



**Report of the AAPOR Task Force on Transitions from Telephone Surveys
to Self-Administered and Mixed-Mode Surveys**

Task Force Report

October 2019

Kristen Olson, University of Nebraska-Lincoln (Chair)
Jolene D. Smyth, University of Nebraska-Lincoln (Co-Chair)
Rachel Horwitz, US Census Bureau
Scott Keeter, Pew Research Center
Virginia Lesser, Oregon State University
Stephanie Marken, Gallup
Nancy Mathiowetz, University of Wisconsin-Milwaukee
Jaki McCarthy, National Agricultural Statistics Service
Eileen O'Brien, US Energy Information Administration
Jean Opsomer, Westat
Darby Steiger, Westat
David Sterrett, NORC at the University of Chicago
Jennifer Su, SSRS
Z. Tuba Suzer-Gurtekin, University of Michigan
Chintan Turakhia, SSRS
James Wagner, University of Michigan

Table of Contents

1	Introduction.....	5
1.1	What is Happening with Telephone Surveys?	6
1.2	Who Transitioned a Survey and Why?	10
1.3	Motivation and Consequences of the Transition	11
1.4	Examples of Surveys Transitioning from Exclusively Interviewer-Administered to Self-Administered or Mixed-Mode Designs	13
1.5	Roadmap for this Report.....	14
2	Coverage and Sample Design	15
2.1	Single Frames.....	16
2.2	Use of Multiple Frames	25
2.3	Use of Nonsurvey Data.....	26
2.4	Summary and Takeaways	27
3	Within-Household Selection and Screening of Respondents	29
3.1	Household Rostering with One Stage of Selection.....	30
3.2	Household Rostering with Two Stages of Selection.....	32
3.3	Any Adult, Most Knowledgeable Person, or Head of Household.....	36
3.4	All Adults.....	38
3.5	Age/Order Selection Methods.....	39
3.6	Last Birthday and Next Birthday	42
3.7	Respondent Selection and Sample Representativeness	47
3.8	Unique Issues in Transitioning from One Mode to Another	50
3.9	Summary and Takeaways	59
4	Questionnaire Design.....	61
4.1	Overview of Relevant Major Mode Features.....	62
4.2	Device Differences within Web Modes	76
4.3	Additional Questions That Are Particularly Hard to Transition.....	81
4.4	Questionnaire Features That Are Hard to Transition.....	91
4.5	Collection of Biomeasures, Environmental Samples, Interviewer Observations and Consent for Administrative Record Linkage	99
4.6	Summary and Takeaways	105
5	Testing Strategies for Getting Questionnaires and Other Materials from One Mode to Another	108
5.1	Expert Reviews	109

5.2	Cognitive Interviews	110
5.3	Web Probing	113
5.4	Usability testing	114
5.5	Field Tests	117
5.6	Experiments	118
5.7	Packages of Testing Strategies in Surveys that Transitioned	120
5.8	Tools to Evaluate Questionnaire Features	122
5.9	Summary and Takeaways	130
6	Recruitment, Nonresponse, and Operational Issues	131
6.1	Modes of Contact in Self-Administered and Mixed-Mode Surveys	137
6.2	Modes of Response in Self-Administered and Mixed-Mode Survey Transitions	143
6.3	Longitudinal Surveys and Transitions to Self-Administered Modes.....	148
6.4	Adaptive/Responsive Designs	151
6.5	Designing Contact Attempts.....	154
6.6	Incentives	159
6.7	Tracking Contacts in All Modes.....	163
6.8	Sample Composition.....	165
6.9	Unique Issues in Transitioning Surveys from Interviewer-Administered Modes to Self-Administered Modes.....	168
6.10	Summary and Takeaways	171
7	Data Preparation, Processing and Management.....	176
7.1	Introduction.....	176
7.2	The Importance of Transparency	180
7.3	A Note on Data Quality Control	181
7.4	Data Capture and Integration	181
7.5	Classification and coding.....	186
7.6	Review and Validate	187
7.7	Editing and Imputation	189
7.8	Weighting.....	193
7.9	Finalizing Data Files	194
7.10	Special Considerations: Longitudinal Data	195
7.11	Summary and Takeaways	196
8	Survey Estimation.....	198
8.1	Null Hypotheses to Test in Telephone vs. Self-Administered Data Comparisons	200

8.2	Assumptions Made by Single Mode and Mixed-Mode Surveys about Mode-Specific Biases	201
8.3	Diagnosing and Adjusting for Measurement and Selection Errors in Mixed-Mode Surveys.....	203
8.5	Summary and Takeaways	219
9	Costs.....	224
9.1	Factors that Might Contribute to Changing Costs	224
9.2	Differential Costs Between Modes	227
9.3	Costs per Complete Versus Sample Size.....	233
9.4	Costs for Bridge Surveys	235
9.5	Timeline as Costs.....	236
9.6	Summary and Takeaways	237
10	Human Subjects Issues	239
10.1	Obtaining Informed Consent.....	240
10.2	Protection of Personally Identifiable Information	243
10.3	Mandatory Reporting of Respondent Abuse or Harm to Self or Others.....	244
10.4	Handling Respondent Distress	245
10.5	Known Adult Respondent.....	247
10.6	Summary and Takeaways	248
11	Communicating the Impact of the Change of Modes	250
11.1	How Do You Talk to the Public and Data Users about a Break in the Time Series?.....	251
11.2	How Do You Communicate about a Break in the Time Series to the People in your Agency or Organization?	252
11.3	What Information Do We Need to Provide to Data Users on Data Files about Mode of Contact and Participation?	252
11.4	Conclusion	253
12	References.....	255

1 Introduction

Since the 1970s, telephone methods have been a ubiquitous way of collecting large scale surveys. This has been especially true for studies with complex questionnaires, surveys requiring screening for special populations, and those requiring smaller area geographic estimates. With the changing environment for telephone surveys, an increasing number of surveys are transitioning from telephone to combinations of multiple modes for both recruitment and survey administration, where phone may be only one of a number of modes that are used, if at all. Survey organizations are conducting these transitions from telephone to mixed modes with only limited guidance from existing empirical literature and best practices. This Task Force report is written with the goal of helping the survey research field navigate these challenges by examining what surveys have done in this transition, what is known, and where open areas are for additional insights and research.

To accomplish this goal, this task force reviewed existing methods reports, technical advisory panel reports, peer-reviewed literature and survey practices to develop a set of best practice recommendations for organizations transitioning ongoing phone surveys to self-administered and/or mixed-mode surveys, as well as identify needed areas of research. In this report, we provide a “lay of the land,” examining which modes are considered for use and being used when telephone surveys are transitioned to mixed-mode surveys, as well as their relative strengths and weaknesses. The goal of the report is not to provide an overview of how to do mixed-mode studies in general, but rather, we specifically focus on issues that emerge when transitioning existing telephone surveys to mixed-mode surveys (and thus requiring potential breaks in time series). In this way, this report is designed to help AAPOR members and other

survey researchers bridge between the address-based sampling (ABS) task force report (Harter, et al. 2016) and the Future of General Population Telephone Surveys task force report (Lavrakas, et al. 2017).

In this report, we evaluate issues related to sample design, household selection and/or screening for eligible respondents, and coverage of different frames and selection approaches; questionnaire design and language of administration; nonresponse and survey operations; survey estimation, including issues related to weighting and measurement error when combining data from multiple modes; and costs. We did this through three approaches. First, we conducted an extensive review of the literature, examining published articles, technical reports, conference presentations, and internal reports conducted by members of the Task Force or their organizations. Second, we reached out to the greater AAPOR community via AAPORNet and asked for any description, papers, or documentation about surveys that had transitioned from telephone to self-administered or mixed-mode approaches or were thinking about making this transition. Finally, we conducted a convenience sample survey (described below) of the AAPOR Community to get more general insights into survey organizations, reasons behind making these transitions.

1.1 What is Happening with Telephone Surveys?

Traditional telephone surveys use a mix of landline random digit dial (RDD) and, recently, cell phone RDD samples. Although landline surveys omitted households without telephones in their homes, in the US, this has been traditionally around 3-4% of the population (Blumberg and Luke 2018). To select a landline RDD sample in the US, area codes traditionally were assigned to specific geographic areas – plain old telephone service (POTS, or households),

commercial use, mobile use, or mixed; and RDD samples are drawn from sets of phone numbers called banks, often defined by the last two to four digits of a telephone number (647-555-xxxx) that have been assigned for household use. Because of the operational inefficiencies in traditional RDD sample designs, a great deal of research in the 1990s led to list-assisted landline RDD designs that improved on the efficiency of the RDD design (Casady and Lepkowski 1993; Brick, et al. 1995; Tucker, Lepkowski, and Piekarski 2002). These designs used directory listings to identify 100-banks (i.e. the last two digits of phone numbers in exchanges assigned to residential service) that had listed numbers in them for stratification purposes, sometimes dropping telephone numbers that were in 100-banks with no listed numbers (unlisted banks) altogether for increased operational efficiency. In the 1990s, the proportion of the population that was in unlisted banks was quite small (less than 4%), and they were not significantly different from the rest of the population on many characteristics, with the important exception of being more mobile (Brick, et al. 1995).

In the 2000s, cellular numbers and alternative telephone services (for example, Voice over Internet Protocol or VoIP and cable companies offering telephone service) grew notably. As shown in Figure 1, although the percentage of adults and children with no telephone service at home has remained relatively steady since 2003, the percentage of adults and children in households with access only through a cellular number has skyrocketed from about 3% in the early 2000s to 56.7% of adults and 67.5% of children in late 2018 being in households that only have a wireless number (Blumberg and Luke 2019).

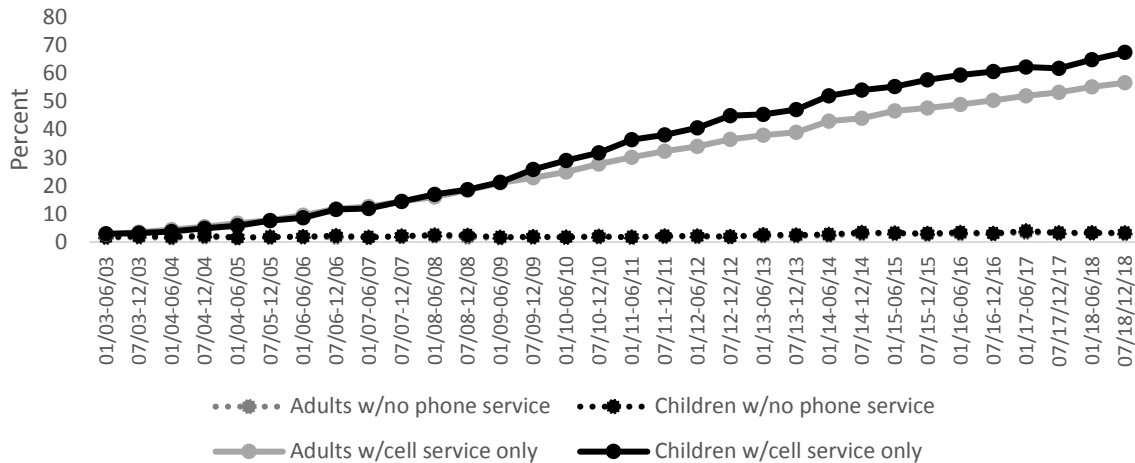


Figure 1.1: Percentage of Adults and Children with Cellular Telephone Service Only and No Telephone Service, 2008-2018, National Health Interview Survey

Source: Early Release Reports on Wireless Substitution, <https://www.cdc.gov/nchs/nhis/releases.htm#wireless>

The widespread use of cell phones had major implications for how telephone samples were designed and for operational efficiency. First, the number of listed landline numbers decreased and the proportion of households in banks with zero listings rose (Fahimi, Kulp and Brick 2009). This led to further declines in the efficiency of stratification for list-assisted designs. Second, it became necessary to include cellular RDD frames in telephone samples through dual frame designs (Lavrakas, et al. 2017). Although cellular RDD frames are functionally similar to landline RDD frames, there are no directory listings for cellular telephone numbers, reducing the efficiency with which cellular samples can be worked and limiting how sample designs can be drawn.¹ Further, post-survey adjustment weights need to be developed that account for the dual frame approach, a difficult task (Brick, et al. 2011).

¹ Cellular RDD frames have been able to add “activity flags” that indicate whether a number is active based on recent calling activity (McGeeney and Kennedy 2015; Dutwin and Malarek 2014).

Other challenges have mounted, making it implausible to use telephone surveys as the only mode of data collection for many surveys. First, response rates for samples from cell and landline telephone frames have dropped precipitously in many telephone survey designs (Lavrakas, et al. 2017). Second, a strong advantage of a traditional landline RDD frame was that geographic targeting of areas as small as a county or even a ZIP code was quite efficient because telephone companies assigned banks of telephone numbers to specified geographic areas. The ability to target landline RDD samples geographically was somewhat diluted by number portability. A Federal Communications Commission order in 2003 allowed telephone customers to keep either landline or cellular numbers when they move or change telephone service providers (Federal Communications Commission 2016). Third, cellular numbers selected into an RDD sample do not have the same geographical associations as landline numbers, with the closest useful proxy of geography the ZIP code for the billing address (Skalland and Khare 2013; Pew Research Center 2015).

Fourth, a new sampling frame providing reasonable coverage of US addresses was made available that could be used to deliver requests to a general population via postal mail. This relatively new sampling frame, known widely as Address-Based Sampling (ABS), based upon the Delivery Sequence File (DSF) of the United States Postal Service (USPS, see Harter, et al. 2016 for details), is increasingly popular. The DSF is the basis for selecting samples of addresses, corresponding to housing units; institutionalized populations are not covered. Although these lists had coverage issues when initially used in the 2000s, these coverage issues are increasingly reduced through changes in requirements for how addresses are listed (Harter, et al. 2016). As a list of addresses, the DSF can be targeted geographically, although addresses need to be geocoded to add Census geography rather than ZIP codes (Census Tract and Block

numbers). To address coverage issues, some surveys have used field staff to update addresses obtained from the DSF (e.g. Lepkowski et al. 2013). However, these issues are relatively inconsequential and the coverage of sampling frames developed from the DSF continues to improve.

These multiple simultaneous changes to the landline and cellular telephone frames and declining response rates have created a perfect storm for survey researchers attempting to measure the household population in the US (and elsewhere in the world). As such, multiple surveys have or are examining transitioning their surveys from a single-mode telephone survey to a self-administered and/or mixed-mode survey, using a combination of mail, web, phone, and/or face-to-face modes of data collection.

1.2 Who Transitioned a Survey and Why?

To understand the current status of surveys that transitioned to a different mode, the AAPOR Task Force on Transitions from Telephone Surveys to Self-Administered and Mixed-Mode Surveys (hereafter, the AAPOR Mixed Mode Task Force) conducted a survey of a convenience sample of organizations that have transitioned one or more surveys across modes, or are planning such a transition in the near future. Participation was solicited on AAPORnet and by personal contacts from members of the Task Force. Data collection began May 10, 2018 and concluded on July 2, 2018.

Representatives of 21 organizations responded to the survey, providing data about a total of at least 25 different data collection efforts. Most of these are specific named studies. Others reported on shifts in the standard data collection mode for the organization. Some of the transitioned studies involve national samples but many are geographically focused and most

target special populations (e.g., children, twins, racial and ethnic minorities) rather than the general public.

This survey includes responses from researchers in government, academia, nonprofit organizations and commercial firms, though at least half of the studies are sponsored by government agencies. Most but not all are surveys of populations in the U.S. Nearly all are household rather than establishment surveys (21 of 23 who answered this question). Most are cross-sectional (N=17) rather than panel surveys (N=7). The survey transitions reported in the study began as early as 2004 and about half of them are still ongoing.

1.3 Motivation and Consequences of the Transition

Data quality topped the list of reasons for implementing the transition of modes. A large majority (17 of 22 responding) said that response rates in the interviewer-administered survey were either “extremely” or “very” important in making a decision to transition. Anticipated response rates in the new modes closely followed (15 “extremely” or “very” of 23). Anticipated frame coverage for the new modes matched this level of importance (15). Ten organizations said that demands for greater precision, such as lower standard errors at the same level of cost, were either extremely or very important.

Table 1.1 Why transition?*Number of respondents choosing each response*

	Extremely important	Very important	Somewhat important	A little/ not at all important
Response rates to the interviewer administered survey	12	5	2	3
Anticipated response rates to the self-administered or mixed-mode survey	10	5	4	4
Anticipated coverage for the self-administered or mixed-mode studies	9	6	3	6
Costs for interviewer administered survey	9	2	3	5
Coverage of the frame of the interviewer administered survey	8	3	5	6
Anticipated costs for the self-administered or mixed-mode survey	8	2	4	6
Desire for greater precision/ lower standard errors / different estimation strategy at lower or same costs	6	4	4	7
Client demands	4	9	3	5
Sponsor or funding agency demands	3	6	3	7

Source: AAPOR Mixed Mode Task Force survey of organizations that have transitioned a survey across modes

The actual outcome of the transition on response rates was mixed. Out of 17 respondents who answered a question about this, seven said rates increased, five said they decreased and five said they stayed the same.

Mixed-mode approaches are often implemented in order to reduce costs, beginning with the most cost-effective contact modes (e.g. self-administered mail) and following-up with more costly contact modes (e.g. face-to-face interviews) to improve response rates (de Leeuw 2005; Dillman, Smyth, and Christian 2014). As such, survey costs were important to many respondents (13 wanted to reduce them, 4 to keep them constant) as a motivation for the transition. And most of the respondents (13 of 19 answering) said that the mode change had reduced the costs. Only one said that the new mode was more expensive than the interviewer-administered survey. Three said the costs are comparable.

Client demands also played a role, with 13 reporting them as either “extremely” (N=4) or “very” (N=9) important. Sponsor or funding agency demands followed (9 extremely or very important).

An open-ended question about lessons learned yielded positive suggestions and assessments about the process. Respondents stressed the importance of close attention to design elements and to test thoroughly. One said it was “win-win initiative” while another said their client was very pleased with the results. But some reported that the mode effect was much larger than anticipated and another said the process to convert questionnaires to mixed modes is lengthy. Surveying a bilingual community presented particular challenges to one study.

1.4 Examples of Surveys Transitioning from Exclusively Interviewer-Administered to Self-Administered or Mixed-Mode Designs

The report contains multiple examples from a wide range of surveys that transitioned from interviewer-administered to self-administered or mixed modes. We provide examples throughout the report from our review of published literature, technical reports, websites, and conference presentations. Early general population surveys examining the possibility of transitioning away from interviewer-administered modes (whether they did or not) occurred in the early 2000s (e.g., Cantor, et al. 2005; Link, et al. 2008; Bailey, Grabowski, and Link 2010; DiSogra, Dennis, and Fahimi 2010), coinciding with the advent of the address-based Delivery Sequence File’s availability as a sampling frame (Iannacchione 2011; Harter, et al. 2016). Many of these early surveys included phone as part of the mixed-mode approach (e.g., Murphy, Harter, and Xia 2010; Brick, Williams, and Montaquila 2011; Jans, et al. 2013). Current work now includes mail and web in the mixed-mode approaches, with some surveys using probability-based web panels or nonprobability opt-in panels as the self-administered mode replacing the telephone survey (e.g., Breton, et al. 2017; American National Election Studies 2018; Brown, et al. 2018; Ghandour, et al. 2018; Penn State Harrisburg, 2018). There are additional surveys that have not

fully abandoned telephone but have incorporated it into one of the possible modes of recruitment and/or data collection along with mail and web (e.g., Amaya, et al. 2018). **Surveys that have transitioned (or studied transitioning) from interviewer-administered to self-administered or mixed modes include both community surveys and large scale, national surveys covering a wide variety of topical domains. Thus, it is not only surveys about one topic that have moved to a self-administered mode from an interviewer-administered mode, or only for special populations.**

1.5 Roadmap for this Report

In this report, we examine the various design features that need to be considered when transitioning from telephone to a self-administered or mixed-mode survey. In doing so, we also review issues related to coverage and sample designs (Chapter 2), within-household selection (Chapter 3), questionnaire design and measurement error (Chapter 4), testing of questionnaires and other materials (Chapter 5), recruitment methods, nonresponse, and operations (Chapter 6), data preparation and processing (Chapter 7), and survey estimation (Chapter 8). We also address what is known about survey costs when transitioning from telephone to different mode(s) (Chapter 9), human subjects issues that change when transitioning modes (Chapter 10) and communicating the impact of the change of modes to the public (Chapter 11). We focus on issues related to transitioning from telephone to other modes; we cite relevant more general mixed-mode survey literature where appropriate.

