level and change.

Recommendation 4. A systematic summarization of information from various surveys should be undertaken on the proportions of respondents for specified parts of populations and for particular questions in stated contexts.

Recommendation 5. Research is needed to distinguish the characteristics of nonrespondents as opposed to respondents and to assess the impact of questionnaire design and data collection procedures on the level of nonresponse.

Recommendation 6. Data sets that permit good estimates of bias and variance to be made when various statistical methods of dealing with nonresponse are adopted should be made publicly available. Such data sets could be used for testing various methods of bias reduction and for assessing effects of the methods on variances. They could also be used for the evaluation of more general methods depending on models.

Recommendation 7. Theoretical and empirical research should be undertaken on methods of dealing with nonresponse in longitudinal and panel surveys.

Recommendation 8. Theoretical and empirical research on the effects of nonresponse on more complex methods of analysis of sample survey data, e.g., multivariate analysis, should be undertaken.

Recommendation 9. A consistent terminology should be adopted for descriptive parameters of nonresponse problems and for methods used to handle nonresponse in order to aid communication on nonresponse problems.

Recommendation 10. Research on response mechanisms that depend on reasons for nonresponse should be undertaken.

Recommendation 11. Data on costs should be obtained and analyzed in relation to nonresponse procedures so that objective cost-effective decisions may become increasingly possible.

OTHER SELECTED RESEARCH TOPICS COMPILED BY THE PANEL

Research Area / Quotation	Source
Theoretical Approaches to Nonresponse	
We conjecture that there may be a direct link between the increase in efforts to contact households and refusals. Many households contacted because of the additional efforts may be more inclined to refuse precisely because of the increased contact efforts. This effect might be especially pronounced in telephone surveys, where the members of households with caller ID can see that numerous attempts have been made to contact them. If so, it is possible that the multiple attempts will predispose the household to refuse when they are finally reached.... This conjecture is consistent with our earlier suggestion that technological barriers may suppress the opportunity actually to hear the survey request. In this case, the barrier would promote refusals by increasing the rate of noncontact over time. Frustration with multiple contact attempts might also partially explain why so many RDD surveys with high nonresponse rates have low nonresponse bias. In terms of a mechanism for nonresponse, frustration with multiple contact attempts is generally not very selective and unlikely to target a particular group or subgroup.	Brick and Williams (2013:55–56)
It is interesting to note that the two most prominent and useful models for thinking about survey nonresponse—social exchange theory and leverage–saliency theory—are actually models of survey participation. They do not explicitly address the relationship between contact efforts and participation efforts. Extending nonresponse models to include the effects of contact and testing these theories might yield valuable practical advice for survey researchers.	Brick and Williams (2013:56)
Perhaps most important in the present study is the finding that the relationship between the type of respondent (cooperative, reluctant) and the attitudinal and background variables was not all in the same direction in all countries. This needs further research and discussion because it creates a serious challenge to any scholar who believes there is a theory of nonresponse that applies cross-nationally.	Billiet et al. (2007:159)

Continued

Research Area / Quotation	**Source**
Nonresponse Bias	
There may be additional hidden costs to the effort to maintain nonresponse rates in the face of mounting resistance. Many survey researchers suspect that reluctant respondents may provide less accurate information than those who are more easily persuaded to take part.... Although the general conditions that produce nonresponse bias in survey means or proportions are known (the bias is a function of both the nonresponse rate and the relation between the response "propensity"—the probability that a given case will become a respondent—and the survey variables), it is not clear what circumstances are likely to yield large nonresponse biases and what circumstances are likely to yield small or negligible ones.	Tourangeau (2003:11)
Most of the survey literature on nonresponse has focused on its impact on means, proportions, and totals. The impact of attrition may be reduced for more complex, multivariate statistics (such as regression coefficients), but clearly more work is needed to document this difference.	Tourangeau (2003:11)
Another kind of study is likely to assume increasing importance in the coming years; these studies will focus on the issue of when nonresponse produces large biases and when it can be safely ignored. Like investigations of measurement error, these studies may involve disruptions of ongoing efforts to maintain response rates (perhaps even lowering response rates by design) in order to assess their impact on nonresponse bias. In addition, it will be important to demonstrate that falling response rates actually matter (at least some of the time) and to understand the impact of nonresponse on complex statistics derived from survey data.	Tourangeau (2003:12)
More research across a range of surveys is needed to answer the question as to whether higher response rates decrease nonresponse bias. Indeed, in the light of our mixed results, we are not able to decide which of the two models, the "*continuum of resistance model*" or the "*classes of nonparticipants model*[,]" finds most support in our data. Further research on the differences and similarities in reasons for refusing cooperation between the two kinds of reluctant respondents (easy- and hard-to-convert refusals) and the refusals who were reapproached and who still refused to participate in a survey is needed.	Billiet et al. (2007:160)

Research Area / Quotation	Source
Interviewer Effects	
First, our results do not go far in explaining the mechanisms through which interviewer experience is related to cooperation. Since experience has a strong effect, further exploration of the mechanisms by which it occurs is of interest. Second, we have not addressed the question of whether experience has a positive effect due to learning or selective drop-out of less successful interviewers. Third, we believe that the lack of effect of inter-personal skills is related to problems in measuring these, rather than to the fact that they are not relevant. The question then is how such skills may be measured more successfully.	Sinibaldi et al. (2009:5968)
What is needed next are studies which address some of the other aspects of the doorstep interaction such as the intonation of the interviewers voice and non-verbal behaviour and the other various intangible things which help to determine the outcome of a request for participation. It would also be useful to try to separate out the subtleties that make a professional interviewer a professional interviewer.	Campanelli et al. (1997:5-4)
The extent to which variation in interviewer practices, sample persons' interactional moves, and the interrelation between these practices and moves have measurable effects on response rates awaits further, quantitative investigation. Nonetheless, this study highlights two challenges for such research. First, if practices are effective because of their deployment in particular contexts, then their effectiveness can be assessed only by experimental designs in which that context is considered. One cannot simply assign some interviewers to do presumptive requests and others to do cautious ones; instead, properly varying the presumptiveness and cautiousness of requests depending on the circumstances may be optimal. Interviewers would need to be trained to recognize these situations—and to do so very quickly. Second, observational studies of practices need to be careful not to confuse the influence of an interviewer's practices on a sample person with the influence of a sample person's behavior on an interviewer.	Maynard et al. (2010:810)
The influences of interviewer behavior, as well as interviewer personality traits, are not yet well understood. It seems advisable to measure interviewer behavior at the interaction level rather than the interviewer level. To better understand the process of establishing cooperation, interviewer call records need to be investigated, which only more recently have become available. It also seems advisable to control for previous interviewer performance, which requires survey agencies to record and use these data. A largely unexplored area is interviewer effects in longitudinal surveys.	Durrant et al. (2010:25–26)

Continued

Research Area / Quotation	Source
Given the apparent importance of the perception and interpretation of voice characteristics, an alternative method is to focus on the perceived interviewer approaches. Since there are probably many combinations of voice characteristics that can convey a similar interviewer approach (e.g., there are multiple ways to express authority), this method might be more fruitful. In that case, more research is needed into how interviewer approaches—as likeability, authority, and reliability—might be expressed and perceived during the introductory part of a telephone interview, and in which conditions they are effective in enhancing cooperation rates.	van der Vaart et al. (2006:497)
In general, more work is needed to assess whether certain types of survey items are more or less susceptible to nonresponse error variance or measurement error variance among interviewers.	West and Olson (2010:1022)
Interviewer incentives are ill-understood and have received little attention in the research literature, relative to respondent incentives. The mechanisms through which they may act on interviewer response rates and nonresponse bias are possibly different from those that act on respondents, as interviewers and respondents have very different roles in the social interaction at the doorstep. Further research is needed to explore how, and under what circumstances, interviewer incentives could help achieve survey goals.	Peytchev et al. (2010:26)
There is also evidence that interviewer motivation is a major contributing factor in maintaining respondents' interest in a survey and preventing break-offs. So studies of interview length should also explore the burden placed on interviewers in different modes and how this impacts on data quality.	Roberts (2005:4)
Mixed Modes	
Another question for future research is the relative power of following the attempts to obtain Web and IVR responses with a mail survey in Phase 2, rather than telephone. In many ways the telephone attempts during Phase 2 were similar to the initial contacts, i.e., both involved interaction by phone. It is reasonable to expect that switching to mail at this stage would have had a much greater impact on improving response to these treatment groups, but remains to be tested experimentally.... Using an alternative mode that depends upon a different channel of communication, i.e., aural vs. visual, to increase response may also introduce measurement differences issues that cannot be ignored. Understanding the basis of these differences should be a high priority for future research.	Dillman et al. (2008:17)