2 Coverage and Sample Design

A key element to consider in any survey is the population that is covered in a sample frame. The issue of the population to which a frame allows inference is critically important when considering transitions from telephone to self-administered or mixed-mode surveys. While the sample frame can restrict or facilitate the survey mode, sampling frames and modes are distinct.

Because survey inference depends on the frame from which the sample is drawn, in ideal circumstances, a perfect frame that can be used to draw inference to all members of the target population exists, independent of the mode of recruitment or data collection. Ideally, the frame has a one-to-one correspondence to the target population. Unfortunately, perfect frames often do not exist, yielding inferences based upon the sample selected for a survey which may differ from the initially desired targeted population. Differences between the respondents and target population can be due to imperfections of the frames or errors due to nonresponse or measurement associated with particular recruitment or interview modes permitted by information on the frame.

As surveys transition from telephone to self-administered or mixed modes of data collection, one important question is whether the population covered by a frame is also changing with the mode switch, and whether a changing frame also shifts the target population. In our survey of organizations and surveys on transitions in modes, 20 respondents indicated that the target population of interest for the survey did not change when the mode changed. However, 3 respondents indicated that the target population changed, generally from a more restrictive target population (telephone households; special populations) to a more inclusive target population (all households; all adults).

In many cases, contact information on the frame limits potential mode choices, either for recruitment or survey completion. For example, landline Random Digit Dial (RDD) samples simultaneously make inference to landline telephone households possible, and limits (at least without augmentation for general population surveys) the mode of data collection and contact for the full sample to telephone interviews. Postal addresses can be identified for a subset of these landline telephone numbers on the RDD frame through identifying a telephone directory listing, but the use of a mailed survey component is then limited to these landlines with listed addresses. Data collection based on other frames, such as an address-based sampling frame, may begin with a different form of contact information (using addresses) and then merge additional contact information (telephone numbers) for a subset of the cases, allowing for the implementation of a mixed-mode survey (e.g. telephone and mail). Other sampling frames such as a list sample for an organization, may have contact information needed for multiple modes and allow survey designers the flexibility to choose which modes of contact and interviewing to apply. In settings where no single frame provides adequate coverage, multiple frames might provide the best approach to obtain the most complete coverage of the population and the most cost-efficient sample design, possibly requiring multiple (or different) modes of data collection for each frame.

In this chapter, we examine the use of both single and multiple frames that have been used when transitioning surveys from telephone to self-administered or mixed-mode surveys. We start by reviewing studies that used single frames, followed by those that used multiple frames.

2.1 Single Frames

Identifying the frame or frames that are available in order to be able to draw inferences to a target population of interest is the first step to obtaining a sample, no matter what mode is used

in contacting sample members for data collection. A number of commonly used frames for general population surveys include a frame of all landline telephone numbers (both RDD and list assisted), a frame of cellular telephone numbers, a dual frame combining cellular and landline telephones, and a frame of addresses provided by survey vendors using information from the US Postal Service. Lists of identified persons or other sampling units are used for a variety of surveys, generally focused on more specialized populations such as registered voters.

The strength in population coverage and ease of sample designs that use the address-based Delivery Sequence File (DSF) have made it a commonly used frame for general population surveys that are transitioning from interviewer-administered to self-administered or mixed modes of data collection. Yet the studies that have used the DSF for a sample frame when transitioning to a different mode are far from uniform in design, modes, and approach.

Some studies that transitioned from telephone or interviewer-administered to self-administered modes that use the Delivery Sequence File as the frame use one or more self-administered modes of mail, web, or web with a mail follow-up to recruit and collect data from individuals (e.g., DiSogra, Dennis, and Fahimi 2010; Brick, Williams, and Montaquila 2011; Montaquila, et al. 2013; Jackman 2015; Kitada 2016; Lesser, et al. 2016; DeBell, et al. 2017). Use of the DSF as a sample frame with only self-administered modes may use the frame as-is or append information for stratification or potential recruitment purposes. For example, Lesser, et al. (2016) describe a series of mode experiments comparing a transportation survey that had traditionally been conducted using a stratified RDD sample with telephone data collection to stratified random samples from the DSF frame with mail-only and web+mail modes of data collection. For both frames, the strata in each mode were defined by geographic regions in the state of Oregon. DeBell, et al. (2017) also used the DSF frame, excluding drop point

addresses where bundles of addresses receive mail together (more likely in urban areas, DeBell, et al. 2017, p. 5), and a simple random sample was selected from the list of US addresses, for the American National Election Studies. In order to examine alternative methods for recruiting individuals for participating in a web version of the American National Election Studies, information on names and other potential characteristics of sample units was purchased from a vendor and matched to the DSF frame. This match allowed the mailing to be targeted to individuals living at particular housing units rather than simply sent to the family at a housing unit or “resident,” but this additional information was not used for sample selection within the household. Most cross-sectional studies transitioning to self-administered or mixed modes from telephone do not use cluster sampling as including face-to-face interviewing as one mode in these mixed-mode surveys is rare. In one example where cluster sampling is used, Biemer, et al. (2018) used the DSF frame, excluding drop point addresses, to draw an equal probability cluster sample in the US when experimentally examining the transition of the Residential Energy Consumption Survey (RECS) from an in-person interview to a web/mail questionnaire.

Using the DSF as a sample frame does not preclude using the telephone as one of the modes of contact or data collection in a mixed-mode survey. In these studies, phone numbers are matched to some addresses from the DSF. Telephone attempts are made to sample units with matched phone numbers, and mail is used to request telephone numbers from those who were not successfully matched. Thus, use of the ABS frame may still use telephone as the primary data collection mode, but with a mixed-mode recruitment attempt. Montaquila, et al. (2013) call this the *ABS Phone-Based Model* (p. 67). For instance, Allison, Stevenson, and Kniss (2014) transitioned the Wisconsin Family Health Survey from a landline RDD stratified sample to a mixed-mode stratified ABS sample. Addresses were matched to a name and phone number (52%

of addresses yielded a phone number); those without a phone number match were mailed a short mail questionnaire requesting a telephone number.

The California Health Interview Survey (CHIS) has examined transitioning from an RDD frame to an ABS frame over a number of years. For an initial test conducted in 2012-2013 (Jans, et al. 2013), two majority Hispanic communities in California that varied in known characteristics of interest were selected. Households were selected from the U.S. Postal Service's DSF, and addresses were merged with telephone directory listings to obtain landline telephone numbers for matched addresses. All households were sent a paper screener questionnaire in both English and Spanish to obtain the household's home and cell phone telephone number(s), even for addresses for which a merged telephone number was available, as well as demographic characteristics to determine eligibility. Across the two communities, 19% returned a mail screener with a phone number; phone numbers were matched to 37% of the households who did not provide a phone number. Households that completed the screener questionnaire or who had a matched telephone number then were called for a telephone interview. Kali and Flores Cervantes (2016) used a very similar protocol in California as a pilot of the 2013-2014 CHIS in Sonoma County. Here, telephone numbers were matched to 48% of sampled addresses from the DSF address-based frame, and the sampled addresses that were not successfully matched to a telephone number were sent a short screener questionnaire to collect telephone numbers (see also Brick, et al. 2013 and Kali and Flores Cervantes 2016 for similar designs). In 2018, the CHIS used ABS and surname list frames to push respondents to a web survey in three California counties. Rather than requesting a telephone number for all respondents, respondents were requested to complete an English-language web survey or to call the CHIS directly to complete

the survey in a language other than English. Addresses with matched telephone numbers were followed up with telephone calls.

This mixed-mode data collection approach can be used in conjunction with field interviewers. For instance, Mayfield, et al. (2015) used a similar approach for a geographically-targeted and child-age specific sample in Los Angeles, adding in-person interviews for those addresses that did not return a mail screener. Sterrett, et al. (2015), examining New York and New Jersey neighborhoods affected by Superstorm Sandy, used an ABS sample to send an initial mailed invitation to participate in a web survey. Those with a matched telephone number, about 45% of the sample, were followed up with telephone calls. Remaining nonrespondents were followed up with field interviewers.

Samples drawn from lists of special populations (list samples) often provide more flexibility in the combination of modes that can be used. For instance, Parast, et al. (2018) compared telephone, mail, and a mixed-mode mail with a telephone follow-up approach as both recruitment and data collection modes for a list sample of caregivers for individuals who had been in hospice (see also Mathews, et al. 2017 for a list sample of emergency department patients with a web survey component along with mail and telephone). Lien (2015) compared phone-only with a mixed-mode web+phone survey for individuals who called into a Tobacco Helpline. Lykes and Meyers (2017) had a sample frame of new vehicle purchasers with a combination of three modes used for either recruitment or survey administration for a survey on auto quality – a paper mailed invitation to complete a web survey and following up with nonrespondents by telephone. Atkeson, et al. (2011) use voter registration files in New Mexico and Colorado as the frame, sending postal mail letters to sampled persons to complete a web

survey, information on how to request a mail survey for those who did not wish to complete the web survey, and a follow-up mail survey to a subset of nonrespondents.

Some survey organizations are turning from the probability-based RDD samples to web-based non-probability samples, either through web panels or social media analysis (Baker, et al. 2010; Murphy, et al. 2014). Rather than use a sample frame with known coverage and a means for assessing probabilities of selection, non-probability samples often use advertisements or commercial partnerships to invite a large segment of the public to participate in a survey/join a panel. Non-probability sample providers try to recruit as many people as possible (rather than specifically selected cases from an existing frame). Non-probability samples then use matching, calibration, and/or post-stratification weighting to external benchmarks to achieve their desired coverage or representativeness for a target population (Wang, Rothchild, Goel, and Gelman 2015; Elliott and Valiant 2017; Mercer, Lau, and Kennedy 2018). As such, errors related to each stage of representation - coverage, selection, and participation - are confounded in nonprobability web samples, making it difficult to identify exactly which error source is at play when estimates deviate from “true” values (Tourangeau 2019).

For instance, the Center for Survey Research at Penn State Harrisburg recently transitioned their omnibus telephone survey of Pennsylvania residents to an opt-in nonprobability web panel with quota sampling after their call yield fell from 14.4% of calls reaching a person in 2012 to 4.5% of calls reaching a person in 2017 (Penn State Harrisburg Center for Survey Research 2019b). The quotas are set by age, sex, and region of the state; the data collection approach also includes a variety of data quality checks to establish residency in the state of Pennsylvania, exclude bots, and respondents who are skipping over questions. In making this transition, the survey was able to increase sample size from n=600 telephone

completes to n=1000 web completes, with similar estimates reported on a variety of topical domains and demographics (Penn State Harrisburg Center for Survey Research 2019a, b). Other approaches blend telephone samples or other high quality surveys with nonprobability samples, using statistical adjustments to select, calibrate and/or combine the nonprobability sample with telephone or other high quality survey estimates (e.g., Ansolabehere, Schaffner, and Luks 2017; Mercer, et al. 2017; AP 2018; Dutwin 2019). AP VoteCast, a survey of the American electorate in the 2018 midterm elections conducted by NORC at the University of Chicago for The Associated Press and Fox News, used a calibration approach featuring multilevel regression and post-stratification models to combine 40,000 interviews from probability samples of registered voters with about 100,000 interviews with registered voters from non-probability online panels. In addition, SSRS has developed a “hybrid” sample that blends data from their Omnibus telephone survey and opt-in nonprobability panel that reduces the cost per sampled case with reported similar estimates on a variety of leisure activity domains (Dutwin 2019). In Spring 2019, Rutgers-Eagleton Poll and Farleigh Dickinson announced a partnership to compare and combine telephone polling data with online probability sample and nonprobability web sample data (Jenkins and Koning 2019).

How well estimates from nonprobability web samples represent the full population compared to an RDD survey varies by estimate, study, and nonprobability sample provider. For instance, Ansolabehere and Schaffner (2014) compare estimates on homeownership, cigarette smoking, and voting in the 2008 election, among others, from a YouGov nonprobability sample, a dual frame telephone survey, and a mail survey with an unknown frame (a “list provided by a data vendor”, p. 287), using propensity weights for each survey that account for standard demographics (age, race, sex, education) as well as political ideology and voter registration

status, finding lower Mean Squared Error of estimates relative to national benchmarks for the nonprobability and mail samples compared to the telephone sample (although all were relatively low). In a review of pre-elections polls for the 2016 Presidential Election, Kennedy, et al. (2018) found that opt-in internet polls had error rates similar to live interviewer RDD polls. Yeager, et al. (2011), in contrast, found that across seven different non-probability internet samples and a variety of topics including health-related topics (e.g., cigarette smoking, alcohol consumption), holding a driver's license or US passport, comparing to benchmark estimates, primary demographics (e.g., age, race, sex), and secondary demographics (e.g., marital status, homeownership, income) almost all of the unweighted estimates were less accurate than those from a probability-based web sample and less accurate than an RDD-phone sample (all surveys had a sample size of around 1000 respondents), and that the weighted estimates did not dramatically improve the accuracy of the nonprobability samples. MacInnis, et al. (2018) replicated the Yeager, et al. findings, showing again that the probability-based web sample and RDD sample were more accurate relative to benchmarks than estimates from six different nonprobability samples. Dutwin and Buskirk (2017) also found that low-response rate RDD surveys were more accurate than two different nonprobability samples on a variety of cross-classified demographic variables. Kennedy, et al. (2016) compared estimates on a wide variety of topics including volunteerism, internet access frequency, health, having a driver's license, among others, from nine nonprobability sample vendors with a probability-based web panel. They found tremendous heterogeneity in the composition of the samples and accuracy of the estimates relative to national benchmarks across the various nonprobability samples, with some having more accurate estimates than the probability-based web panel and some having less accurate

estimates. Thus, there is mixed evidence on the quality of nonprobability samples as an option as a frame and sampling method for transitioning away from RDD surveys.

An additional frame option used by those transitioning from telephone to self-administered or mixed-mode is a probability-based web panel (Blom, et al. 2016; Bosnjak, Das, and Lynn 2016; DiSogra and Callegaro 2016). Some surveys use existing probability-based web panels that were developed and built by another company, and pay a fee to conduct a survey based on the number of panel members invited to participate and/or the number of minutes of respondent time (e.g., a nonexhaustive list includes Ipsos Knowledge Panel, AmeriSpeak, Understanding America Study, American Life Panel in the US; LISS in the Netherlands, German Internet Panel and GESIS Panel in Germany). Other organizations that need frequent surveys have built their own probability-based panel. For example, the Pew Research Center transitioned their regular dual-frame RDD telephone surveys to the probability-based web-based American Trends Panel, providing internet access to non-internet households. Initially, American Trends Panel participants were recruited via an RDD request from a dual-frame sample design; in 2018, this recruitment changed to a mailed survey request selected using a stratified address-based sample from the DSF (Keeter 2019; Pew Research Center 2019). In Germany, the GESIS panel sample was developed using municipal registers (Bosnjak, et al. 2018). Then, a high-intensity recruitment effort was undertaken, including face-to-face interviewing. Subsequent waves of the panel were administered either via web or paper. Likewise, NORC's probability-based AmeriSpeak Panel, which uses NORC's address-based national sample frame, incorporates phone, mail, and face-to-face interviews during the panel recruitment and then administers subsequent surveys to the panelists via phone and web (Bilgen, Dennis, and Ganesh 2019; Dennis 2019).

2.2 Use of Multiple Frames

In some studies, multiple frames may be necessary to improve coverage and more efficiently survey the population. Use of multiple frames may not necessitate using multiple modes. For example, dual-frame telephone surveys can be used to combine landline and cellular phone numbers, with optimal allocation of sample to each frame (Lohr and Brick 2014), but the mode of data collection (telephone) is consistent across devices (landline and cell phones).

Some surveys transitioned from an RDD frame to a multi-frame design to gain efficiencies in data collection. The entire sample can be (optimally) allocated and interviewed across two frames . For example, one frame that produces a sample with good coverage properties but also has high data collection costs (possibly due to low eligibility or due to costs associated with contacting sampled units), and a second frame that produces a lower quality sample in terms of coverage but has less expensive data collection costs (possibly due to high eligibility or ease of contact). Multi-frame designs can also be simply used combining addresses available on a list frame with an ABS sample to supplement missing portions of the population from the list in order to obtain a representative general population sample.

For example, in recent studies conducted by the National Oceanic Atmospheric Association, the Coastal Household Telephone Survey (CHTS) transitioned from a landline RDD in a limited subset of counties to using a dual frame mail survey design in order to estimate fish catch in four states (Brick, Andrews, and Mathiowetz 2016). In this study of anglers, one frame used a list of state licensed anglers, with an expected higher rate of respondents who had fished. This list frame does not have complete coverage of all anglers, as not all anglers need to purchase a license, leading to coverage issues for this frame. An ABS sampling frame provided

high coverage of the states' population but had a lower chance of obtaining an angler. A sample of addresses was selected from the ABS frame to supplement the list sample. The angler list was merged with the ABS sample, and the addresses that did not merge with the list frame were subsampled at a lower rate than those that did merge with the list frame. This approach provided a more efficient design with better coverage as compared to using only one of these frames. Similar creative solutions may be useful when other rare subpopulations are of interest. For instance, addresses with household members who may speak a particular language (e.g., Spanish, Korean) may be identified through a compiled surname listing (e.g., Zuckerberg and Mamedova 2012; Brick, et al. 2013; Wells, et al. 2018). In a field test to transition the California Health Interview Survey from dual frame RDD to a mixed-mode web+phone ABS sample, Wells, et al. (2018) used Spanish, Korean, and Vietnamese surname lists to potentially identify non-English speaking households in three counties in California.

2.3 Use of Nonsurvey Data

Two reports were released by the National Academies of Sciences, Engineering, and Medicine in 2017 examining the potential for improvement in federal statistics through the use of alternative sources of data, including both government and private-sector sources. The first report discussed the multiple types of additional data sources, such as federal and state administrative data, electronic health records, web scrapings, credit card transactions, satellite images and sensor data (National Academies of Sciences, Engineering, and Medicine 2017a). The second report assesses alternative approaches for implementing procedures that would combine diverse data sources from both government and private-sector sources (National Academies of Sciences, Engineering, and Medicine 2017b). Although the National Academies

reports focused on the use of these alternative data sources for estimation purposes, they can also be used to improve the efficiency of sample frames. For example, the 2016 National Survey of Children’s Health transitioned from RDD to an ABS design using the Census Bureau’s Master Address File as the frame. The Census Bureau identified potential addresses with children and those who receive social security benefits using administrative records, as well as information about the poverty level of the block group, to stratify the addresses on the frame (U.S. Census Bureau 2018a, b; Ghandour, et al. 2018). Invitations to a web survey with mail survey follow-up were then sent more efficiently based on this stratification. As noted in the National Academies reports, more research and development is needed to evaluate these sources of data for stratification and estimation purposes.

A number of researchers outside the federal government are also exploring the availability of data provided by commercial vendors who assemble data from multiple sources, such as credit reporting agencies, magazine subscriptions, property records, and so forth (Harter, et al. 2016; Couper 2017). Commercial data are incomplete, but available for a large proportion of households (Pasek, et al. 2014; West, et al. 2015; Harter, et al. 2016). Given their imperfections, commercial data have been used in a dual frame approach or to stratify the population into groups likely to be eligible and likely to be ineligible (Valliant, et al., 2014; Brick, Andrews and Mathiowetz 2016). These commercial data can allow for additional modes of contact to improve response rates and coverage. For example, Link and Burks (2013) appended commercial data to identify housing units with young adults, in particular racial/ethnic groups, in block groups with particular demographic characteristics, and those that are matched to a telephone number to evaluate different mixed-mode strategies combining web and mail.