Research Area / Quotation	Source
Mixed or multiple mode systems are not new, but new modes emerge and with them new mixes. This means that we have to update our knowledge about the influence of modes on data quality. We need comparative studies on new modes and mode effects, and preferably an integration of findings through meta-analysis.	De Leeuw (2005:249)
Multiple mode contact strategies are employed to combat survey nonresponse. Still we need more research on the optimal mixes, preferably including other indicators besides response rate, such as bias reduction and costs.	De Leeuw (2005:249)
Adjustment or calibration strategies for mode mixes are still in an early phase, and more research is needed.	De Leeuw (2005:250)
Not much is currently known about people's preferences for different data collection modes. What modes would respondents prefer to use when participating in a survey? Meta-analyses of mode preference data have found that people tend to "over-prefer" the mode in which they were interviewed, but when mode of interview is controlled for, there is an overall preference for mail [surveys]. It is likely that these findings are now out of date, yet the apparent popularity of the Internet as a mode of data collection may well reflect an overall preference among respondents for self-completion. More research into public attitudes to data collection modes would shed light on this issue and might help guide survey designers in making mode choices.	Roberts (2005:3)
Offering different survey agencies/countries or respondents a choice from a range of data collection modes will be a realistic option only once it is known that a questionnaire can practicably be administered in each of the modes on offer.… Not enough is known, however, about the extent to which modes are differentially sensitive to questionnaire length (and people's tolerance of long interviews), so any survey considering the feasibility of mixing modes will need to examine this problem. [Some] survey organisations impose a limit on the permissible length of phone interviews (e.g., Gallup's "18 minute" rule). But research has shown that people's willingness to respond to long surveys depends on their motivation and ability to participate which, to a large extent, will vary by survey topic. There may also be cultural variation in tolerance of interview length (e.g., norms regarding the duration of phone calls), and these should be investigated.	Roberts (2005:4)

Continued

Research Area / Quotation	Source
We need to understand better the non-response mechanisms associated with each mode. For example, non-response in self-completion surveys is often linked to variables of interest. A weakness of face-to-face interviewing is that we get greater non-response in urban populations than in rural ones. Each mode has weaknesses, and we need to be aware of what those weaknesses are.	Roberts (2005:7)
Cell Phones	
In terms of nonresponse, cell phone response rates trend somewhat lower than comparable landline response rates, but the size of the gap between the rates for the two frames is closing. This is thought to be due to landline response rates continuing to drop faster than cell phone response rates. Research needs to be conducted to more fully understand the size and nature of differential nonresponse in dual frame telephone surveys and the possible bias this may be adding to survey estimates. Future research needs also to seek a better understanding of how dual service users (those with both a cell phone and a landline) can best be contacted and successfully interview via telephone.	American Association for Public Opinion Research (2010a:109)
Responsive Design	
While we were quite successful in predicting response outcome prior to the study, surveys vary in the amount of information that is available on sample cases. Exploring external sources of information is needed, particularly for cross-sectional survey designs that do not benefit from prior wave data and may also lack rich frame data. Similarly, more research will be needed on how to apply these data prior to any contact with sample cases. Two alternatives are to apply model coefficients from similar surveys, or to estimate predictive models during data collection as proposed under responsive survey design (Groves and Heeringa, 2006).	Peytchev et al. (2010:26)
New and effective interventions for cases with low response propensities are needed in order to succeed in the second step of our proposed approach to reducing nonresponse bias. Such interventions are certainly not limited to incentives as their effectiveness varies across target populations, modes of data collection, and other major study design features.	Peytchev et al. (2010:26)
Further research is needed into the whole sequence of the survey process and how the protocols at each stage (e.g., screening) interact with those applied on other stages (e.g., refusal conversion or interviewing) of the process. The dynamic treatment regimes approach offers a roadmap for [how] this research might be conducted. The results developed here suggest that such a research program could be successful.	Wagner (2008:76)