2.4 Summary and Takeaways

- 2.4.1 Many surveys that have transitioned from telephone to self-administered or mixed-mode approaches have used the Delivery Sequence File alone or in combination with list frames.
- 2.4.2 Use of the Delivery Sequence File as a frame does not preclude use of telephone as one of the modes in a mixed-mode survey.
- 2.4.3 Sample designs using the DSF for self-administered and mixed-mode data collections are often simple random samples or stratified samples. Cluster samples for these types of surveys are rare.
- 2.4.4 Although nonprobability web samples can be used when transitioning a telephone sample to a self-administered or mixed-mode survey, their use is not (yet) ubiquitous. Those that use nonprobability web samples all require high quality census data or probability-based samples (often collected via telephone or some other mode(s)) for purposes of sample selection or post-survey adjustment. Incorporating the use of nonsurvey data with probability-based surveys is an important area of future research.

3 Within-Household Selection and Screening of Respondents

A challenging decision related to sampling when moving from telephone to self-administered or mixed-mode surveys is how to select respondents within a household. Without an interviewer to assist the respondent, the selection decision moves out of the hands of the survey organization and into the hands of the sampled household. As such, respondent selection methods vary by mode of data collection, ranging from full probability-based to quasi probability-based methods to non-probability-based methods. Probability-based methods minimize selection bias, but require knowledge of eligible persons within the household, can be intrusive, and may result in higher nonresponse. In an interviewer-administered mode, an interviewer can assist with implementing probability-based methods, including those that require a household roster. In self-administered modes, however, the rostering of a household must be completed by a household informant. To put control over selection back in the hands of the survey organization, in some surveys, the selection process of rostering and selecting an individual is often separated into two steps – the household completes a roster, sends it back to the survey organization, and the survey organization selects the sampled person. These methods increase the length of the survey field period, resulting in higher survey costs and potentially lower response rates. Quasi- and non-probability-based methods have some level of selection bias, but they tend to be less burdensome on respondents, have higher response rates, and be more cost-efficient compared to probability-based methods (e.g., Marlar, et al. 2018).

Survey organizations make decisions about which respondent selection methods to use depending on the mode of data collection, target population, and information available on the

frame. As a survey moves from telephone to self-administered modes, the within-household selection method may also change. In our convenience sample of surveys that are transitioning from telephone to self-administered or mixed modes, 32% reported that the method of respondent selection changed when modes changed, and 40% of the respondents reported that the survey screens for special populations. Screening for special populations does not necessarily occur in two steps of household selection – only half of the surveys reported using two steps for selection screening in special populations such as children, teens, or individuals with a particular characteristic.

3.1 Household Rostering with One Stage of Selection

Full rostering methods are common in face-to-face surveys and in some telephone surveys (Gaziano 2005; Smyth, Olson and Stange forthcoming). Among probability-based methods, the Kish (1949) household roster method involves enumeration of all adults in the household by sex and age, with random selection of one respondent. After a household roster has been constructed, follow-up questions may be required to reduce household coverage error. Although this method ensures that every eligible member in the household has equal probability of selection, it imposes respondent burden and increases the likelihood of nonresponse, particularly for telephone surveys. Rizzo, Brick and Park (2004; we will refer to this as the Rizzo method) proposed a modified version of household rostering for use on the telephone where respondents are initially asked how many adults currently live in the household, and then randomly generates a selection based on the number of adults in the household. Full household rosters (or any other selection method) are used only in households with three or more adults in instances where the phone answerer is not randomly selected to be the respondent. At the time of

development of this method, only about 15% of households in the United States required full rosters. Beebe et al. (2007) compared the Rizzo method of respondent selection to the “next birthday” method, finding the same average number of attempts to interview for each method (5.6 attempts for the last/next birthday method v. 5.7 attempts for the Rizzo method), but showed a lower refusal rate and higher response and cooperation rates for the birthday method.

The Kish method is not generally used in mailed invitations for self-administered surveys. With a move to web-push designs in which the sampled household completes a roster online, the selection method with a full roster may look like the Kish method in the web mode. We know of only one implementation of a Kish method of selection for adults in a small scale mail survey, with no details provided as to how it was actually implemented by the respondent (Reich, Yates, and Woolson 1986). Gallagher, Fowler, and Stringfellow (1999) evaluated the use of a modified Kish selection procedure to identify a randomly selected child as the target of a questionnaire completed by parents. In this procedure, adults were asked if they had any children who met an eligibility criteria (aged 17 or younger and on a health care policy), and then were asked to list children’s names, ages, and sex from oldest to youngest. They then had one of six different selection instructions printed under the grid directing them to the targeted child about which to answer questions. Although this seemed complex, it yielded identical response rates to a condition in which the child was pre-selected by the researchers.

Web surveys may ask household informants to complete a household roster to select respondents within the household. Bosa, Gagnon and Caron (2017) report that the “standard” method for web surveys at Statistics Canada was that a household informant completes a full roster online, with an individual randomly selected from the household roster to be the selected respondent. In an experiment for a pilot for the Canadian National Travel Survey, the response

rate for this full roster was 13.6% (statistically lower than that for a last birthday and age-order method), with only about 3% of respondents selected incorrectly (statistically better than the last birthday and age-order methods). As part of a larger experiment (detailed below), the recruitment pilot test for the 2016 American National Election Studies (ANES) included a method akin to a modified Rizzo-Brick-Park method for households that did not match to commercial databases (DeBell, et al. 2017). Respondents were asked to report the number of adult citizens in the household, and one person was randomly selected from this number. If the individual completing the screener was selected (roughly 2/3 of sampled households), the survey continued. If the individual completing the screener was not selected (roughly 1/3 of sampled households), then the screener respondent completed a household roster to identify the appropriate respondent from the remaining adults in the household.

3.2 Household Rostering with Two Stages of Selection

In a self-administered context, full household rosters are often used in mail or web/mail surveys with two stages of selection. Looking only at mail surveys, Montaquila, et al. (2013) call this the *ABS Two Phase Mail-Based Model*. Examinations of two-stage within-household selection methods may focus on selecting an adult within a household, but often also are interested in selecting a particular subpopulation, such as children, older adults, veterans, people who engage in a certain activity, or those who speak a language other than English. In two-stage within-household selection methods, a screener is first sent to a household to obtain a household roster and limited additional information, and the survey organization uses this information to identify the respondent or focal child/teen/household member, as appropriate for the set of topical questionnaires for that household. The household is then sent a “topical” or

“main” survey, containing the survey content of interest for key survey estimates, usually with the instructions for the sampled adult to complete the questionnaire or the household informant to complete the survey for the targeted child, teen, or other household member. When screening for a particular population, other respondent selection techniques might be used as a starting point to identify a screener respondent, and then a screening questionnaire is used to ascertain an appropriate respondent meeting the more restrictive criteria for a given study.

Gallagher, Fowler and Stringfellow (1999) used a two-stage selection procedure in a mail survey to select children among households that subscribed to a healthcare plan. Parents provided a roster of children in the household. The two-stage selection did not statistically reduce response rates compared to pre-selection of a child or use of a modified Kish procedure used by a parent to select a child overall, although this two-stage procedure did reduce response rates when the parent also completed a survey about themselves.

Brick, Williams and Montaquila (2011) and Montaquila, et al. (2013) used a two-phase approach to selecting persons within a household in the mail survey version of the National Household Education Survey (NHES; see also Han et al. 2010 for a two-phase approach to sample veterans). Here, addresses were selected from an ABS frame. Selected households were asked to complete a screener questionnaire, including whether there were any and the number of children in the household and a full roster of the children in the household (name, age, sex, type of school, and year in school). Household members who met the eligibility criteria for the NHES were identified from the screener questionnaire, and a target child was selected to report on in the questionnaire. The 2009 NHES pilot study yielded a 58.7% response rate to the screener, with a 60.6% response rate from the addresses that had only mail attempts. The topical questionnaire yielded a response rate to the mail survey of 71.1% among selected screener

respondents. Both of these response rates were the same as or higher than the previous RDD version of this survey.

In 2016, a mixed-mode experiment was added to the NHES in which a subset of sampled addresses was randomly assigned to complete the screener and main questionnaire on the web rather than in the two-stage paper mail survey approach (McPhee, et al. 2018; Wilkinson-Flicker, et al. 2016). Addresses in this condition were asked to complete the screener providing information about all members of the household on the web (the 2016 questionnaire contained a topical questionnaire on adult education), and then were notified automatically about what child was selected to be the focus of the questionnaire (to be answered by a knowledgeable adult) or which adult should answer for their own educational experiences. Nonresponding households assigned to the web were sent a paper questionnaire during nonresponse follow-ups. The weighted screener response rate was 62.1% for addresses initially assigned to web, compared to 67.2% for addresses initially assigned to mail. Over 85% of the selected knowledgeable adults for the children and/or adults completing for themselves who completed the screener online continued to complete the topical surveys online. This was higher than the addresses who completed the screener by paper, either through initial assignment to that mode or through completing the screener during the nonresponse follow-ups.

Brick, et al. (2012) (see also Mathiowetz, et al. 2010) used a two-phase mail approach for surveying saltwater anglers. In addition to interest-getting questions that could also be used for nonresponse adjustment, households were asked to complete a full household roster for all members of the household, identifying the sex, age, race/ethnicity, and number of days fishing during specific months from shore or boat in a specific state (Fishing Effort Survey). Eligible households were sent a survey packet for an identified angler. Marlar, et al. (2017) expanded this

two-stage approach for this population, using a mail screener, but incorporating a mailed invitation to participate in a web survey for the topical questionnaire, followed by a mail survey sent to nonrespondents, which was completed online by 68% of respondents in two states.

In the 2016 National Survey of Children's Health (NSCH), households were asked to complete an online screener questionnaire; households without children simply identified that they did not have any children (Ghandour, et al. 2018; U.S. Census Bureau 2018). Those households with children reported the number of children in the household, the language spoken at the household, and then information for each child in the household, including the child's name, race/ethnicity, age, sex, English-speaking ability, and a variety of questions on the children's medical history to identify children with special needs. Households with children automatically had a child subsampled for the topical survey. Nonrespondents to the web survey request then received a mail screener questionnaire. Completed paper screeners were sent back to the survey organization (the Census Bureau) for processing and selection of the child.

The California Health Interview Survey (CHIS) 2018 web experiment pilot used a household roster to identify potential eligible teen and child respondents, although adults were selected using quasi-probability methods (Wells, et al. 2018). The selected adult respondent completed a household roster, and eligible children and teens were identified. The adult answered questions about the selected child and was asked for permission to survey and for contact information, including email addresses and phone numbers, for eligible teens identified from the roster. In the pilot, the adults successfully completed interviews for 79 of 136 eligible children (58.1% unweighted completion rate; 64.9% weighted completion rate), similar to the rate observed in the 2017 telephone-based CHIS (63.7%). Of the 125 eligible teens who were identified, parents provided permission for 38 teens, or 30.4%, similar to that of the 2017

telephone CHIS. Only 12 of the 38 teens completed an interview, for a 14.0% teen response rate, much lower than the 23.4% response rate from the 2017 CHIS.

DeBell, et al. (2017) experimentally evaluated a variety of selection methods for the 2016 ANES pilot study. One of the experimental methods evaluated included a mailed two-stage selection with a web follow-up, where mail respondents completed a mailed two-page screening questionnaire with a roster, and one respondent was randomly selected from this roster (condition 3). Nonrespondents to the mailed screening questionnaire were asked to follow an age-order selection procedure rather than send in a mailed household roster and complete a web survey. Other experimental conditions included a variety of web-based selection methods that incorporated information from commercial sources on names and size of household. Although the response rate to the mailed screener (54%) was higher than that for the web-based screeners (47% or 48%), the overall response rate for the web-based surveys was higher due to a larger drop off at the mailed topical questionnaire (59%) compared to the online questionnaires (83% to 89%).

The 2017 National Household Travel Survey (NHTS) used a mailed “recruitment survey” rather than a screening survey to identify the number of people in the household who would need to complete a one-day travel log (Federal Highway Administration and Westat 2018). The recruitment survey included a number of questions about transportation and the household, as well as a household roster, with a weighted response rate of 30.4%.

3.3 Any Adult, Most Knowledgeable Person, or Head of Household

Early transitions from telephone to mail surveys considered options that did not require an explicit probability model for selecting a household respondent. One of these methods was to

allow any adult in the household to participate. For instance, when examining a transition to a mail survey for the Behavioral Risk Factor Surveillance System (BRFSS), Battaglia, et al. (2008) included an “any adult” method of within-household selection experimental condition, stating “This survey should be completed by any adult, age 18 or older, living in your household except a college student living away at school; anyone in a prison, mental hospital or nursing home” (p. 468). The “any adult” method in this experiment yielded a higher response rate than the next birthday and all adults methods (described below), and yielded a somewhat less representative sample based on demographic characteristics.

Another non-probability method to obtain an individual to represent the views of the household involves asking the most knowledgeable person, decision maker, or head of household to participate in the survey. This method is typically used when the research question for the study requires knowledge about a particular issue. This respondent selection technique might involve interviewing multiple members of the household, and the person most knowledgeable about a topic area then provides the responses to the survey questions. For instance, when evaluating whether the Survey of Consumers could be transitioned from RDD to a self-administered mail survey, Elkasabi, et al. (2014) asked for “the head of the household or his or her partner complete the questionnaire” (p. 743). Biemer, et al. (2018) report on a pilot study to transition the Residential Energy Consumption Survey (RECS) to mixed web/mail modes in which a knowledgeable adult was asked to report on the energy use in the home (see also Residential Energy Consumption Survey n.d.). The use of a knowledgeable household reporter is appropriate when inference is made to the household level, but deviates from a probability sample when inference at the adult level is needed.

3.4 All Adults

An alternative method for within-household selection in self-administered surveys is to ask all eligible adults to complete a questionnaire. This can be useful when the goal is to obtain information from multiple people in the household. Here, households are not asked to select a single individual. Rather, all people who meet the eligibility criteria are asked to complete a survey. The costs of this selection method thus are higher in the mail mode due to the extra need to print and mail additional copies of a questionnaire.

Battaglia, et al. (2008) examined whether all adults in a household would complete a BRFSS survey. Households were instructed “This survey should be completed by every adult, age 18 or older, living in your household.” About 33% of households returned at least one questionnaire, but only 85% of eligible adults completed a questionnaire, resulting in an overall response rate of 28%. This was lower than the any adult and next birthday methods in this experiment, but this condition yielded a respondent pool that was more representative of young adults and of males overall. Replicating Battaglia, et al. (2008), Hicks and Cantor (2012) compared the all adult and next birthday methods of within-household selection in a mail survey version of the Health Information National Trends Survey (HINTS). They also found similar household-level response rates for the two methods (35% all adults; 39% next birthday) and that 85% of adults completed the all adult questionnaire, yielding a cumulative response rate for the all adults condition of about 30%.

Medway and Battle (2015) compare the implementation of the National Adult Training and Education Survey (NATES) in which all eligible persons in a household aged 16 to 65 who were not in high school were asked to complete a NATES adult topical questionnaire in a two-phase approach where households returned a screener, and then one adult was selected from the

household. They found that the two-stage selection yielded lower response rates for people within the household (79% compared to 96% for all adults), although households participated in some form at about the same rate in the two approaches (between 65% and 69%). Even though all adults were asked to participate, not all adults in the household completed a questionnaire, with full-household participation rates declining as the number of eligible persons in the household increased (about 90% or more for households with 1, 2, or 3, adults; 16% for households with 4+ adults).

Brick, Andrews and Mathiowetz (2016, p. 385) collected information on all adults in the household, but did not require these adults to answer for themselves. Rather, in this study of anglers, a proxy reporter was permitted to report on fishing trips made by all household members. Thus, this approach combines an all-adults level of data reported with the “any adult” method of household reporting. This approach is similar to that used in the NHTS in which reports were made for all household members about their travel during a given day, collected either via self- or proxy-report (Federal Highway Administration and Westat 2018). A mailed travel log was to be completed during the day, and then entered into a web survey or reported by telephone for each member of the household; parents were instructed to provide information for children under the age of 16. Conditional on answering the recruitment survey, the response rate for the travel log was 51.4%.

3.5 Age/Order Selection Methods

Selection methods based on the age of household members and their relative position in the household, sometimes incorporating gender of the respondent (e.g., youngest male/female; youngest male/oldest female methods; Troidahl and Carter 1964; Hagan and Collier 1983), have

been used in telephone surveys and have seen some use in self-administered and mixed-mode surveys. Age-order methods are non-probability-based respondent selection approaches commonly used for phone-based surveys typically involving short field periods. On the telephone, the interviewer asks first to speak to “the youngest male (or female), 18 years of age or older, who is now at home” where the gender is based on a rotation. If no adult male (or female) is at home, the interviewer asks to speak to the youngest female (or male), 18 years of age or older. This method targets not only younger age groups, it also targets respondents at home. The distribution of the sample therefore is heavily dependent on when the calls are made. This method also assumes that adults in the household identify with binary gender categories, something that may not be universal across all members of a household.

The age-order approach has been used as surveys transition from telephone to self-administered or mixed-mode. In these modes, the focus is not on adults at home at the time of the call, but all adults who live in the household. In a mailed survey, the wording can be quite complex and requires multiple recruitment letters to fully reflect all of the combinations of age-order in the household (see example wording below in Table 3.1). Some approaches start by asking for the number of members of the household, and then provide guidance to the respondent based on the number of adults in the household, similar to the approach used by Gallagher, Fowler and Stringfellow (1999) in their modified Kish selection of children. Other approaches simply inform the householder which person matching a particular age-order and/or sex combination has been selected for the survey.

Table 3.1. Example wording from age-order selection methods in self-administered surveys

Study	# letters	Example Wording
Bosa, Gagnon and Caron (2017)	6	<p>Oldest adult: “Who should complete this survey?”</p> <ul style="list-style-type: none"> •If you are the only person in your household who is 18 years of age or older, you have been selected to participate in the survey. •If your household has two or more members 18 years of age or older, the oldest member among them has been selected.” <p>3rd oldest adult: “Who should complete this survey?”</p> <ul style="list-style-type: none"> •If you are the only person in your household who is 18 years of age or older, you have been selected to participate in the survey. •If your household has two members 18 years of age or older, the older member of them has been selected. •If your household has three or more members 18 years of age or older, list those members in order of oldest to youngest. <p>1. _____ 2. _____ 3. _____</p> <p>The third person on the list has been selected.”</p>
DeBell, et al. (2017)		<p>We would like to ask the [oldest/youngest] [male/female] in your household who is 17 or older their views on a variety of topics related to life in the United States today. If there is no [male/female] there, then we would like the [oldest/youngest] [male/female] who is 17 or older to take the survey. The mail version of the survey is over, but the survey can still be completed online in the next two days.</p>
Olson and Smyth (2014)	6 x 2 modes	<p>Mail: To make sure we hear from all different types of Nebraskans, please share this letter with the <oldest/youngest/second youngest> adult (age 19+) <sex> in the household and have them complete the enclosed questionnaire.</p> <p>Web: To make sure we hear from all different types of Nebraskans, please share this letter with the <oldest/youngest/second youngest> adult (age 19+) <sex> in the household and have them go to the website listed below to complete the questionnaire.</p>
Olson, Stange, and Smyth (2014)	2	<p>Oldest adult: In order to make this study more scientific, we ask that the enclosed survey be completed by the adult (age 19 or older) in your household who is the oldest adult in your household.</p> <p>Youngest adult: In order to make this study more scientific, we ask that the enclosed survey be completed by the adult (age 19 or older) in your household who is the youngest adult in your household.</p>
Wells, et al. (2018)	6	<p>Condition 2C:</p> <p>Step 1: Identify who should complete the survey</p> <p>How many adults, 18 years of age or older, are in your household?</p> <p>One adult: You should complete the survey.</p> <p>Two adults: The older adult should complete the survey.</p> <p>Three or more adults: List the three oldest adults in order from oldest to youngest. The third person on the list should complete the survey.</p> <p>1. _____ 2. _____ 3. _____</p>

In an experimental comparison of web, mail, and mixed web/mail surveys, Olson and Smyth (2014) found no statistical difference in selection accuracy in an age/order selection

method across modes of data collection, with between about 15% and 20% of respondents inaccurately selected among respondents to a previous survey. Olson, Stange, and Smyth (2014) compared the youngest adult and oldest adult selection methods as part of an experimental evaluation of within-household selection methods in a mail survey. The oldest adult method yielded a response rate of 37.4%, compared to the significantly lower response rate of 32.0% for the youngest adult method. Across all households, about 30% of the selections were inaccurately made in the oldest adult procedure and about 35% were inaccurately made in the youngest adult procedure, not statistically different. Bosa, Gagnon, and Caron (2017) mailed letters identifying the adult by age position in the household, yielding a 20.4% response rate, which was higher than that for a full roster, with 13% of selections inaccurately conducted, but significantly better than the last birthday method. Bosa, Gagnon, and Caron report that the age-order method will be used for two surveys in Canada (the National Travel Survey and the Canadian General Social Survey). The CHIS piloted a web-based instrument with an age-order selection method as one of the experimental conditions for within-household selection of an adult in the household (Wells, et al. 2018). This condition yielded a weighted response rate of 13.6%. The CHIS included a household roster that allowed them to evaluate the accuracy of the selection within the household; 30% of the respondents were inaccurately selected overall. DeBell, et al. (2017) used an age-order selection for nonrespondents to a screener in one experimental treatment for the 2016 American National Election Studies pilot.