Research Area / Quotation	Source
Incentives	
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. However, such effects have been demonstrated in a number of studies in which the use of incentives has brought into the sample larger (or smaller) than expected demographic categories or interest groups.	Singer and Ye (2013:134)
Clearly, there is still much about incentives that is unknown. In particular, we have not examined the interaction of respondent characteristics such as socioeconomic status with incentives to see whether they are particularly effective with certain demographic groups. Geocoding telephone numbers in the initial sample might permit analysis of such interaction effects (cf. King [1998], who applied a similar method to face-to-face interviews in Great Britain). And we need better information on the conditions under which incentives might affect sample composition or bias responses. Such analyses should receive high priority in future work.	Singer et al. (2000:187)
The number of incentive experiments that could be designed is legion; unless they are guided by theory, they will not contribute to generalizable knowledge.… One question often asked is how large an incentive should be for a given survey. The issue here is the optimum size of an incentive, given other factors affecting survey response. If experiments varying the size of the incentive are designed in the context of a theory of survey participation that allows for changes in motivation over time, some generally useful answers to this question may emerge. In the absence of such theoretically based answers, pretesting is the only safe interim solution.	Singer (2000:241)
Research is also needed on how paying respondents for survey participation affects both respondent and interviewer expectations for such payments in the long run.	Singer (2000:25)
Research is needed on the conditions under which incentives not only increase response rates but produce a meaningful reduction in nonresponse bias. Because they complement other motives for participating in surveys—such as interest in the survey topic, deference to the sponsor, or altruism—it is reasonable to hypothesize that incentives would serve to reduce the bias attributable to nonresponse. Whether the use of incentives for this purpose is cost-effective is less easily answered, however, and research is needed on this topic, as well.	Singer (2000:25)

Continued

Research Area / Quotation	Source
Weighting and Nonresponse Adjustment	
Including many auxiliary variables and using the fullest cross-classification of these variables possible in the weighting will quickly result in small numbers of respondents in at least some of the weighting cells. Guidance on appropriate cell sizes for calibration weighting is very limited. The appropriate cell size is a trade-off between the potential reduction in nonresponse bias associated with increasing the information in calibration weighting and the potential increase in the variance and ratio biases of the estimates. More research is needed in this area.	Brick and Jones (2008:72)
Another area that requires more research is the effect of nonresponse on multivariate methods such as measures of association and linear and logistic regression parameters when the survey weights are used to compute these measures. The analytic results for odds ratios imply that the bias in this type of statistic could be sensitive to varying response propensities. Simulation studies on these multivariate statistics could prove very enlightening.	Brick and Jones (2008:72)
The challenge of weighting adjustment, for survey researchers and practitioners, lies in the search for an appropriate set of auxiliary variables that are predictive of both response probabilities and survey variables of interest. We encourage survey researchers to engage actively in identifying an appropriate set of auxiliary variables in developing non-response adjustment weights. This should include identifying measures at the design stage that can be obtained on both respondents and non-respondents and that are good proxy variables for one or multiple survey variables. In the past, attention was often focused on finding variables that are associated with response although small R^2-statistics are very common in response propensity models.... The results of this paper show that a renewed focus on correlates of the key survey outcome variables is warranted. An avenue that is worth exploring is statistics derived from call record data or other types of paradata that were not discussed here.	Kreuter et al. (2010:405–406)

Research Area / Quotation	Source
Paradata	
Regarding further research, we make several suggestions. First, we suggest looking to new technologies to further assess paradata validity and quality. If possible, the use of computer-assisted recorded interviewing (CARI) might be implemented. Ideally, we could record the pre-interview door-step interactions so we could have the "truth" against which to compare [content history instruments (CHI)] entries. However, given the legal and policy requirements to obtain informed consent prior to using CARI, this may prove impossible. An alternative is to have trained observers shadow interviewers, record their own versions of CHI, and then compare their records and the interviewer's. Second, we recommend bringing interviewer characteristics into the equation when assessing paradata quality (e.g., years of experience, gender, education). Since recording interviewer–respondent interactions is a rather subjective undertaking, interviewers are undoubtedly a source of systematic variance. To date, there is very little research regarding interviewer impact on the collection of paradata.	Bates et al. (2010:103)
We encourage future work in this area that might include indicators for time and part of the day or other features that would be correlates of respondent attributes related to contactability and cooperation.	Kreuter and Kohler (2009:224)
This paper did not consider the measurement error properties of the interviewer observations and record variables. We made a simplistic assumption that there is no measurement error in those variables. Of course, this assumption is debatable in the real world. Future research is needed to examine the effect of the potential measurement error in auxiliary variables on survey estimates and on the bias–variance trade-off. Although it will be difficult to do so, research is also needed on the presence and effect of selective measurement error, e.g., if measurement error in the auxiliary variables is correlated with response.	Kreuter et al. (2010:405)
Administrative Records	
Administrative records are another avenue agencies are pursuing for use as sampling frames, as survey benchmarks, as sources of auxiliary data for model-based estimates, and for direct analysis. This is a promising area for future research, Abraham said, but she added a word of caution about treating administrative records as the "gold standard" of data, because little is known of their error properties.	National Research Council (2011:7), summarizing workshop presentation by Katharine Abraham (University of Maryland, College Park)