3.6 Last Birthday and Next Birthday

The quasi-probability “last birthday” and “next birthday” methods are commonly used for both telephone and self-administered surveys. The two versions of next and last birthday

differ as to whether the person who had the most recent birthday (the “last birthday”) or will have an upcoming birthday (the “next birthday”) is the selected person and asked to complete the survey. As such, telephone surveys that initially used this method do not need to transition to a different method when incorporating a self-administered questionnaire. The birthday methods avoid intrusive questions required for household rostering, but do not deliver fully randomized respondent selection due to anchoring the selection birthday date to a field period. If the month of birth was randomly assigned, then all persons would have an equal chance of selection. For instance, one could imagine randomly selecting a month within a year, and assigning that month as the “eligible” month for birthday closest to the month for a sampled household. We know of no studies that have done this approach.

Although surveys commonly state that they are using a next birthday method, how the method is implemented varies across studies. As shown in Table 3.2, some of the next birthday methods define the age for an adult (e.g., 18 years of age or older). Some embed the birthday selection within a questionnaire asking for the number of adults in the household first. Others provide some justification for the birthday method (e.g., “In order to make sure we get response from a random sample of people...”). Others provide more directions, including the need to follow the selection instructions only if more than one person lives in the household. Still others link the next birthday date to a particular calendar date.

Table 3.2. Example wording from birthday selection methods in self-administered surveys

Study	Which birthday	Example Wording
Battaglia, et al. (2008)	Next	<p>This survey should be completed by one adult living in your household.</p> <p>1. How many adults, age 18 or older, live in this household? Note: Please include yourself.</p> <p>__ __ Number of adults</p> <p>Not counting</p> <ul style="list-style-type: none"> • college students living away at school • or anyone in a prison, mental hospital or nursing home. <p>If only one adult lives here, that person should complete the survey.</p> <p>If more than one adult lives here, the one with the next birthday should complete the survey.</p>
Bosa, Gagnon and Caron (2017)	Last	<p>Who should complete this survey?</p> <p>The person in your household who had the most recent birthday, and is 18 years of age or older, has been selected to participate.</p>
Hicks and Cantor (2012)	Next	<p>1. Is there more than one person age 18 or older living in this household? Yes/No</p> <p>2. Including yourself, how many people age 18 or older live in this household? __ __</p> <p>3. The adult with the next birthday should complete this questionnaire. This way, across all households, HINTS will include responses from adults of all ages.</p> <p>4. Please write the first name, nickname, or initials of the adult with the next birthday. This is the person who should complete the questionnaire. _____</p>
Westat (2013)	Next	<p>In order to make sure we get responses from a random sample of people, we ask that the adult in your household with the next birthday complete and return this questionnaire in the next two weeks.</p>
Westat (2018)	Next	<p>In order to make sure we get responses from a random sample of people, we ask that the adult in your household with the next birthday complete and return this questionnaire in the next two weeks.</p>
Olson, Stange, and Smyth (2014)	Last, next	<p>Last birthday: In order to make this study more scientific, we ask that the enclosed survey be completed by the adult (age 19 or older) in your household who most recently celebrated a birthday.</p> <p>Next birthday: In order to make this study more scientific, we ask that the enclosed survey be completed by the adult (age 19 or older) in your household will be the next to celebrate a birthday.</p>
Olson and Smyth (2017)	Next, Next with cover, Next with confirmation question	<p>Next: “To assure that we have heard from people of all types, we ask that the adult (age 18 or older) in your household who will have the next birthday complete the enclosed survey.”</p> <p>Next w/cover: Advance letter: “To assure that we have heard from people of all types, we ask that the adult (age 18 or older) in your household who will have the next birthday complete the enclosed survey.” Additional instructions on the questionnaire cover: “Thank you for your help! Please have the adult age 18 or older in your household who will have the next birthday complete this survey.”</p> <p>Next w/confirmation q’n: Advance letter: “To assure that we have heard from people of all types, we ask that the adult (age 18 or older) in your household who will have the next birthday complete the enclosed survey.” Additional question on the questionnaire cover: “Thank you for your help! Are you the adult age 18 or older in your household who will have the next birthday? Yes -> Please continue. No-></p>

Table 3.2. Example wording from birthday selection methods in self-administered surveys

Study	Which birthday	Example Wording
Stange, Smyth, and Olson (2016)	Next	Please have the adult in your household who will have the next birthday complete the survey.”
		No calendar: Please have the adult age 19 or older in your household who will have the next birthday that will take place after July 1st, 2012, complete the questionnaire and return it in the enclosed envelope. Hearing from the person with the next birthday is very important because it ensures that we get responses from all different types of Nebraskans—men and women, the young and old, those who typically read the mail and those who do not.
		Calendar: Please have the adult age 19 or older in your household who will have the next birthday that will take place after July 1st, 2012, complete the questionnaire and return it in the enclosed envelope. Hearing from the person with the next birthday is very important because it ensures that we get responses from all different types of Nebraskans—men and women, the young and old, those who typically read the mail and those who do not. We have printed the calendar at the right in case it helps you identify the right person in your household.
		Standard: To make sure that our results accurately reflect the opinions of all Nebraskans, we ask that the enclosed survey be completed by the adult (age 19 or older) in your household who will be the next to celebrate a birthday.
		Explanatory: Some people like filling out surveys and others do not, but hearing from only certain types of people can lower the quality of our results. To make sure that our results accurately reflect the opinions of all Nebraskans, we need to randomly pick someone within your household to answer the survey. Because the timing of birthdays is pretty random, we can use them to determine who should answer. Please take a moment to think about the birthdays of all the adults (age 19 or older) in your home. Who will be the next to celebrate a birthday? We ask that the enclosed survey be completed by the adult (age 19 or older) in your household who will be the next to celebrate a birthday. To ensure the quality of our results, it is very important that this is the person to complete the survey.
Wells, et al. (2018, 2019)	Next, Next with confirmation question	Next in cover letter: Please have the adult, age 18 years of age or older, in your household who has the next birthday complete the survey. Next in cover letter + verification question: Please have the adult, age 18 years of age or older, in your household who has the next birthday complete the survey. Verification question: “Are you the adult 18 or older in your household who will have the next birthday?”

The birthday methods have received a great deal of attention for how to improve the accuracy of selection, with mostly little success. For instance, in a mail survey, Stange, Smyth, and Olson (2016) focused on the next birthday method, examining two different methods to improve representativeness and selection accuracy. First, they included a calendar as a visual

display for birthdays; this calendar had no effect on the demographic composition of the sample, and the calendar yielded less accurate selections (47% inaccurate in 2+ person households) than not including the calendar at all (37% inaccurate in 2+ person households). Second, they examined extensive explanatory language for the importance of the within-household selection method – this also had no effect on the composition of the sample.

Success in improving the accuracy of selection for the birthday methods has been found using a verification question to confirm that the selected respondent is the correct one. Olson and Smyth (2017) focused on the next birthday method in a mail survey, including the instructions (1) in the cover letter alone, (2) on the cover of the questionnaire, and (3) with a verification question to confirm that the selected respondent was in fact the person in the household with the next birthday. The next birthday method including the verification question had a lower response rate, but had the highest rate of selection accuracy out of the three methods. Wells, et al. (2018, 2019) included an experimental comparison of the next birthday method and the next birthday method with a confirmation question in the CHIS pilot. The next birthday method with the verification question yielded the highest response rate of the three methods (15%, compared to 13.9% for next birthday and 13.6% for age-order) and a substantial improvement in selection accuracy (10% inaccurately selected, compared to 30% inaccurately selected for the other methods) (this is replicated statewide in Wells, Hughes, Park, and Ponce 2019). It is unknown whether the verification question improves selection accuracy for other selection methods.

3.7 Respondent Selection and Sample Representativeness

3.7.1 Demographic Composition Differences Depending on Within-Household Selection

Method

Most of the studies examining within-household selection compare alternative methods of selecting an individual from a household *within* a mode, rather than *across* modes.

Surveys with a household informant do not require randomly selecting an eligible person within the household. Thus, respondent-level characteristics may differ from benchmarks simply because the process of identifying the household respondent in a self-administered survey (e.g., opening the mail, reading the cover letter, completing the survey) is not random and because self-selected mail survey respondents and self-selected interviewer-administered survey respondents may be, on average, different. For instance, the National Pilot test for the RECS examined four different self-administered mode combinations with a knowledgeable adult respondent in each of the mode treatments. Across the four treatments, “renters, older housing units, single-person households, and apartments” were underrepresented, especially in the mode protocol that only offered web as the mode of completion (Biemer, et al. 2018, p. 18).

The Wisconsin Family Health Survey (FHS) is a household-level survey of Wisconsin households, collecting information on all household members (Allison, Stevenson, and Kniss 2014). In 2012, the FHS switched from an RDD survey to using an ABS frame with a goal of finding a cost-effective method of identifying cell-phone only households, and started by matching addresses to listed landline phone numbers. For addresses that did not match to a landline phone number, a short mail screener was sent, with the goal of obtaining a phone number to complete a full survey, including cell-phone only households. The FHS obtained 30% of their interviews with respondents on a cell phone. Among those households who were

identified in the part of the ABS frame that was not matched to a landline telephone number, 65.3% were cell-only. Among in the part of the ABS frame that was matched to a landline telephone number, 7.8% were cell-only households. The “unmatched” sample included significantly younger respondents, respondents who were more likely to be single, more likely to be non-white, have children, live below the poverty level, and rent.

Examinations of one-stage within-household selection methods generally compare quasi-probability methods of selecting an adult within a household. As summarized by Smyth, Olson and Stange (forthcoming), **studies of within-household selection methods in telephone surveys or in self-administered surveys found few demographic differences across within-household selection methods within each of those modes.** These studies generally compared any adult, next birthday, all adults, and an age-gender position selection method (e.g., youngest adults; oldest female). Including a demographic characteristic in the selection procedure (e.g., youngest male) yielded more people with that characteristic in the sample in both telephone and self-administered modes (Marlar, et al. 2014; Olson, Stange, and Smyth 2014). Additionally, mail surveys tend to overrepresent older adults, adults with higher levels of education, and non-Hispanic white adults relative to population benchmarks (Smyth, Olson and Stange forthcoming). In a web-only survey, Bosa, Gagnon, and Caron (2017) found that the age-order selection method yielded a respondent pool that was more similar to age/sex distributions in Canada than a household roster, and was similar to the last birthday method.

Stange, Smyth and Olson (2016) found no difference in demographic composition for respondents to the next birthday method when they received additional explanatory instructions about the importance of the within-household selection method or not. They also found no difference in demographic composition for those who received a calendar embedded in the cover

letter to help with the selection task. Similarly, Olson and Smyth (2017) found no differences in demographic composition over three different placements of instructions for the next birthday method.

When using a two-stage approach to select households, the target population often changes from a general population survey to a special population. For example, the NHES uses a two-stage approach to screen households for the presence of children in certain age ranges (Zukerberg and Mamedova 2012; Montaquila et al. 2013). In pilot work that compared a 2009 national mail survey with the 2007 RDD survey conducted two years earlier, Brick, Williams and Montaquila (2011) found that the mail survey garnered more lower income and renter households, but fewer Hispanic households, differences that the authors attribute to coverage of cell-only households and English-only screening instruments in the pilot.

The NSCH also uses a two-stage approach to identify children overall and those with special health care needs in particular. A nonresponse bias analysis of the 2016 NSCH found that screener respondents were from Census Block Groups or Tracts that had slightly higher socioeconomic status and had slightly more white respondents than the full sample, and that this trend continued for topical respondents.

3.7.2 Responses to Survey Questions by Respondent Selection Methods

Few studies examining one stage within-household selection in self-administered or mixed-mode surveys within a survey across different methods have found notable differences in key survey estimates. For instance, Battaglia, et al. (2008, Link et al. 2006) found that eight health variables showed no differences across three respondent selection conditions. Similarly, Hicks and Cantor (2012), examining 24 variables in the transition of the

HINTS to mail found that only two differed significantly between the all adult and next birthday selection methods, differences that disappeared after weighting adjustments. Stange, Olson and Smyth (2016) found no difference in item nonresponse rates or substantive estimates about social attitudes or about trees and forests across different methods of implementing the next birthday method in two different mail surveys.

Olson and Smyth (2017), examining three alternative methods of implementing the next birthday method in a mail survey, found no difference in item nonresponse rates across the three experimental treatments. They also found that survey estimates related to household tasks (e.g., being a mail opener, paying bills) differed significantly across accurately and inaccurately selected respondents across the within-household selection methods. This suggests that estimates on topics for which household members differ in perceptions or behaviors are likely to be the most sensitive to within-household selection method differences.

3.8 Unique Issues in Transitioning from One Mode to Another

Transitioning a survey from an interviewer-administered mode such as phone to a self-administered mode poses unique sets of challenges.

Handoff issues. The first issue is the hand-off to another selected adult in the household. For phone, an interviewer is able to monitor the hand-off to the selected respondent. For self-administered modes, researchers have to rely on the respondent to follow the respondent selection instructions and therefore the risk of respondent self-selection is higher. It is consistently found that **accuracy of within-household selection is lower in larger households** (Battaglia, et al. 2008; Olson and Smyth 2014; Olson, Stange, and Smyth 2014; Olson and Smyth 2017; Smyth, Stange, and Olson forthcoming), **generally because the handoff to a**

respondent other than the initial informant is considered to be challenging. Battaglia, et al. (2008) report that respondents to a mail survey who were not those with the next birthday in the household completed the survey because “the person with the next birthday did not want to fill out the questionnaire” (p. 466). Brick, Andrews, and Mathiowetz (2016) avoid this issue by requesting proxy reports for all members in a household, but this may not be feasible or desirable for all studies.

Mode of screener. There are few studies that experimentally examine mode differences for completion of screening instruments separate from completion of the overall questionnaire, largely because this requires using a two-stage screening design for within-household selection. **Those that do examine screener mode differences tend to find that mail screening instruments tend to yield higher completion rates than other modes of data collection.** For example, in an experimental comparison of a two-stage sample of veterans, the response rate for a paper screener was almost 7 percentage points significantly higher than that of a web screener, and the effective coverage rate of the target population was significantly better (Han, et al. 2010). In the telephone-based 2007 NHES, the screener response rate was 53%, compared to a screener response rate of 69% for a mail-based NHES in a 2011 Field Test for the same survey (Montaquila, et al. 2013). The NHES experimentally compared web and mail modes of collecting information for the screener in 2016, and the mail screener response rates was about 5 percentage points higher than the web screener response rate (McPhee, et al. 2018; p. 87). Amaya, et al. (2015) experimentally examined screener completion rates for a sequential mixed-mode telephone-mail survey from an address-based sample matched to telephone numbers in six US communities targeting particular racial/ethnic groups. The experimental treatment in which mail surveys were sent initially with a phone follow-up yielded a higher screener completion rate

(48.7%) than the experimental treatment where phone attempts were followed up by a mail questionnaire (44.8%), with notably more mail screeners being completed than telephone screeners. The conditions for the ANES recruitment pilot that included screening by mail were higher (54%) than those that screened via a web instrument (about 48%), but the response rate for the topical survey was lower for those who were screened by mail and asked to go online for the main survey compared to those who were screened and interviewed all online (59% topical response rate for mail; >81% topical response rate for web survey) (DeBell, et al. 2017).

Screener incentives. With a two-stage selection approach, an open question is whether to and at what level to provide incentives for completing the screener questionnaire. **In the existing experimental comparisons, prepaid incentives improve response rates to mailed screening questionnaires.** Montaquila, et al. (2013) and McPhee (2012) compared two levels of prepaid incentives (\$2 vs. \$5) for the screener questionnaire in a field test for the 2011 NHES. Unsurprisingly, the incentives increased the screener response rate initially (46.1% responded to the first mailing with a \$5 incentive, compared to 39.7% with a \$2 incentive), a difference which held even with additional mailings (final response rate \$5 68.9% vs. \$2 65.0%). This experiment was replicated on the NHES in 2016, with less than a 2 percentage point difference between the two experimental conditions on screener response rates, but notable differences in screener completion rates across groups that varied simultaneously on incentive level and predicted response propensity (McPhee, et al. 2018). Adding a \$1 incentive with a screener survey for anglers increased response rates to the screener survey by about 12 percentage points in data collection efforts during the second half of a year compared to data collected during the first half of the year with no incentives (Andrews, Brick, and Mathiowetz 2013). In a one-stage mail survey that screened for anglers, higher incentive levels raised the response rate for the mail

survey at a decreasing rate (Brick, Andrews, and Mathiowetz 2016). Prepaid incentives also increased screener completion rates in the web/mail National Survey of Children's Health (\$2: 53.2%, \$5: 55.3%) over no incentives (50.3%) (US Census Bureau 2018).

Content of the screener. **In a two-stage approach, what information to include in a screening questionnaire above and beyond the questions needed to determine eligibility is an open question, with few studies replicating the same design features or yielding the same results.** In the 2009 NHES and 2011 NHES pilot studies, alternative content of the screener questionnaire was evaluated, comparing a short questionnaire that contained the necessary information for determining eligibility, but nothing else (the “screen-out” version) to longer questionnaires that contained information about the topic (the “engaging” and “core” versions) (Brick, Williams, and Montaquila 2011; Montaquila, et al. 2013). Although in the 2009 NHES pilot study, the screen-out version yielded a higher response rate, the engaging version saw a higher response rate in the 2011 NHES pilot. For a two-stage sample of veterans, including one single question inquiring about being on Active Duty significantly increased response rates by 3.4 percentage points compared to not including that question at all, but did not change the effective coverage of veterans (Han, et al. 2010).

Design of the survey mailing package. **A number of methods have been used to attract the attention of respondents, sometimes aimed at a particular subgroup of interest, with very little difference in response rates.** For instance, among respondents to a paper questionnaire on veterans, including an insert attempting to garner attention to the survey had no effect on response rates, but significantly improved coverage of the target population (Han, et al. 2010). Similarly, Stange, Smyth, and Olson (2019) found that including images of LGB adults and families significantly improved coverage of the LGB population in a one-stage mail survey,

with no difference in response rates compared to the “default” condition with no images of LGB adults and families. In the CHIS, a sponsor logo on the exterior of an envelope depressed screener response rates in one geographic area, but not in another (Jans, et al. 2015). There was no difference in return rates when a Health Resources and Services Administration’s Maternal and Child Health Bureau logo appeared in a follow-up mailing compared to a Census logo in the National Survey of Children’s Health (US Census Bureau, 2018b).