Continued

Research Area / Quotation	Source
For many years, members of the statistical community have said that administrative records can and should be used more fully in the federal statistical system and in federal programs. The use of administrative records in the Netherlands and other countries gives a good flavor of the kinds of things the statistical system can envision doing in the United States to varying degrees. There are also areas, however, in which substantial work has already been done in the U.S. context. Most notably, administrative records have been used in economic statistical programs since the 1940s. There are also good examples of administrative data use with vital statistics, population estimates, and other programs across several federal statistical agencies.	National Research Council (2011:41–42), summarizing workshop presentation by Rochelle Martinez (U.S. Office of Management and Budget)
[Another] barrier is administrative data quality. Although they are not perfect, with survey data, agencies have the capability to describe and to understand the quality of what they have. In other words, there are a lot of measurement tools for survey data that do not yet exist for administrative records. Some have assumed that administrative data are a gold standard of data, that they are the truth. However, others in the statistical community think quite the opposite: that survey data are more likely to be of better quality. Without a common vocabulary and a common set of measurements between the two types of data, the conversation about data quality becomes subjective.	National Research Council (2011:44), summarizing workshop presentation by Rochelle Martinez (U.S. Office of Management and Budget)
Another significant data quality issue for statistical agencies is the bias that comes with the refusal or the inability to successfully link records. In addition to the quality of the administrative data as an input, the quality of the data as they come out of a linkage must be considered as well.	National Research Council (2011:44), summarizing workshop presentation by Rochelle Martinez (U.S. Office of Management and Budget)
For the future, Trépanier said, using administrative data to build sampling frames is of particular interest. There is the risk of coverage error in using an administrative database in constructing a frame, but if it is done in the context of using multiple other frames and calibration to correct coverage error, this is probably less of an issue. The ideal goal is a single frame, which is the approach used in building Statistics Canada's Address Register, but this does not preclude the inclusion of auxiliary information. A single frame would allow for better coordination of samples and survey feedback, she said.	National Research Council (2011:49–50), summarizing workshop presentation by Julie Trépanier (Statistics Canada)

Research Area / Quotation	**Source**
For data collection, one of the goals related to administrative data is to enable tracing. Statistics Canada wants to centralize the tracing process leading to the linking of all administrative data sources to make available the best contact information possible. This will require substantial effort, including a process to weigh the quality of the different sources and determine what contact information is most likely to be accurate. Another goal for administrative data could be to better understand the determinants of survey response and improve data collection procedures based on this information. For example, administrative data can provide guidance on preferred mode of data collection if one can assess whether persons who file their taxes electronically are also more likely to respond to an electronic questionnaire.	National Research Council (2011:50), summarizing workshop presentation by Julie Trépanier (Statistics Canada)
Statistics Canada has been successful in using substitution of income data from tax records, and this is likely to be continued. It is yet unclear, however, whether other information is available that could replace survey data. Investigating these options is done with caution because of the risk discussed. There is also the problem of ensuring consistency between survey and administrative data across variables.	National Research Council (2011:50), summarizing workshop presentation by Julie Trépanier (Statistics Canada)
Administrative data can also assist researchers in better understanding nonresponse bias and the impact of lower response rates. Finally, they can help both reduce the volume of data collected in surveys and improve estimation. Now that Statistics Canada has the omnibus record linkage authority in place, exploring all of these options has become a much easier process.	National Research Council (2011:50), summarizing workshop presentation by Julie Trépanier (Statistics Canada)

Appendix C

Biographical Sketches of Panel Members

Roger Tourangeau (*Chair*) is a vice president at Westat. Before joining Westat, he was a research professor at the University of Michigan Survey Research Center and director of the Joint Program in Survey Methodology (JPSM) at the University of Maryland, College Park. Earlier in his career, he worked at the Gallup Organization and the NORC at the University of Chicago. He has helped design and conduct studies involving a wide range of topics, including secondary and postsecondary education, labor force participation, privacy attitudes, health-care costs and utilization, and sexual behavior. He is well known for his methodological research on the impact of different modes of data collection and on the cognitive processes underlying survey responses. He is the lead author of a book on the latter topic (*The Psychology of Survey Response*, coauthored with Lance Rips and Kenneth Rasinski, 2000); this book received the 2006 American Association for Public Opinion Research (AAPOR) Book Award. He is also one of the coeditors of a collection of papers (*Cognition and Survey Research*, 1999) from a conference on cognitive aspects of survey response. In addition, he has published many papers on mode effects and other methodological issues in surveys. Dr. Tourangeau has received several awards over the course of his career. In 2002, he received the Helen Dinerman Award for his work on the cognitive aspects of survey methodology. This is the highest honor given by the World Association for Public Opinion Research. He received the 2005 AAPOR Innovators Award (along with Tom Jabine, Miron Straf, and Judith Tanur). He was made a fellow of the American Statistical Association in 1999 for his work on survey measurement error and his contributions to federal surveys as a sampling statistician. He is a member of the Committee on National Statistics and has served on the Panel on Residence Rules in the Decennial Census and the Panel on Design of the 2010 Census Program of Evaluations and Experiments. He holds a Ph.D. in psychology from Yale University.

Nancy Bates is a senior researcher for survey methodology at the U.S. Census Bureau. She has responsibility for conducting original research and making developmental contributions to the field of survey methodology. She conducts experimental design and testing with scholars within and outside the federal government, private research, and academia, and investigates new theories and practices in data collection methodology. She also serves as principal consultant to Census Bureau staff on methodological problems. She holds an M.A. in applied sociology from the University of Oklahoma.

Suzanne M. Bianchi is distinguished professor of sociology at the University of California, Los Angeles, and holds the Dorothy Meier chair in social equities. During 2010–2011, she was a Russell Sage Visiting Scholar. She was chair of sociology at the University of Maryland, College Park, from 2005 to 2009. She is a former director of the Maryland Population Research Center. She is a past president of the Population Association of America and past editor of the journal *Demography*. Prior to joining the Maryland faculty in 1994, she served as assistant chief for social and demographic statistics in the population division of the U.S. Census Bureau. Her research focuses on the American family, time use, and gender equality. She has coauthored four books that investigate the changes in family life and gender equality in the latter half of the 20th century. She received her Ph.D. in sociology from the University of Michigan, Ann Arbor.

J. Michael Brick is a senior statistician, vice president, director of survey methods, and associate director of the statistical staff at Westat. With more than 30 years of experience, he has special expertise in sample design and estimation for large surveys, the theory and practice of telephone surveys, the techniques of total quality management and survey quality control, nonresponse and bias evaluation, and survey methodology. He has contributed to the statistical and substantive aspects of numerous studies and to statistical methodology research in several areas, including establishment, education, transportation, and product injury studies. Dr. Brick is a fellow of the American Statistical Association, an elected member of the International Statistical Institute, and a research professor in the Joint Program in Survey Methodology at the University of Maryland, College Park. He received his Ph.D. in statistics from American University.

Douglas D. Heckathorn is professor of sociology at Cornell University. He has conducted research in formal sociological theory, policy analysis, social psychology, and quantitative methods. His research focuses on developing means for studying the structure of very large social networks using a new network-based sampling method, respondent-driven sampling. This method provides the means both for drawing probability samples of hard-to-reach and hidden populations and for studying their network structure. It has been applied in studies of a variety of populations, including injection drug users and jazz musicians. He received his Ph.D. in sociology from the University of Kansas.

Larry Hedges is Board of Trustees professor of statistics and social policy and a faculty fellow at the Institute for Policy Research at Northwestern University. He holds appointments in statistics, psychology, and education and social policy. Previously, he was the Stella M. Rowley distinguished service professor at the University of Chicago. His research straddles many fields—in particular, those of sociology, psychology, and educational policy. He is best known for his work to develop statistical methods for meta-analysis (a statistical analysis of the results of multiple studies that combines their findings) in the social, medical, and biological sciences. It is a key component of evidence-based social research. He is an elected member of the National Academy of Education and is a fellow of the American Academy of Arts and Sciences, the American Statistical Association, the American Psychological Association, and the American Educational Research Association. He is vice chair of the board of trustees of the Russell Sage Foundation, cochair pro tem of the board of the Society for Research on Educational Effectiveness, and president of the Society for Research Synthesis Methods. He holds a Ph.D. in mathematical methods in educational research from Stanford University.