Languages. Studies face challenges when administering surveys in multiple languages. In an interviewer-administered mode, the interviewer is able to identify cases with a language barrier, and an appropriate bilingual interviewer is then able to call back the respondent to complete the survey. For self-administered modes, the survey organization either has to have translate the survey into multiple languages or consider other modes for rare language cases.

Self-administered and mixed-mode surveys conducted in multiple languages often include cover letters and survey questionnaires in these multiple languages from the initial mailing. Asking non-English-speaking respondents to call into a language-specific telephone survey is less successful. The NHES conducted telephone interviews in both Spanish and English, totaling about 5% of phone-based screener interviews (Zuckerberg and Mamedova 2012). To transition from telephone to mail, in 2011, the NHES tested an English-only, a Spanish-only, and a bilingual screening form for a nationally representative sample of households and a Spanish-targeted sample, focusing on linguistically isolated Census tracts and individuals with a Spanish surname who lived in non-linguistically isolated Census tracts. The NHES found that the timing of the Spanish screener form affected both response rates and who participated in the screener and the topical survey, recommending that surveys include a Spanish-language screener with an English-language screener in each mailing of the screener to

better identify Spanish-speaking households (see also Montaquila, et al. 2013; Brick, et al. 2012). In particular, respondents were more likely to be white and less likely to speak Spanish as the primary language when both English and Spanish screener forms were included in the second mailing compared to when both forms were included in all mailings, but did not differ in parental education, household tenure, or household income. Additionally, among Spanish linguistically-isolated Census tracts, offering a Spanish-language screener yielded more Hispanic respondents, lower levels of education, higher levels of renting, and higher household income than those in an English-only screener. As part of a screener to identify children eligible for a homeschool questionnaire, NHES changed the wording of items identifying whether a child is homeschooled, a concept that was difficult to translate accurately into Spanish (Battle, Megra, and Wan 2017).

In contrast, a pilot study for the National Crime Victimization Survey in Chicago experimentally compared mailing bilingual screening materials (Spanish language and English language screening surveys) versus English-only screening materials to addresses in areas that were not linguistically isolated and did not have a Hispanic surname (Brick, et al. 2013). The response rates were about four percentage points lower for the bilingual screeners, and only 4 respondents completed the materials in Spanish.

The 2012-2013 test for the CHIS contained a Spanish-language and English-language screener form in every mailing, and translated the cover letter from English to Spanish, placing them on opposite sides of the paper (Jans, et al. 2013). The CHIS recently tested a transition from phone to a web-push/phone survey, in which English language questionnaires are initially attempted via the web, and speakers of other languages are asked to call into a phone line to talk with an interviewer who speaks Spanish, Chinese, Korean, Vietnamese, or Tagalog (Wells, et al. 2018). All non-respondents from the initial web-push phase were followed up with a telephone

call to attempt a telephone interview when a telephone number was matched to the address. The 2018 CHIS experiment also included surname lists to target Spanish and Korean/Vietnamese households. The surname lists yielded a 5.8% cooperation rate (compared to 9.1% for the ABS sample), and only 11 interviews were conducted as part of this experiment in a language other than English (Wells, et al. 2018).

The NSCH also provided an English and Spanish version of the screener and the topical survey. Spanish-language translations were printed on the back on the invitation letters, and respondents could request a Spanish-language paper screener and topical questionnaire. The web survey included an option to switch between English- and Spanish-language instruments, with about 350 web screeners completed in Spanish and about 250 web topical questionnaires completed in Spanish (US Census Bureau 2018a,b; Ghandour, et al. 2018). These versions were not experimentally varied.

Minors. Research with minors has inherent challenges. Persons under age 18 are a protected class of people by law and interviewing minors comes with stringent consent requirements. Parental consent is required to interview any minor children under the age of 18. This means that the researchers must identify and interview two different people: the parent or legal guardian to obtain consent and the minor. Nonresponse can occur at two places: first with the parent refusing to allow their child to be interviewed and second due to non-contact or refusal by the minor to participate in the survey. For a single-mode telephone survey, for instance, this requires more phone calls, voicemails, and follow-up messages to first speak with the adult to obtain consent and then to reach the minor for whom the parent has consented to be interviewed. The type of telephone can also further complicate data collection, where a parent might be reached on a cellular phone but their 17-year old teenager is more easily reached at a different

cell phone number. The preferred language for taking the survey for the parent and their minor child can further pose challenges and increase the level of effort by the survey organization.

Transitioning an existing phone survey to a self-administered (web or mail) mode for research to screen and identify minors faces a unique set of challenges. As with identifying a sample of adults, identifying a sample of children or teenagers can be done in a single-stage or two-stage self-administered survey. Which approach is best depends on whether parents/guardians provide proxy reports for all of their children, a single child, or the child is asked to report for themselves. For example, in the redesigned web and mail-based NSCH, household informants completed a screener questionnaire to identify whether there were any children in the home, including those who met particular survey criteria of having special health care needs or being young. Focal children in the household were then randomly selected from the screening questions, and the adult household informant completed a survey about the child. In the web version of the NSCH, children were selected automatically via the web instrument; in the paper version, a two-stage selection occurred – the household returned the screener questionnaire, and then a topical questionnaire was sent with the identified child in the household prefilled in the questionnaire (US Census Bureau 2018a,b; Ghandour, et al. 2018). This approach for an adult respondent providing proxy reports for children is similar to the design for the NHES (e.g., Brick, et al. 2011; Montaquila, et al. 2013; McPhee, et al. 2018) and the child portion of the web-based CHIS pilot (Wells, et al. 2018).

Difficulty with transitioning to a self-administered mode increases substantially when the minor is a teen who is requested to answer survey questions for themselves. Here, the parent must provide both permission to contact and interview the teen. Difficulty arises when attempting to accurately capture an email address or cell phone number to request that the teen

complete an online survey, or engaging a teen respondent who is unaware of or disinterested in research activities and must rely on their parent to inform them about the mail or web survey they are being asked to complete. Questions about data quality could also become an issue if survey research topics are sensitive in nature and parents are present when their children participate in an online survey, without researcher knowledge.

To our knowledge, few studies that have transitioned from telephone to self-administered modes have attempted this difficult task. The CHIS pilot collected data from teens on the web by first asking parents for permission and contact information for a selected teen respondent, and then following up with the teens. Out of the 125 eligible teens, parents provided permission for only 38 of these teens, and completed interviews were obtained from only 12 of them, yielding about a 10% cumulative response rate among the eligible teens (Wells, et al. 2018). Cantrell, et al. (2018) report on an ABS-screener to identify youth and young adults aged between 15 and 21, with 1,293,801 households sent survey invitations to complete a web-based screening questionnaire by a household respondent who completed a household roster. Age-eligible household members were identified and one teen or young adult was randomly selected from the household. Parents provided consent for teens aged 15-17, as well as contact information. Of the 1,293,801 households contacted, 40,464 completed the web screener questionnaire (3.1% of the total sample), 12,882 were identified as eligible (31.8% of the screener completes, Cantrell, et al. 2018, Table 1), and 10,257 completed the questionnaire. The National Survey of Children's Exposure to Violence (NatSCEV) is conducting methodological research to transition from telephone to self-administered modes, with parental reports for children aged 2 to 11, and self-reports for children aged 12 and older (Brick, Steiger, et al. 2018). Recruiting and successfully

interviewing teens in an ABS-sample, self-administered mixed-mode survey requires further investigation.

3.9 Summary and Takeaways

- 3.9.1 As in interviewer-administered surveys, there is no single method for selecting a respondent within the household in self-administered and mixed-mode surveys.
- 3.9.2 Surveys that transition from telephone to self-administered or mixed modes may use the same methods of selecting a respondent within a household or may change the methods of selecting a respondent within a household.
- 3.9.3 Full household rosters are often used in mail or web/mail surveys with two stages of selection. Mailed household rosters are more likely to be completed than other modes for a screener questionnaire. However, household rosters that are completed online seem to successfully transfer respondents to the online instrument at higher rates.
- 3.9.4 A variety of probability, quasi-probability, and non-probability methods are used in self-administered or mixed-mode surveys with one stage of selection. Selection within a household is often inaccurately made where accuracy can be evaluated; asking respondents to verify that they meet the selection criteria can help reduce inaccurate selections.
- 3.9.5 Studies of within-household selection methods in telephone surveys or in self-administered surveys found few demographic differences across within-household selection methods within each of those modes, each over- or underrepresenting groups in similar ways. There are few evaluations directly comparing representation of different demographic groups across modes for different within-household selection methods.

- 3.9.6 Few studies examining one stage within-household selection in self-administered or mixed-mode surveys across different methods have found notable differences in key survey estimates.
- 3.9.7 The type of information to include in a screening questionnaire above and beyond the questions needed to determine eligibility is an open question, with few studies replicating the same design features or yielding the same results. Similarly, experimental designs examining properties of a one-stage survey vary, with few consistent design features or outcomes.
- 3.9.8 Prepaid incentives improve response rates to mailed screening questionnaires.
- 3.9.9 Successful self-administered and mixed-mode surveys conducted in multiple languages include cover letters and/or survey questionnaires in these multiple languages from the initial mailing. Asking non-English-speaking respondents to call into a language-specific telephone survey is less successful.
- 3.9.10 As with identifying a sample of adults (including special population groups), identifying a sample of children or teenagers can be done in a single-stage or two-stage self-administered survey. Which approach is best depends on whether parents/guardians provide proxy reports for all of their children, a single child, or the child is requested to report for themselves. Parental reports for children occur at about the same rates as in a telephone survey, but more research is needed on obtaining successful cooperation from teen respondents in a self-administered or mixed-mode survey.

4 Questionnaire Design

Each survey mode is made up of features that affect what types of questions can be asked, other measurements that can be collected, and the quality of these measurements (de Leeuw 2005). Any two modes may have some of these features in common and others that differ. Thus, in transitioning from telephone to self-administered or mixed modes, a major challenge is determining if it is possible to collect the necessary information at the required quality level in the new mode(s) and how to do so. It is also important to consider respondent characteristics that, through their interaction with these mode features, may make surveying in a specific mode more or less difficult or accurate.

We first provide an overview of questionnaire design features that may differ when transitioning from telephone to self-administered or mixed-mode surveys. We then briefly review potential differences in measurement quality for different types of devices within modes (e.g., landline versus cell phone telephone interviews; desktop/laptop computer versus mobile devices for browser-based web surveys) and what is known about surveys that transitioned to a mode including the web. Next, we turn to additional types of questions and questionnaire features that are problematic to transition. We end with discussing collection of biomarkers, environmental samples, and consent to link survey data to other records, as well as our summary and takeaways about questionnaire design when transitioning from telephone to self-administered or mixed modes of data collection.

4.1 Overview of Relevant Major Mode Features

There are three primary dimensions on which modes differ: whether they are interviewer- or self-administered, aural/oral or visual (or both), and computerized or not computerized. These three characteristics, individually and in concert, have implications for how respondents experience a questionnaire and thus the responses they give. **All surveys that transition from telephone to self-administered or mixed modes must transition questionnaires from an interviewer-administered to a self-administered mode, and from an aural administration to a visual administration. Some surveys that transition from telephone to self-administered or mixed modes also transition from a computerized mode to a non-computerized mode.**

4.1.1 Interviewer-Administered Versus Self-Administered Modes

Telephone and in-person interviewers can take advantage of the social basis of surveys by listening and/or watching for cues that the respondent is not understanding questions and then providing clarification or by following up inadequate answers with feedback or probes (Schwarz, et al. 1991). Interviewers can also motivate the respondent to complete the survey or answer optimally through probing or other behaviors, and can order the items as presented in the questionnaire. Self-administered surveys do not have these benefits of having an interviewer in administration, clarification, motivation, or order of presentation of items.

In interviewer-administered questionnaires, “don’t know” and “refused” options can be available for respondents without explicitly offering them aloud, accepting a volunteered “don’t know” or “refused” response after an initial probe is unsuccessful. In a web or paper questionnaire, when interviewer presence is not possible, offering a “don’t know” or “refused” response as an explicit response option is the only way to communicate to the respondent that the

response is a valid one. Explicitly offering these nonsubstantive response options in self-administered modes often results in a higher selection of them than occurs in interviewer-administered modes where they are accepted on a voluntary basis (Nicolaas and Tipping 2006; Jones, et al. 2015) or even where the options are read aloud (Klausch, Hox, and Schouten 2013). In our survey of survey organizations that transitioned from telephone to self-administered or mixed modes, 4 organizations reported having explicit don't know responses only in the interviewer-administered mode, 6 used explicit don't know options only in the self-administered mode, 10 had them in both modes, and 3 reported not using them at all. When nonsubstantive options are not explicitly offered in self-administered surveys, respondents can simply leave items blank, although the analyst has no means of knowing if the respondent didn't know the answer, didn't want to give the answer, or just accidentally skipped a question.

In general, since self-administered modes are normally more prone to item-nonresponse than interviewer-administered modes (Nicolaas and Tipping 2006; Heerwegh and Loosveldt 2008; Heerwegh 2009; Klausch, Hox and Schouten 2013; Breton, et al. 2017), **surveys experience slightly higher item nonresponse rates when transitioning to self-administered modes.** Figure 1 shows examples of average (mean or median) item nonresponse rates before and after mode transitions for the National Household Education Surveys (NHES) and Residential Energy Consumption Survey (RECS) surveys.

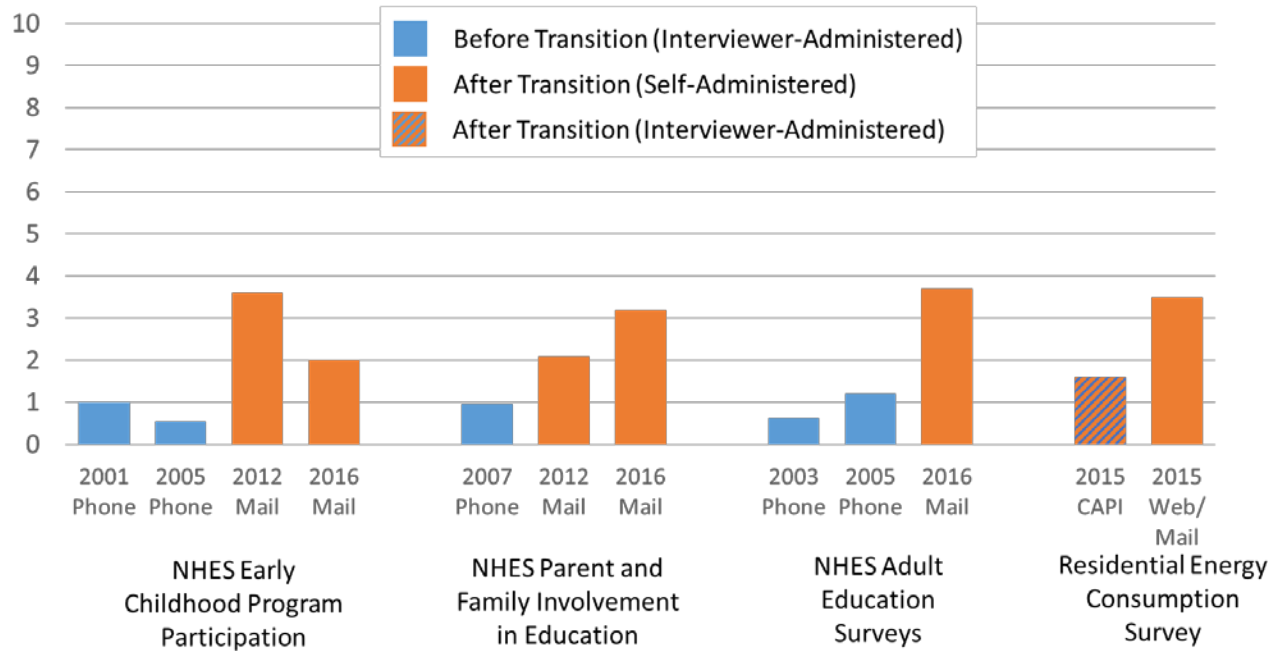


Figure 1: Item nonresponse rates by survey mode before and after transitions

Exceptions to the trend of higher item nonresponse rates in self-administered modes may be on sensitive questions (Nicolaas and Tipping 2006; Liu 2018). However, for some sensitive questions, “don’t know” may be the more embarrassing response, resulting in fewer people selecting it (e.g., the number of times one has had sex recently, Olson, Smyth, and Ganshert 2019).

For knowledge questions, a “don’t know” response may be more accurate than a guess or may be a legitimate answer. But, respondents to web surveys may be able to look up answers and thus, transitioning to a web-based mode may have unintended consequences on knowledge items. For example, in the American National Election Studies (ANES), web respondents had higher levels of political knowledge than face-to-face respondents on 10 of 13 political knowledge questions (Liu and Wang 2014; see also Chang and Krosnick 2009 and Ansolabehere and Schaffner 2014 for similar findings on political knowledge). This pattern is attributed to web respondents being able to look up answers for factual questions on the internet (Clifford and Jerit

2016). Fricker, et al. (2005) found higher levels of science knowledge for respondents who participated via web compared to those interviewed via a telephone, and that it took about four more minutes for respondents to complete open-ended knowledge questions on the web compared to the telephone; here, the authors did not attribute the mode differences to information seeking on the internet. Domnich, et al. (2015) found a significant difference in health-related knowledge for timed vs. untimed administration in a web survey for items that were easily searchable on the internet, but no difference on items that were not easily searchable on the internet. A randomized controlled mode of interview experiment conducted by Gooch and Vavreck (2019) as part of pilot research for the ANES found that respondents in the web-based self-administration condition scored better on knowledge questions than those in the face-to-face interview condition. In this study, which was conducted at the CBS research facility in Las Vegas and not at respondents' homes, only two of the 505 respondents assigned to the self-administered mode had looked up the answers. It is not clear whether these findings translate to a more general population survey. Thus, **surveys with knowledge questions may see an increase in estimated knowledge when transitioning from telephone to web-based self-administered modes.**

While interviewers can improve measurement, they can also have negative effects on measurement quality such as when they introduce interviewer bias (i.e., estimates are artificially low or high) or variance into measures. Interviewer bias occurs when responses are influenced by interviewer characteristics such as gender (Groves and Fultz 1985; Catania, Binson, Canchola, Pollack, and Hauck 1996) or race (Hyman, Cobb, Feldman, Hart, and Stember 1954; Hatchett and Schuman 1975; Schuman and Converse, 1971; Krysan and Couper 2003). Biased measurements can also result when the simple presence of an interviewer evokes a social norm

such as social desirability (Hochstim 1967; Dillman and Tarnai 1991; Aquilino 1994; Tourangeau and Smith 1996; Tourangeau and Yan 2007; Kreuter, Presser, and Tourangeau 2008; Preisendorfer and Wolter 2014) or acquiescence (Schuman and Presser 1981; Javeline 1999) that changes how respondents answer. In contrast, self-administered surveys can be answered without others hearing the answer, including the interviewer or other household members, which helps minimize socially desirable and acquiescent responding (Schwarz, et al. 1991; de Leeuw 2005). As a result, respondents are more likely to give answers that cast them in a positive light in interviewer-administered than in self-administered modes (e.g., Hochstim 1967; Dillman and Tarnai 1991; Tourangeau and Yan 2007; Kreuter, et al. 2008). When compared to records, self-administered modes generate more accurate reporting of autobiographical sensitive information than interviewer-administered modes (Tourangeau and Yan 2007; Kreuter, et al. 2008; Preisendorfer and Wolter 2014). Additionally, several studies have found that respondents are more likely to agree with items in interviewer-administered than self-administered modes (Dillman and Tarnai 1991; Greene, Speizer, and Wiitala 2008).