Arthur Kennickell is an assistant director of the Division of Research and Statistics at the Board of Governors of the Federal Reserve System. In this position, he serves as head of the official statistical unit and oversees the conduct and development of the Survey of Consumer Finances and other Federal Reserve surveys. He has been on the staff of the Board of Governors of the Federal Reserve System since 1984. His areas of expertise are data collection and estimation methodology, microeconomics, and macroeconomics. He is a fellow of the American Statistical Association and the 2007 recipient of the Julius Shiskin Award for Economic Statistics. He has B.A. and M.A. degrees from the University of Chicago and a Ph.D. in economics from the University of Pennsylvania.

Kristen Olson is an assistant professor of survey research and methodology and sociology at the University of Nebraska–Lincoln. Her areas of research include nonresponse bias and nonresponse adjustments, the relationship between nonresponse and measurement errors, and interviewer effects. Her research has appeared in *Public Opinion Quarterly,* the *Journal of the Royal Statistical Society, Series A,* and *Environmental Health Perspectives* and is forthcoming in *Sociological Methods and Research* and *Field Methods*. She is currently coinvestigator on a grant funded by the National Institutes of Health to study health, mental health, and HIV risk behaviors among homeless women in three cities in the United States and is conducting research funded by the Bureau of Labor Statistics and the U.S. Department of Agriculture with the National Science Foundation. She is also editor of the research synthesis section of *Public Opinion Quarterly*. She has taught short courses on nonresponse bias studies for AAPOR, the Washington, DC, chapter of AAPOR, the Joint Program in Survey Methodology, and the Southern Association for Public Opinion Research. She earned her B.A. in mathematical methods in the social sciences and sociology from Northwestern University, her M.S. in survey methodology from the Joint Program in Survey Methodology at the University of Maryland, College Park, and her Ph.D. in survey methodology from the University of Michigan, Ann Arbor.

Nora Cate Schaeffer is the Sewell Bascom professor of sociology at the University of Wisconsin–Madison, where she also serves as faculty director of the University of Wisconsin Survey Center, teaches courses in survey research methods, and conducts research on questionnaire design and interaction during survey interviews. She currently serves as a member of the *Public Opinion Quarterly* Advisory Board of AAPOR and of the General Social Survey Board of Overseers. Her service for the National Research Council includes the Panel on the Design of the 2010 Census Program of Evaluations and Experiments, the Committee on National Statistics, the Panel to Review Research and Development Statistics at the National Science Foundation, and the Panel to Evaluate Alternative Census Methods. Other previous service includes the American Statistical Association Technical Advisory Committee on the Survey of Income and Program Participation; the Technical Review Committee for the National Longitudinal Survey of Youth; the National Science Foundation Advisory Committee for the Social, Behavioral, and Economic Sciences; and the governing council of AAPOR. She has also served on the editorial boards of *Public Opinion Quarterly*, *Sociological Methods and Research*, and *Sociological Methodology.* She is a fellow of the American Statistical Association. She holds a Ph.D. in sociology from the University of Chicago.

Frank Stafford is a professor of economics at the University of Michigan, Ann Arbor. He is a fellow of the Society of Labor Economists and is principal investigator of the Child Development Supplement of the Panel Study of Income Dynamics. His active research areas include issues of time allocation, the economics of childcare, and cross-national comparative studies on the role of information technology. Other research interests include family decisions about wealth, pensions, and savings as they relate to individual mental and physical health through time. He received his Ph.D. in economics from the University of Chicago.

COMMITTEE ON NATIONAL STATISTICS

The Committee on National Statistics (CNSTAT) was established in 1972 at the National Academies to improve the statistical methods and information on which public policy decisions are based. The committee carries out studies, workshops, and other activities to foster better measures and fuller understanding of the economy, the environment, public health, crime, education, immigration, poverty, welfare, and other public policy issues. It also evaluates ongoing statistical programs and tracks the statistical policy and coordinating activities of the federal government, serving a unique role at the intersection of statistics and public policy. The committee's work is supported by a consortium of federal agencies through a National Science Foundation grant.