Thus, surveys that transition from telephone to self-administered or mixed modes may see changes in their survey estimates (perhaps with increases in accuracy) related to socially (un)desirable issues or that are subject to acquiescence. For instance, Cernat, Couper and Ofstedal (2016) found that web respondents to the traditionally interviewer-administered Health and Retirement Study (HRS) had higher rates of endorsement of negatively rated items and lower rates of endorsement of positively rated items than interviewer-administered (telephone or face-to-face) respondents in a commonly used depression scale, even after accounting for the latent trait of depression. In a repeated cross-sectional Transgender Acceptance survey transitioned from telephone (in 2017) to web (in 2018) by Langer Research

Associates, the percent of respondents reporting being comfortable with transgender people declined 10 percentage points and the percent reporting being uncomfortable increased 17 percentage points with the move from phone to web. The online survey also produced a 12 percentage point increase in reports that students should use the bathroom that matches their sex at birth (Sinozich, et al. 2019).

Additionally, surveys that measure items potentially influenced by interviewer characteristics (e.g., race- or gender-related attitudes) are likely to see changes in response distributions when moving to a self-administered mode because interviewer characteristics will not be a cue for answering. Interviewer vocal characteristics and paralinguistic cues such as interviewer speaking speed also can affect respondent perceptions of interviewers and data quality (Charoenruk 2015; Charoenruk and Olson 2018). How exactly these changes manifest, however, depends on the composition of the interviewer and respondent pool in the interviewer-administered mode. For instance, web respondents in the ANES provided cooler responses to the Feeling Thermometer questions about various political figures and greater endorsement of Blacks and Latinos as lazy and as unintelligent, more racial resentment, and lower ratings on feeling thermometers toward racial groups than face-to-face respondents (Liu and Wang 2015; Abrajano and Alvarez 2019). Similarly, the Pew Research Center found that web respondents reported less satisfaction with their quality of life and they were less likely to indicate that minority groups experienced “a lot” of discrimination than telephone survey respondents (Keeter, et al. 2015). Importantly, there are notable differences in these mode effects on racial attitudes across respondent racial/ethnic groups (Keeter, et al. 2015; Abrajano and Alvarez 2019).

Interviewer variance occurs when different interviewers administer questions in different ways, leading to artificially high variation in respondent answers (Groves and Magilavy 1986; Fowler and Mangione 1990). This is most likely to occur when interviewers have more need or discretion to assist respondents such as on attitude, sensitive, ambiguous, complex, and open-ended questions (Schaeffer, Dykema, and Maynard 2010; West and Blom 2017). For example, Klausch, Hox and Schouten (2013) found less random measurement error, and thus more reliable measurements, on attitudinal items administered via self-administered web or mail surveys than in interviewer-administered telephone or face-to-face surveys in a mixed-mode experiment for the Dutch Crime Victimization Survey. When examining reports of depression in a mixed-mode HRS (face-to-face, telephone, and web), however, Cernat, Couper, and Ofstedal (2016) found no differences in reliability of measurement across modes. **More research is needed to identify exactly how and what variable errors change and on what types of questions when transitioning from a telephone survey to self-administered or mixed-mode study.**

4.1.2 Aural versus Visual Stimuli

Interviewer-administered surveys tend to be primarily delivered through oral communication channels. While visual cues such as body language and show cards can also be used in face-to-face surveys, interviewers and respondents have to rely entirely on aural stimuli in telephone surveys (Schwarz, et al. 1991; de Leeuw 2005). This means respondents have to hold the question and any response options in working memory while also generating a response, making such surveys more difficult from a respondent cognitive processing/working memory perspective and leading to more top-of-the-head responses (Schwarz, et al. 1991). This can be particularly difficult for respondents with lower cognitive abilities such as older respondents and

those with low education (Krosnick 1991). In contrast, mail and web-based self-administered surveys are primarily visual. For these modes to work, respondents have to be literate enough to read the questions and response options without the assistance of an interviewer (although in computerized self-administered modes audio reading of questions can be offered; Couper 2005). Additionally, self-administered surveys may be preferred or even necessary for people with hearing limitations but may be problematic for those with vision limitations.

One persistent mode effect is that **ordinal scale attitude/opinion items produce more extreme positive responses in interviewer-administered modes, especially telephone, than in self-administered modes** (e.g., Tarnai and Dillman 1992; Krysan, et al. 1994; Christian, Dillman, and Smyth 2008; Dillman, et al. 2009; Ye, Fulton, and Tourangeau 2011). For example, Dillman, et al. (2009) found that phone respondents were about twice as likely as mail respondents to choose the extreme positive response option when asked for overall satisfaction with their long distance telephone service. In an evaluation of web versus face-to-face respondents for the ANES, Liu (2018) found higher levels of reports of “favor” on a three point scale for a series of abortion-related attitudes for face-to-face respondents than for web respondents. Keeter and his colleagues (2015) found that phone respondents were less likely to use the extreme negative rating (“very unfavorable”) compared to web respondents when rating high-profile political figures. Several explanations have been offered for this mode effect, including primacy/recency effects, social desirability, acquiescence, differential cognitive processing of information obtained orally versus visually, and reluctance to give negative evaluations to interviewers, although tests of these alternative explanations are inconclusive (Krosnick and Alwin, 1987; Schwarz, Hippler, and Noelle-Neumann 1992; Dillman, et al. 1995; Ye et al. 2011; Dillman, et al. 2014). Whatever the mechanism, **surveys that transition ordinal**

scale opinion questions from telephone to self-administered or mixed modes will likely see less extreme positive reports on these items.

In addition to the benefits of visual communication, respondents can answer self-administered surveys at their own pace and do not have the social pressure to avoid long silences that might occur for respondents in a telephone survey (Schwarz, et al. 1991). This allows respondents to read questions and response options at their own pace rather than the pace set by the interviewer and have the time needed for recall and answer formation. For example, the American Community Survey (ACS) asks respondents how much they pay in real estate taxes, information that is likely not easily recalled from memory. Seeskin (2016) found that the difference in self-reported property taxes on the ACS versus administrative data were less variable for respondents to the mail questionnaire than for respondents to either the telephone or face-to-face mode, possibly because mail respondents can take the time to look up the information online or locate past statements. In a web test with highly cooperative respondents from the Panel Survey of Income Dynamics (PSID), more than half of the respondents report using records, and those that used records had web interviews that were about 27% longer than those who did not use records, compared to 11% longer in CATI (McGonagle, et al. 2017) Thus, **surveys switching from interviewer-administered to self-administered may see better quality of responses on autobiographical items that can be identified from records (for motivated respondents), although more research is needed to evaluate this hypothesis.**

One advantage of web and mail modes is that they allow researchers to take advantage of visual design to more effectively communicate with respondents. **Visual self-administered surveys allow for the use of graphics such as maps, ladders, smiley faces, or thermometers to try to help respondents understand questions that are not possible or very difficult to**

implement in telephone surveys. For example, in the National Household Transportation Survey (NHTS) transition, researchers were able to capitalize on the visual and dynamic nature of the web by integrating mapping functions (using Google Maps API) for the origin, destination, and shortest path distances of respondent reported trips (Federal Highway Administration and Westat 2018). Likewise, in the RECS, a question about the number of cooktops in the home proved problematic in the web/mail pilot because respondents misinterpreted the item as asking for the number of separate burners. After testing, this problem was resolved by adding a picture of a modern cooktop and wording the question, “An example separate cooktop is displayed above. How many separate cooktops do you have in your home? (Count the entire cooktop, not the number of burners. Do not include cooktops that are attached to an oven)” (Murphy, et al. 2015). The same survey also included images of CFL, LED, and incandescent light bulbs to help respondents accurately report how many of each type of bulb they have in their home. Both of these surveys were able to use the visual communication channel of self-administered modes to improve their data collection.

4.1.3 Computerized versus Not Computerized Instruments

When transitioning from interviewer-administered to self-administered modes, researchers often use a mix of web and mailed paper questionnaires. Although web questionnaires share many of the same features of programmed telephone or face-to-face instruments, paper questionnaires cannot accommodate many of these computer-assisted features. Computerization is important from a questionnaire design perspective because it allows automation and advanced design features. **When transitions involve the use of mail surveys,**

designers lose the ability to use a package of automation methods to assist respondents in that mode.

Skip patterns are ubiquitous across modes – in our survey of survey organizations that transitioned from telephone to a self-administered or mixed mode, virtually all (19 of 23) had skip patterns in both modes. A computerized questionnaire removes the responsibility for navigation from both interviewers and respondents, automatically taking people to the appropriate next question. Provided the programming is correct, in addition to the benefits of interviewers in helping navigation, automation can virtually eliminate navigation errors and considerably reduce interviewer training and workload. Computerization also can make it much easier to manage topical modules that apply to particular sub-populations.

Although a web survey can easily duplicate routing used in a telephone survey, mail surveys are much more constrained; skip patterns are limited to what can easily be conveyed to the respondent using text and graphics. In addition, respondents on the telephone cannot hear (and on the web, cannot see) what items are being skipped (or easily anticipate that their answers are triggering follow-up questions). In contrast, respondents filling out a mail questionnaire can see every item and may be discouraged from completing a lengthy-looking survey - even if many of the questions would not apply to them - or may choose answers that allow them to minimize the number of follow-up questions that they will receive (i.e., motivated misreporting). Thus, if mail surveys are going to be used at all in a self-administered or mixed-mode survey, the questionnaire may need to be simplified or abbreviated in order to avoid complex skip patterns (Berkthold, et al. 2018). For example, the ACS asks questions about marital status for all mail respondents, but implements an age-based skip pattern for telephone and face-to-face respondents (US Census Bureau 2014). Skip pattern errors include errors of omission

(not answering items that should have been answered) and commission (answering items that should have been skipped). Such errors may be more prevalent among those with lower levels of education or income or among youth who are less familiar with how to navigate a questionnaire (Redford and Hastedt 2011).

Possible solutions are to eliminate complex skip patterns in the paper version of the instrument, even if this means creating some repetitiveness for respondents or eliminating modules of items that are only reachable through complex skip patterns. The transition of the NHES from telephone to mail, for example, required the simplification or removal of many complex skip patterns that had been built into the CATI questionnaire. In the Parent and Family Involvement in Education component of the NHES, researchers decided it made more sense to move a set of questions about homeschooling into an entirely new topical questionnaire because leaving them in the main questionnaire would have required complicated skips and the questions only apply to about 3 percent of K-12 children (Chapman and Hagedorn 2009).

When responses to items in a set of related questions trigger follow-up questions, an additional consideration is whether to “interleaf” the stem (i.e., filter) and leaf (i.e., follow-up) items (asking the stem and leaf for each item before moving to the next stem) or to ask all stem items first, then move to the follow up items for each endorsed stem (grouped). What works well on phone may not work the same on web/paper. Major interviewer-administered surveys such as the Behavioral Risk Factor Surveillance System (BRFSS) and the Consumer Expenditure Survey (CE) use an interleaved format (Bureau of Labor Statistics n.d.; Centers for Disease Control and Prevention n.d.), whereas other major surveys such as the National Comorbidity Survey uses a grouped format (National Comorbidity Survey n.d.). Kreuter, et al. (2011) found that telephone respondents are more likely to affirmatively answer filter questions when the items are asked in a

grouped format rather than interleaved. In the interleaved format, they learn to alter their answers to filter questions in order to avoid follow-up questions later in the interview. In a web format, Mavletova and Couper (2016) also found that, if given an option, respondents will choose a strategy that will minimize their effort. This is a concern for surveys transitioning to mail for some or all of their data collection because the grouped format requires complicated skips that are really only feasible with computerization.

In addition to navigation, computerization opens up possibilities for customizing and personalizing the questionnaire such as by using information from a previous survey (Mathiowetz and McGonagle 2000; Jackle and Callegaro 2008; Lugtig and Jackle 2014), a previous answer in the same survey, or from the sample frame to create personalized routing and/or question wording. Although very effective and widely used in interviewer-administered and web questionnaires, fills are not possible on a paper questionnaire. A mail version of a survey requires more generic item wording, or construction of a version for each fill, which greatly complicates survey production and management. For instance, in the National Survey of Children's Health (NSCH), computerization is used for skip patterns, range checks, "pick lists," fills, required responses for screening questions, soft edit prompts, and online help screens in the web mode. On the mail questionnaire, researchers were able to include identifying information taken from the screener about the sampled child (name, initials, or nickname; age, and sex), but were unable to use any of the other automation tools (U.S. Census Bureau 2018b). In our survey of organizations that transitioned a survey from telephone to self-administered or mixed modes, 11 studies reported having fills in both the interviewer- and self-administered modes, 3 in only the interviewer-administered modes, and 6 in neither mode. **While there is not a specific**

literature on this, questionnaire designers should be thoughtful in handling fill language when converting from a telephone instrument to a paper instrument.

Computerization also allows real-time validation of inconsistent responses within a survey or between two surveys of the same person. Because of the lack of automated consistency checks and edit prompts in mail surveys, irresolvable inconsistencies can sometimes occur. For example, the 2012 NHES data file user's manual cites reports of children with both birth mothers and foster fathers at home and with age and grade mismatches, such as a 12 year old in 12th grade or a 17 year old in first grade (McPhee, et al. 2015). There are several ways these inconsistent reports can be dealt with such as treating them as missing, imputing new values, or, as was done in the NHES, leaving them in the data file for analysts to deal with on a case-by-case basis.

Surveys that transition to a computerized self-administered mode can take advantage of dynamic question formats in web surveys such as drag-and-drop questions where respondents can move items around to keep track of order (Blasius 2012), slider (or visual analogue) scales (Couper, Conrad, and Singer 2006; Funke, Reips, and Thomas 2011), and automatic calculation tools that keep running totals of numeric responses (Conrad, Couper, Tourangeau, and Galesic 2005). For example, as noted above, the web-based 2017 NHTS used the Google Maps API to calculate the distance traveled for each trip rather than self-reported distance traveled (Federal Highway Administration and Westat 2018).

Another benefit of computerization is that computerized assistive programs can allow those with vision impairments to still complete self-administered web surveys. This is important given Section 508 requirements that federal agencies ensure accessibility of their surveys for persons with disabilities (<https://www.section508.gov>).

Surveys that transition to a computerized self-administered web mode may experience lower item nonresponse rates than those that transition to a mail survey alone.

Previous research shows that mail surveys often have higher item nonresponse rates than web surveys (Israel and Lamm 2012; Lesser, Newton, and Yang 2012; Messer, Edwards, and Dillman 2012; Millar and Dillman 2012; Marken, Auter, and Marlar 2018). In web surveys, respondents can be prompted to give a response if they leave an item blank, a practice that has been shown to reduce item nonresponse (DeRouvray and Couper 2002; Al Baghal and Lynn 2015). Moreover, at least one study has shown that such prompting, when done immediately, can reduce item nonresponse to the same level as in face-to-face interviews (Al Baghal and Lynn 2015).

4.2 Device Differences within Web Modes

In addition to considering mode features when transitioning modes, telephone and web surveys have different types of devices that are used within modes to answer surveys. These devices may complicate questionnaire design when transitioning and the resultant measurement error and data quality. Smartphones are sometimes used to answer self-administered web surveys and provide a potentially different stimulus to the respondent than using a computer or laptop. Estimates of the percent of mobile phone web completes vary based on the survey topic and target population, but can be as high as 40% or more (questionpro.com n.d.). In addition, some respondents will answer web surveys on other mobile devices like tablets. The prevalence of mobile device usage in web surveys has led some to argue that all web surveys are mixed-device surveys (Toepoel and Lugtig 2015). The prospect of mixed device web surveys involving small mobile screens has implications for questionnaire design for any survey transitioning to a self-administered or mixed modes that contains a web component.

Although many surveys report collecting paradata on the device used to complete the survey (e.g., McPhee, et al. 2018), few of the studies that transitioned from telephone to a web mode contain indicators in public use data files for whether the survey was completed on a desktop or laptop, mobile phone, or mobile tablet. The PSID Well Being and Daily Life Supplement is an exception. It was collected via mail, web, and telephone, and the public use data file contains an indicator for both mode of interview and, for the web respondents, device used to complete the questionnaire, including whether the survey was logged into multiple times and the devices used for each log in (Freedman 2017). Furthermore, few of the studies that transitioned to self-administered modes that included a web component reported how the design changed on mobile devices and whether data quality differed for those who completed the questionnaire on a mobile device. Additionally, few studies that we reviewed provided screenshots of any part of the web instrument overall or the mobile instrument in particular. **As studies transition to self-administered modes that contain web, planning for questionnaire display and response on mobile devices in addition to that for desktop web instruments is critical. Screenshots of the questionnaire on both web and mobile devices should be captured and reported as part of methodology reports to allow data users to understand differences in questionnaire format and design across devices, and how these differences may have affected measurement quality.**

A few surveys report overviews of how a survey was optimized for mobile use. For instance, the ACS, in developing a mobile device version of the questionnaire, conducted multiple rounds of usability testing, providing screenshots of a few items from the usability testing. This testing included evaluating a “mobile optimized” browser, including removing a sidebar that appears on the desktop web to help with navigation, reducing banner width, having

larger text and less white space, and increasing the padding between response options. Even with these “optimized” features, usability testing identified several design features where the mobile instruments’ “optimized” version still failed to be usable (Olmstead-Halawa, Nichols, and Myers 2017). The National Survey of College Graduates (NSCG) had 4% of respondents complete the 2015 administration on a mobile device, and also increased font size and padding around response options in the mobile modes (National Academies of Sciences, Engineering, and Medicine 2018).

Empirical literature examining measurement and data quality differences across web devices is growing. Simply fitting the questions on the screen becomes a challenge (Peytchev and Hill 2010) as does minimizing design differences across devices and across modes (Smyth, et al. 2018). The two most consistent device differences reported within this literature are that **mobile respondents (especially smartphone respondents) break off at higher rates and take more time to respond than computer respondents** (e.g., Mavletova 2013; Buskirk and Andrus 2014; Keusch and Yan 2016; Lambert and Miller 2015; Antoun, et al. 2017b) .

Otherwise, there are few consistent measurement or data quality differences across respondents who answer via desktops/laptops and mobile devices. For example, response distributions tend not to differ across devices provided questions are asked in the same way (e.g., Baker-Prewitt and Miller 2013; de Bruijne and Wijnant 2013; Wells, Bailey, and Link 2014; Keusch and Yan 2016), including for socially desirable questions (Mavletova 2013; Mavletova and Couper 2013; Toninelli and Revilla 2016; Antoun, et al. 2017a). However, some mobile-specific question formats can be more problematic than the corresponding formats used for computers, for example, date pickers on mobile devices compared to month, day, and year drop down boxes on computers (Antoun, et al. 2017). Likewise, reliability and validity of answers

have not been found to differ across devices (Sommer, et al. 2016; Tourangeau, et al. 2017; Mavletova, Couper, and Lebedev 2018; Grady, Greenspan, and Liu 2019).

Differences in item nonresponse rates across devices are generally equivocal. Most studies report no difference in item nonresponse rates or the use of nonsubstantive response options (i.e., don't know, prefer not to answer, etc.) across devices (Mavletova 2013; Buskirk and Andrus 2014; Andreadis 2015; Revilla and Couper 2017; Schlosser and Mays 2017; Toepoel and Lugtig 2014; Tourangeau, et al. 2017; Olson, Smyth, and Phillips 2018). A few studies find small, but statistically significant item nonresponse rate differences, although not always in the same direction (Guidry 2012; Keusch and Yan 2016). Most studies report no differences in the length of responses to open-ended questions (e.g., Buskirk and Andrus 2014; Toepoel and Lugtig 2014; Schlosser and Mays 2017) or longer response on computers (Mavletova 2013; Peterson, et al. 2013; Wells, et al. 2014; Revilla and Ochoa 2016), but one study has reported longer responses on mobile devices (Antoun, et al. 2017a).

Nondifferentiation in battery items depends on both the device used and the format in which the battery is displayed – that is, displayed in a grid versus item-by-item – although there is little replication in which device and display have the highest nondifferentiation rates. Some studies find no difference in nondifferentiation rates across devices (Antoun, et al. 2017a; Revilla and Couper 2017; Tourangeau, et al. 2017; Liu and Cernat 2018; Maveltova, Couper and Lebedev 2018; Olson, et al. 2018; Grady, Greenspan, and Liu 2019). Still other studies find the highest nondifferentiation rates in smartphone grid formats (Baker-Prewitt and Miller 2013; Struminskaya, Weyandt, and Bosnjak 2015; Stern, Sterrett, and Bilgen 2016), and others find the highest nondifferentiation rates in computer grid formats (Peterson, et al. 2013; Lugtig and Toepoel 2016; Richards, et al. 2016). More work is needed here to understand how question

content, number of response options (e.g., Liu and Cernat 2018; Grady, Greenpsan, and Liu 2019), type of scale used, and number of items (Mavletova, Couper, and Lebedev 2018; Grady, Greenpsan, and Liu 2019) displayed in the battery affect answering across devices.

In sum, despite much worry about device differences for web surveys, the empirical literature to date has not shown large or consistent differences in the quality of data obtained from mobile versus web respondents. It must be noted that much of this literature is based on studies using volunteer panels and limited to people who have both computers and mobile devices (i.e., excluding those with only one or the other device). Additionally, few studies compare answers from mobile web respondents to responses via a mail survey. As such, general population studies that transition from a telephone mode to a web mode or a mixed-mode study that includes both web and mail may see different response patterns and data quality issues that arise than those found in prior work. To help future research on data differences across devices, **collecting information about the device used to complete the survey through paradata or respondent reports and including this information on public release files will facilitate understanding of how device of completion affects measurement quality.**

While they can pose challenges from a questionnaire design perspective, mobile devices create new measurement opportunities as described in the AAPOR Task Force Report on Mobile Technologies (Link, et al. 2014). These opportunities include the ability to collect location and activity data (e.g., Krenn, et al. 2011; Mavoa, et al. 2011; Wagner, Olson, and Edgar 2017), social network data (e.g., Boonstra, Larsen, and Christensen 2015), photos and videos (Gotschi, Delve, and Freyer 2009), and physical measurements (e.g., Gregoski, et al. 2013; Link 2013). In addition, data collection apps allow for more frequent and timely reporting, reducing recall bias, such as in time-use surveys (Lai, et al. 2010; Link, et al. 2014). For example, the National

Oceanic and Atmospheric Administration has developed apps to collect angler reports of catches for specific types of fish and for fishing vessel operators to submit Vessel Trip Reports (National Oceanic and Atmospheric Administration n.d.).

4.3 Additional Questions That Are Particularly Hard to Transition

Other questions may pose particular challenges when transitioning from telephone to self-administered or mixed-mode surveys. In our convenience sample survey, organizations were asked about the questionnaire design features of studies in both the original mode and in the transitioned mode. A few common types of questions were present in most of the responding studies before and after the transition. For example, 16 respondents reported having open-ended questions both before and after the transition. These questionnaire design features are summarized in Table 4.1.

Table 4.1 Questionnaire features of transitioned surveys

Number of respondents indicating presence of each feature

	In interviewer- administered only	In self- administered only	In both	In neither
Skip patterns	0	2	19	2
Matrix/grid Qs	2	6	10	5
Mark all that apply Qs	2	4	15	2
Interviewer-coded questions	9	2	5	6
Optional instructions	10	2	9	2
Fills	3	0	11	6
Explicit DK options	4	6	10	3
Open-ended Qs	1	4	16	2
Definitions	3	2	15	3
Long questionnaire	5	4	9	5
Multiple languages	1	1	11	10
Sensitive subject matter	1	2	10	10

Source: AAPOR Mixed Mode Task Force survey of organizations that have transitioned a survey across modes

4.3.1 *Numeric Reports and Complex Recall*

In interviewer-administered surveys, numeric data such as age, number of children, number of adults currently living in the household, dates, measurements (e.g., height or weight), amounts, or expenditures are relatively straightforward to collect (although are not necessarily easy to answer!). Speakers often clearly state units, which help interviewers be clear about the format of an answer (e.g., “twenty five dollars and sixteen cents”, “five feet, six inches”), and the verbal interaction allows interviewers to verify responses or follow up unclear responses. Interviewers also accurately enter numeric responses into the telephone or face-to-face instrument (Smyth and Olson Forthcoming).

Surveys that transition from telephone to self-administered surveys should carefully consider how to ask about numeric values, especially when web-based mobile devices or paper mail surveys are included. In a web-based survey, some of these types of items can be collected via drop-down question formats provided by the researcher. In mail surveys and for questions where the drop-down format is impractical in web surveys, well-designed answer boxes and good verbal instructions have been shown to drastically improve open-ended numeric reports (Couper, Traugott, and Lamias 2001; Christian, Dillman, and Smyth 2007a; Fuchs 2009a; Fuchs 2009b; Couper, Kennedy, Conrad, and Tourangeau 2011; Dillman, et al. 2014). Moreover, in web surveys, edit checks, placeholder examples within a box, and error messages can be used successfully to prompt respondents to provide properly formatted answers (Christian, et al. 2007a). However, even with these tools, open-ended numeric questions can be problematic on mobile devices because on-screen keyboards may take up considerable screen space, or the size of the number box is reduced, making it difficult for the respondent to see what they entered. For instance, in examining usability properties of a mobile instrument for the American Community

Survey, Olmsted-Hawala, Nichols, Holland and Gareau (2017) found that respondents had difficulty seeing numbers that were entered (thus misentering the number of zeros) and often missed a “.00” in the cents area of a number box.

No automated tools like validation checks or error messages are available in a mail survey, making administration of open-ended numeric items more complicated, especially for items for which there are multiple conventional ways of formatting (e.g., dates, monetary values, telephone numbers, etc.), those that can be reported in different units (e.g., income, height, etc.), or whether whole numbers are sufficient or decimal numbers are needed. Even with well-designed input boxes, respondents sometimes write in answers that are too vague or do not make sense to researchers, such as, for example, writing cents into dollar boxes or providing expense amounts that seem too high such that it is not clear if they are correct or if they are missing a decimal point (Breidt, et al. 2018; U.S. Census Bureau 2018b). For instance, one of the data editing steps for the mail-based Health Information National Trends Survey (HINTS) (Westat 2018) is to account for people reporting height in feet and inches (with two boxes provided) in the wrong box or with the wrong units (e.g., centimeters rather than feet). Additionally, the self-administered 2016 NSCH methodology report lists write-in items such as age, birth weight, BMI, year entered the U.S., and similar items as having higher item nonresponse rates than many other questions in the survey (6% missing or higher) (U.S. Census Bureau 2018). In the telephone administration of these same questions, the item nonresponse rates were less than 5% (BMI), and generally much lower (2011/12 National Survey of Children’s Health 2013).

In some cases, respondents may write illegibly or they may write outside of the answer box. Both of these errors make data entry more difficult and may inhibit the use of scanning. Even when the data are provided correctly, scanning vendors may charge for scanning open-

ended responses at a higher rate than closed-ended responses, which can increase the total cost of collecting these data. Finally, respondents may be less likely to enter a numeric response than they are to check a box (Olson, Smyth, Phillips, and Stenger 2019).

The challenges of asking open numeric questions in mail surveys is exacerbated when the items are particularly complex, such as those that require respondents to consult multiple data sources or to do complicated calculations. Thus in surveys that transition from interviewer administered to self-administered and include mail surveys, it may be necessary to limit such requests, break them down into smaller and simpler pieces, and/or provide considerable context to help respondents understand the task and complete it accurately (Redline 2011; 2013). An example of a complex numeric question that likely suffers from increased measurement error after the survey mode was transitioned from in-person to mixed modes (in-person, web and mail) is a question ascertaining the square footage of housing units in the RECS (Amaya, Biemer, and Kinyon 2018; Murphy, Biemer, and Berry 2018). The definition of what counts in this measure is rather complex, with some spaces (attics and garages) only being counted in specific circumstances (attics if they are heated, cooled, or finished and garages if they are heated or cooled AND directly attached to the housing unit) (U.S. Energy Information Administration 2017). In computer-assisted personal interviews (CAPI), interviewers could ensure that respondents understood what counted and did not count toward housing unit square footage for their particular home and could help respondents estimate it using the official definition. Moreover, in previous CAPI administrations and in the CAPI portion of the 2015 RECS (42.5% of completes), square footage was collected two ways – by respondent self-report and by interviewer-taken measurements (with respondent consent) (Amaya, et al. 2018; U.S. Energy Information Administration 2017). Comparisons of CAPI self-reports and interviewer

measurements from both 2009 and 2015 RECS indicate that respondents consistently underestimate the square footage of their homes by about 400 square feet with even larger underestimates for single family detached homes, even when interviewers were present to help them understand the instructions (U.S. Energy Information Administration 2017). During the transition, based on these findings and the danger that bias would be even larger in self-administered modes, researchers opted to report interviewer measures of square footage for CAPI respondents but to use imputed measures rather than self-responses for web and mail respondents.

4.3.2 Multiple Answer Questions

In telephone surveys, multiple answer questions are often asked in a forced-choice (yes/no) format, allowing interviewers to administer the items one-at-a-time and eliminating the need for respondents to remember all of the items at once. When converting to self-administered modes, these items can be converted to a “mark-all-that-apply” or “check-all-that-apply” format rather than the forced-choice format. The check-all format are thought to be easier for respondents to read and answer, and therefore reduce burden. In our survey of organizations that transitioned a survey from telephone to self-administered or mixed modes, 15 respondents reported having check-all-that-apply questions both before and after the transition, 4 had them only in the self-administered mode, 2 in only the interviewer-administered mode, and only 2 did not use them at all.

However, a growing body of research indicates that check-all-that-apply formats are subject to shallower cognitive processing and more satisficing. The forced-choice question format tends to result in endorsement of significantly more response options than does the check-

all question format because the items are processed more deeply (i.e., more optimal response behavior) in this format (Lau and Kennedy 2019; Smyth, et al. 2006; Thomas and Klein 2006). This phenomenon holds both across modes and within modes (Smyth, Christian and Dillman, 2008). As such, **multiple answer questions should be asked with a forced choice format when transitioning to self-administered or mixed modes of data collection.**

One potential problem that has been identified with the forced-choice format is the phenomenon of respondents marking answers only in the affirmative column and leaving the negative column blank, essentially treating the item as a check-all question. When this happens, it is unclear if the missing items were overlooked (i.e., truly missing) or intended to be “not affirmative” responses (note that this confusion about missing items is always the case in the check-all format - Rasinski, Mingay, and Bradburn 1994). For example, the 2016 NSCH methodology report indicates that the items with the highest missing data rates were forced-choice items about reasons needed health care was not received, sources of health insurance, and reasons for not having health insurance, and that the primary reason for the missing data was respondents only using the affirmative response option (i.e., treated the item as a check-all) (U.S. Census Bureau 2018b). Review of the identified questions, however, reveals that while the response options were formatted as forced choice, the question stems were written using check-all wording, a combination that previous research suggests can increase the likelihood of people treating forced-choice items as check-all items (Dillman, Smyth, and Christian 2014, but also see Smyth and Olson 2019). For example, the question stem for the item about reasons a child was not covered by health insurance in the 2016 NSCH 12-17 year old questionnaire is, “Indicate whether any of the following is a reason this child was not covered by health insurance DURING THE PAST 12 MONTHS:” (Data Resource Center for Child and Adolescent Health n.d.). This

question stem emphasizes the affirmative response option (“is a reason”), ignores the negative response option, and reinforces the impression that a check-all answering strategy is needed with the words, “any of the following”. A forced-choice equivalent of this question is, “Indicate whether or not each of the following is a reason this child was not covered by health insurance DURING THE PAST 12 MONTHS”. Unlike the check-all wording, this wording emphasizes the need for an affirmative or negative response (“whether or not”) for every item (“each of the following”) and thus should reduce the incidence of respondents treating it like a check-all question. The HINTS notes that editing is needed for a number of forced-choice format questions, where questions are presented in a grid with “yes/no” responses (Westat 2018). The items where this editing is reported also use the “any of the following” wording or fail to include directive wording at all “was there a time when you...”. Thus, **surveys that transition to a self-administered mode should be aware of the independent potential influence of the question wording along with the response option format for multiple answer questions and ensure that the question wording reinforces the response task dictated by the response option format.**

While the forced-choice format (with forced-choice wording) is generally recommended over the check-all format in both interviewer- and self-administered modes, a check-all approach may still be more appropriate for a limited set of items in a self-administered questionnaire. For example, when questions ask for factual information (e.g., race/ethnicity, language spoken at home, etc.) so that the primary response task is simply searching the list for the appropriate items (i.e., respondents do not need to read and consider every item), a check-all format is appropriate and less tedious than a forced-choice format. However, for check-all items, it is helpful to add a

"none of these" option. Otherwise, it becomes difficult to interpret these types of items when they are left blank.

4.3.3 *Open-Ended Narrative and Field-Coded Questions*

In interviewer-administered questionnaires, asking open-ended narrative (i.e., non-numeric) questions can be an effective means of gathering information without potentially influencing respondents by presenting predefined response options or for questions where a predefined list is unavailable. Sometimes, interviewers are asked to record the text of the response verbatim. For other items, “field coded” questions are used, in which respondents provide open-ended answers and the interviewer immediately interprets the response and categorizes it into one of the existing response options (although recording these accurately can be quite difficult; Smyth and Olson forthcoming). For example, the PSID uses this approach to categorize occupation by collecting information about the respondent’s major activities and duties, and then following up with additional questions and probes to ensure the responses are sufficiently detailed (McGonagle, et al. 2012). In our survey of organizations that transitioned, 16 used open-ended questions and 5 reported using field-coded questions in both modes (although how this was operationalized is unclear), with nine organizations using field-coded questions only in the interviewer-administered mode.

Open-ended items are more challenging in self-administered questionnaires for several reasons. First, without an interviewer present, there is no way to ask follow-up questions to clarify a respondent's answer (which may be unusable in its initial formation) or to redirect the respondent if they do not provide an answer to the specific question asked (McGonagle, et al. 2017). Secondly, because there is no interviewer to code open-ended responses into predefined

categories, a great deal of costly data cleaning post data collection is required. Third, respondents often write very brief responses or skip the open-ended item altogether. Finally, interviewers can encourage a response to sensitive questions or questions that raise privacy concerns. In fact, the lost ability to probe open-ended questions was highlighted as one of the problems experienced when transitioning to a self-administered or mixed modes of data collection in our survey of organizations that transitioned.

Field coded questions cannot be used in mail-based self-administered surveys, so the response options need to be converted into multiple open-ended questions that probe on relevant areas or to questions with closed-ended response options that will make sense to survey respondents and allow them to easily map their responses. In a web or mail administration, lists of categories can be administered via a series of closed-ended items that successively narrows the set of appropriate categories via skip patterns or a combination of open- and closed-ended questions (e.g., field of study in the (discontinued) National Survey of Recent College Graduates, Pierzchala, Wright, Wilson, and Guerino 2004; Tijdens 2014, 2015). For example, in the PSID, this was accomplished through maintaining three open-ended questions about the respondent's major activities and duties (rather than converting these to closed-ended questions). While nearly three times fewer characters were provided in the responses to the web instrument than the interviewer-administered instrument, levels of agreement of the occupational coding were high (McGonagle, et al. 2017).

One exception where retaining the open format is necessary is when researchers do not want to unduly influence respondents. For instance, if researchers want to measure what a respondent remembers hearing in the news yesterday, it may be better to ask an open-ended question. Another exception would be for questions in which no predefined response can be

populated such as re-contact information (i.e., email addresses or phone numbers). In cases when an open-ended question is the only available option, researchers should use visual design and motivational instructions in self-administered surveys to improve reports (Dillman, et al. 2014; Smyth, et al. 2009). Moreover, it has also been shown that structured probes that can be anticipated ahead of time can improve responses to open-ended questions in web surveys (Holland and Christian 2007; Oudejans and Christian 2011). However, in general, **researchers who transition from telephone to self-administered surveys should anticipate lower item response rates to open-ended questions and additional processing costs to retrieve information provided by respondents.**

4.3.4 Matrix or Grid Questions

Battery items, or individual items that share the same question stem and response options, are often converted to a grid or matrix format for self-administered questionnaires. Grids are an efficient visual format because they don't require respondents to read the same text (particularly the response options) over and over. They also take up less space in a paper questionnaire than writing each item out separately, which can reduce perceptions of questionnaire length and burden and save costs. In our survey of surveys that transitioned from telephone to self-administered or mixed modes, matrix or grid questions were used by most organizations, but six appear to have added them only after the transition to self-administered modes. When asked to elaborate on questionnaire elements that were particularly problematic in the transition, respondents mentioned concerns about administering grid items on the web because of people responding on mobile devices.

Several studies point to a reduction in completion time when grids are used, rather than single items (Couper, Traugott and Lamias 2001; Tourangeau, Couper and Conrad 2004; Callegaro, Shand-Lubbers and Dennis 2009; Toepoel, Das, and van Soest 2009). However, other studies, particularly of mobile devices, find that grids take longer to fill out than other types of questions (Couper and Peterson 2017) and can be particularly problematic for mobile device users (e.g., de Bruijne and Wijnant 2013; McClain and Crawford 2013; Peterson, et al. 2013). In a meta-analysis of break-off rates among mobile web respondents, for example, Mavletova and Couper (2015) found that complex grids increased the odds of the respondent breaking off the survey. However, as noted above, rates of straightlining or nondifferentiation are not consistently higher in a grid format compared to an item-by-item format across devices. **Yet, with higher breakoff rates, many recommend limited use of grids (Dillman, Smyth and Christian 2009) or finding ways to improve their design in order to mitigate their negative effects (Tourangeau, Conrad, and Couper 2013).**

4.4 Questionnaire Features That Are Hard to Transition

In addition to types of questions that are difficult to transition, there are more general survey features that can be challenging during mode transitions such as optional instructions, definitions, questionnaire length, and multiple languages. In our survey of organizations that transitioned surveys, 19 organizations reporting having skip patterns both before and after the transition and no organizations reported eliminating skip patterns in the transition (Table 1 above). Fifteen organizations reported having definitions at both time periods, three reported having definitions only in the interviewer-administered mode, and two reported having definitions only in the self-administered mode. While most organizations reported using optional

instructions (i.e., “if needed” messages), 10 had them only in the interviewer-administered version, not the self-administered version of the questionnaire; nine organizations had them in both. Nine studies reported that their questionnaires were longer than 20 minutes in both interviewer- and self-administered versions, while five reported this was the case only in the interviewer-administered version and four reported it was the case only in the self-administered version. In open-ended comments, respondents mentioned that transitioning interviewer-coded “don’t know” responses led to difficult choices about the provision of explicit “don’t know” responses in the self-administered versions and that skip patterns had to be simplified in several of the studies. We turn next to the difficulties of transitioning these types of questionnaire features.

4.4.1 *“If Needed” Information*

Interviewer-administered questionnaires often include “if needed” information that is provided only to those for whom it applies. Sometimes the interviewer has discretion over when such information is provided, such as with some definitions, clarifications, and instructions. Other times automation can be used to provide or not provide the information based on previous answers, which allows some “if needed” information to be used in web surveys. **For self-administered questionnaires, researchers need to decide whether “if needed” information should be included or not, knowing that in some cases, including the information means it will be there for everyone.** The use of italics, parentheticals, bolding, and other visual techniques can help differentiate this optional information from the main question text (Redline, et al. 2003; Christian and Dillman 2004; Tourangeau, Couper and Conrad 2004, Dillman and Christian 2005).

For example, one question in the RECS is about how many full bathrooms are in the household. A complication with this question is that depending on type of household, some respondents may need to be reminded to think about spaces they commonly overlook like finished attics or finished basements, but others do not need this reminder. In the interviewer-administered version of the RECS, the instruction to “Include bathrooms in finished attics or finished basements” is automatically added or excluded from the question depending on previously established housing type so that only those to whom the instruction applies are exposed to it. In the mail version of the questionnaire, this instruction is visible to everyone, regardless of housing type (U.S. Energy and Information Administration n.d.).

4.4.2 *Definitions*

In an interviewer-administered survey, definitions can be read to all respondents or to help only those respondents who exhibit signs of difficulty. In self-administered surveys, respondents themselves decide what information to read from definitions and when to read them. Thus, **an important and common decision when transitioning a survey from an interviewer-administered method to a self-administered method is where and how to display definitions.** Although common practice in some self-administered surveys is to include definitions at the beginning of the survey in a call-out box or glossary format, respondents are more likely to read and use definitions if they are strategically placed to facilitate use when/where they are needed in the instrument, for example at the actual survey item to which they pertain (Christian and Dillman 2004). For some items, it may be more effective to place the definition before the question stem than after it (Redline 2013).

We could not find a systematic evaluation of how surveys that transitioned from telephone to self-administered modes addressed the placement of definitions and how that varied across modes. **Some surveys that transition from telephone to mixed modes or self-administered surveys may strategically change the placement of definitions in a question.** For example, the telephone-based 2007 NSCH directly incorporated a definition of “specialists” into the first part of a question prompt, to be read to all respondents: “Specialists are doctors like surgeons, heart doctors, allergy doctors, skin doctors, and others who specialize in one area of health care. [During the past 12 months/Since [his/her] birth], did [Sampled Child] see a specialist [IF K4Q22 = 1, THEN INSERT: other than a mental health professional]?” In the 2017 self-administered version, this definition came in italics after the main question: “DURING THE PAST 12 MONTHS, did this child see a specialist other than a mental health professional? *Specialists are doctors like surgeons, heart doctors, allergy doctors, skin doctors, and others who specialize in one area of health care*” (US Census Bureau 2018). In addition, the 2015 ACS provided interviewers with a definition of who is included in the household, and interviewers ask a general question about whether there are people in the household who meet that definition. In contrast, the mailed ACS survey includes a definition of household membership in the front cover of the questionnaire, whereas the web mode turns those definitions into individual questions answered by the respondent (Clark 2017).

In web surveys, the presentation of definitions can be executed several different ways including their inclusion on every screen, a clickable reference, or a rollover feature whereby respondents are able to rollover a term to receive a definition. For example, in the web version of the ACS, definitions for the household roster and residence rules (“help text”) require clicking on specific help links at the top of the screen (Clark 2017). An analysis of the paradata for the web

version of the ACS indicates that fewer than 3% of respondents accessed any of the definitions during the household roster, and generally less than 1.5% of respondents accessed them at any point (Clark 2017). Definitions are more likely to be attended to if they are easier to access (Peytchev, et al. 2006; Galesic, et al. 2008). For example, Conrad, et al. (2006) found that few web survey respondents (about one in six) accessed definitions at all, and the more effort it required to get the definitions, the less likely respondents were to consult them. Fewer respondents opened definitions when it took two mouse clicks to access them than when it took just one. Those respondents who did obtain definitions might not have attended to the details of the definitions (Tourangeau, et al. 2006). Thus, **even with computerization available, definitions should be placed where they are needed and should be immediately available with no necessary user-action to access them.**

4.4.3 Long Questionnaires

Long questionnaires can be difficult to transition to a self-administered mode. In a lengthy telephone-administered survey, interviewers can be quite effective at encouraging participation in the survey, particularly when respondents show signs of fatigue or exhibit frustrations with the survey length. In a self-administered mode, respondents can complete the survey when it is convenient, in some cases even pausing and returning to lengthy surveys as time permits, but motivational prompts such as those used by interviewers cannot be used. In addition, respondents to self-administered questionnaires (especially mail) can see the entire questionnaire at once and thus may perceive the questions as more burdensome or intimidating than would be the case for the same questions in an interviewer-administered mode, potentially

leading to break-offs. For this reason, **a big challenge in transitioning surveys to self-administered modes is managing questionnaire length.**

Many surveys that transition from telephone to self-administered modes shorten the questionnaire. For example, the RECS transition shortened a 40 minute face-to-face survey to a 20 to 30 minute web and paper questionnaire by focusing on only the most critical content and asking for less detail in the self-administered modes (Murphy, Biemer, and Berry 2018). For example, whereas the interviewer-administered mode asked for information about up to three refrigerators in a household, the self-administered modes were capped at two refrigerators (U.S. Energy Information Administration 2017). Likewise, the transition of the NHTS reduced the number of response categories for questions about the purpose of trips and the means of transportation used (Federal Highway Administration and Westat 2018). The NHTS also attempted to reduce respondent burden by taking advantage of web technology in their trip rostering section. Since household members sometimes travel together, if one household member had already reported a joint trip, the other household members simply had to confirm and/or edit the details of the trip, saving them time and burden (2017 NHTS Data User Guide 2018). Similarly, the 2007 HINTS introduced a mail instrument to the existing RDD telephone survey, reducing the length of both from a 40 minute interview to a 30 minute interview (Cantor, et al. 2009). Others have attempted to deal with remaining questionnaire length issues after shortening surveys for a transition by offering the new version in two separate modules versus one longer survey. However, they found this to be an ineffective strategy as it decreased response rates and increased data collection time and costs (Liao, et al. 2019).

In some instances, efforts to shorten questionnaires have led to unanticipated problems in the self-administered data collection. For example, the National Center for Education Statistics

undertook an intensive review of questions in the NHES surveys as part of their transition process. This review involved identifying and dropping questions that were of secondary importance or that were too difficult for self-administration (Montaquila, et al. 2013). One result was that they asked fewer questions to verify homeschooler status on the screener questionnaire than had previously been used in the interviewer-administered screener, resulting in possible parent misreports of homeschooled children being in public or private schools in the 2012 NHES screener (McPhee, et al. 2015).

4.4.4 Single versus Multiple Languages

When surveys are designed and administered in multiple languages, interviewers help identify the need to administer a questionnaire in the appropriate language, to conduct the interview if they are bilingual interviewers in the appropriate language, or to ensure follow-up with a bilingual interviewer. Administering self-administered surveys in multiple survey languages is slightly more complicated than in interviewer-administered surveys. In a mail survey, researchers often send multiple versions of the same survey in different languages (e.g., the RECS, NHES, NSCH, ACS, etc.), or a dual-language survey, perhaps formatted as a “swim-lane” (side-by-side) questionnaire (e.g., 2010 U.S. Census [Rothhaas, et al. 2011]). The inclusion of multiple languages significantly increases the costs associated with printing and mailing mail survey packages, especially in cases when two or more alternative languages are required. In an effort to minimize the total number of survey packages printed, many researchers use sample information to predict the likelihood the respondent will require another language in order to participate in the survey (e.g., 2010 U.S. Census [Rothhaas, et al. 2011]; HINTS 4, Cycle 2 [Westat 2013]; NHES 2016 [McPhee, et al. 2018]). In a web-based survey administration, this

process is more efficient when respondents may select their own language, but translation and programming costs are still required.

In the surveys that transitioned from telephone to self-administered or mixed modes, most were administered in only English or only English and Spanish. From our survey of surveys that transitioned, multiple languages were offered for about half of the studies, the vast majority of which offered them in both interviewer- and self-administered modes. For example, the self-administered NSCH used only English and Spanish language materials because the prior telephone administration, which used a language-line service for languages other than English and Spanish, found that 0.2% of interviews were conducted in Mandarin, Cantonese, Vietnamese or Korean (Bramlett, et al. 2017; Ghandour, et al. 2018). The California Health Interview Survey web survey pilot was administered only in English and asked respondents to call into the telephone center for languages other than English, yielding only 11 non-English interviews (Wells, et al. 2018).

Respondents in languages other than English may experience challenges with an instrument that do not occur for respondents who complete the English-language questionnaire, either because of translation problems or because of other visual design problems. One challenge with bilingual self-administered questionnaires, such as the 2010 U.S. Census questionnaire, which included both English and Spanish side-by-side in a “swim-lane” design, is that respondents may not answer questions in a single language. Rather, they may enter responses in both languages, raising challenges for data entry and processing. In the 2010 Census, 3.4 percent of returned bilingual questionnaires had this problem (Rothhaas, et al. 2011). Web questionnaires can allow respondents to switch between languages, as was done in the web experiment of the 2016 NHES (McPhee, et al. 2018). As such, data users who want to know which language was

used to complete the questionnaire may need item-specific flags; alternatively, the survey organization may need to make a decision on how to assign language used. For example, in the 2016 NHES web experiment, language of interview in web surveys was identified as the language used for the last item completed in the questionnaire (McPhee, et al. 2018).

When transitioning interviewer-administered survey instruments into self-administered questionnaires in languages other than English, it is critically important to test and evaluate all parts of a questionnaire, including formatting and visual design, launch pages to a web survey, and question wording itself. These tests may reveal myriad problems that may be different in self-administered modes than in interviewer-administered modes. For instance, the Spanish-language respondents answering the mail questionnaire for the HINTS showed difficulties in completing grids (Westat 2013). Cognitive testing for Spanish-language respondents to the 2020 Census Barriers, Attitudes, and Motivators Survey (2020 CBAMS) revealed that the phrase “beginning the survey” on the web survey’s launch page was actually translated as “after the survey” (Lykke and Garcia Trejo 2018). Other issues may arise.

4.5 Collection of Biomeasures, Environmental Samples, Interviewer Observations and Consent for Administrative Record Linkage

Some measurements are facilitated by having an interviewer-administered survey (Kreuter 2013). Having interviewers present in-person, for example enables the use of show cards and card sorting measurement techniques and the collection of physical measurements of people (e.g., height, weight, etc.) or housing units (e.g., square footage) or physical samples (i.e., biological or environmental samples such as saliva, stool, water, and air). Interviewers can also make important observations about neighborhood characteristics, housing conditions or other

factors. Finally, interviewers can assist with the integration of less traditional types of measurement such as by installing passive data loggers (e.g., meters that measure television viewing, energy use, light, air quality, etc.) or obtaining record linkage consent. These measurement techniques are not possible in mail or web surveys and – in some instances – are more difficult on the telephone.

For example, as described above, interviewers in the RECS measure housing unit square footage, providing better measures than those that can be obtained through respondent self-report. They also record, based on their own observation, the housing type of the respondent and then a number of details about the housing (e.g., if it is an apartment, what floor it is on and whether the door opens to a hallway or outside); respondents have to be asked for this information directly in the self-administered version. Likewise, interviewers in the ANES made observations about respondent characteristics like skin tone, apparent intelligence, cooperation, suspicion, interest in the interview, and sincerity as well as took notes about any visible political or campaign signs at the residence (American National Election Studies 2015; 2018). In addition, interviewers are sometimes asked to collect administrative records or even to install passive data loggers. **Transitioning away from interviewer-administered and to self-administered modes raises challenges for interviewer observations as a critical part of data collection.**

One way to continue to collect observational or biological measurements when transitioning to self-administered or mixed modes is to send a separate observational team to collect the assessments, but consent rates may decrease substantially and more research is needed to minimize the losses. In the first four waves of National Longitudinal Study of Adolescent to Adult Health (Add Health), researchers collected extensive biological measures, including height, weight, BMI, DNA, pulse, and blood pressure and tested for sexually

transmitted infections, HIV, immune function, inflammation, and diabetes, requiring the taking of physical measurements and collection of blood, urine, and saliva samples. In wave 4 of data collection, measures and samples were taken by in-person interviewers in an approximately 30 minute procedure that took place immediately after the interview (Add Health Wave IV n.d.); 96% of respondents consented to providing saliva and 95% consented to providing blood samples (Harris 2018). In wave 5, the survey was transitioned from in-person to a mixed-mode design that started with web and mail data collection followed by telephone non-response follow-up. Researchers sought consent for the physical and biomarker collection during the initial web, mail, or phone survey and then had a biomarker subcontractor visit respondents for actual collection. Using this two-step process, consent rates were considerably lower, with only 66% consenting for the biomarker visit (Harris 2018).

The HRS also faced mode-related limits to collection of these types of measures. HRS started collecting physical and biomarker data in 2004 and since 2006 has used interviewers to collect breathing tests, hand strength tests, walking tests, balance tests, height, weight, waist circumference, blood pressure, saliva, and blood spots in their biennial surveys. Interviewers are also able to administer cognitive performance tests and to provide observations such as information about the mode of response, how much help respondents received with the interview and from whom, notes about respondent difficulties with the questionnaire, and notes about factors that might impact respondent recruitment in future surveys (Health and Retirement Study Questionnaires n.d.). In off-years HRS, has collected considerable data using self-administered modes, but is generally unable to collect physical measures, biomarkers, some cognitive performance tests, and interviewer observations in these efforts (Fisher and Ryan 2018; Health and Retirement Study Questionnaires n.d.).

Without interviewers to collect biomeasures, researchers are left with a few options. **One option for surveys that transition away from interviewers is to ask sample members to go to a clinic to give samples.** While collecting samples at a clinic maximizes the types and quality of samples that can be taken, this method requires sample members to be in geographic proximity to a clinic, is expensive, and is prone to low cooperation rates among sample members (Sakshaug, et al. 2015).

Some studies have attempted to collect biological samples via self-administration.

These studies typically have lower participation rates than those using interviewers, although not always (Sakshaug, et al. 2015). Participation rates to such requests vary widely from 15% to 92%, and likely depend on how the request is made, of whom, and what samples are collected (blood and urine tend to have more nonresponse than saliva and buccal cells [i.e., cheek swabs]) (Gatny, Couper, and Axinn 2013; Sakshaug, Couper, and Ofstedal 2010; Dykema, et al. 2017). For example, the Wisconsin Longitudinal Study was able to collect saliva samples from 54% of sampled participants via mailed saliva kits using a protocol that started with prenotice phone calls, followed by postal mail saliva kits, a reminder postcard, and a final reminder telephone call (Dykema, et al. 2017). In the Danish Nurse Cohort Study, Hansen, et al. (2007) found 76% to 80% of nurses asked to mail in buccal cells did so as did 72% of those asked to mail in saliva samples. This compares to 31% of those asked to go to a central location to have their blood venously collected. They further found that most self-administered samples of buccal cells and saliva contained the desired DNA; only 2.6% were failed samples containing no DNA. However, the DNA quality from the buccal cells was too low for genotyping whereas the DNA from the blood samples and about three quarters of the DNA from the saliva samples could be genotyped. Rylander-Rudqvist, et al. (2006) also found high rates of saliva sample returns among Swedish

men (80%) and high DNA quality in the samples. Clements and Parker (1998) similarly showed that concentrations of cortisol in saliva samples that were exposed to simulated postal mail conditions were virtually the same as those frozen within one hour of collection, and Durdiakova, et al. (2013) found that salivary testosterone levels were unchanged after 1 day, 1 week, and 1 month of sample storage at room temperature, 4°C, -20°C, and -80°C (i.e., neither storage time nor temperature degraded the samples). These studies suggest that saliva can be successfully collected via self-administered and mail-back methods in order to test common biomarkers like DNA, cortisol, and testosterone, although we are unaware of any studies directly comparing the quality of samples collected via interviewer- and self-administration in a population survey context.

There is some evidence that, at least among certain populations, blood samples can also be collected in self-administered surveys. In 2003, the HRS conducted a survey among people diagnosed with diabetes in which they mailed blood-collection kits to sample members with instructions to mail the completed blood sample to a lab. The blood completion rate for this study was 52% (Sakshaug, et al. 2015). The study demonstrated that it is possible to collect blood via self-administered means, but it is notable that the completion rate is considerably lower than the comparable 80% to 87% for interviewer-administered HRS surveys around the same time frame, even though the survey population is made up of people who commonly have their blood monitored or monitor it themselves (Sakshaug, et al. 2015).

While it is by no means exhaustive, in our review of surveys that have transitioned, we did not come across any that attempted to collect environmental samples such as soil, water, or dust using self-administered modes.

In addition to making interviewer observations and collecting biological and environmental samples, researchers have also begun to rely on interviewers to collect consent for administrative record linkage, such as linking medical records to survey responses. **Successful consent to record linkage is obtained at higher rates in face-to-face interviews than in any other mode of data collection.** As just one example, the HRS links survey data to Social Security Administration records on earnings and benefits, to the Centers for Medicare and Medicaid Services claims information, to Veteran's Affairs health care utilization information, and to the National Death Index for mortality and cause of death information (Fisher and Ryan 2018; Health and Retirement Study n.d.). Their in-person linkage consent rates range from 78 to 84% (Sonnegga, et al. 2014). Fulton (2012) reviewed 22 U.S. surveys conducted between 1982 and 2010 that utilized record linkage. Most of these surveys (18) were conducted using interviewer-administered modes. Those conducted in-person had average record linkage consent rates of 75% compared to 63% for those conducted by phone. Three of the surveys were conducted by mail; these three had a substantially lower average record linkage consent rate of 49%. These results are consistent with the findings of an experimental comparison of record linkage consent rates (to employment data) in the 2012/13 Legitimation of Inequality Over the Life Span (LINOS) panel survey in Germany. In this experiment, 94% of those responding to an in-person interviewer consented to the record linkage compared to only 54% of those responding by mail or web, a finding that held even after controlling for differential nonresponse across the modes. In addition, while the linkage consent bias was small for all modes, it was larger for the self-administered modes (Sakshaug, et al. 2017). **These results suggest that transitioning from interviewer-to self-administered modes can be problematic for record linkage. More work**

is needed to figure out how to increase consent rates for record linkage in self-administered modes.

4.6 Summary and Takeaways

- 4.6.1 Different survey modes have different features that affect what can be measured and how measures work. Major mode features of consequence for measurement are interviewer-versus self-administration, visual versus aural communication channels, and computerized versus not computerized instruments.
- 4.6.2 Surveys that transition from telephone to self-administered or mixed modes generally experience slightly higher item nonresponse rates in the self-administered modes.
- 4.6.3 Surveys that transition from telephone to self-administered or mixed modes may see shifts in their survey estimates related to socially desirability, acquiescence, ordinal scale items, or items that are related to interviewer characteristics.
- 4.6.4 More research is needed to identify exactly how and what variable errors change and on what types of questions when transitioning from a telephone survey to self-administered or mixed-mode study.
- 4.6.5 Surveys switching from interviewer-administered to self-administered modes may see better quality of responses on autobiographical items that can be identified from records, although more research is needed to evaluate this hypothesis.
- 4.6.6 Surveys with knowledge questions may see an increase in estimated knowledge when transitioning from telephone to web-based self-administered modes.

- 4.6.7 Visual self-administered surveys allow for the use of graphics such as maps, ladders, smiley faces, or thermometers to try to help respondents understand questions that are not possible or very difficult to implement in telephone surveys.
- 4.6.8 Questionnaire and question features that are particularly challenging to transition to mail surveys include skip patterns, fills, open-ended and field-coded items, and definitions. Thoughtful planning about implementing questions with these features, including the need to simplify questions and skip patterns, differ across web and mail modes.
- 4.6.9 Screenshots of the questionnaire on both web and mobile devices should be captured and made available with documentation of the questionnaire.
- 4.6.10 The empirical literature to date has not shown large or consistent differences in the quality of data obtained from mobile versus computer web respondents. The most consistent pattern is that mobile respondents (especially smartphone respondents) break off at higher rates and take more time to respond than computer respondents. To facilitate learning about possible device effects on measurement, surveys should collect the type of device used to complete the survey through paradata or respondent report, and include this information on public release files.
- 4.6.11 To the extent possible, multiple answer questions should be asked with a forced choice format. Surveys that transition to a self-administered mode should align the question wording to the response option format for multiple answer questions.
- 4.6.12 Because of higher breakoff rates, use of grids should be limited or their design improved to mitigate negative effects.
- 4.6.13 Many surveys that transitioned from telephone to self-administered modes shortened the questionnaire.

- 4.6.14 Most surveys that transitioned from telephone to self-administered or mixed modes were administered in only English or only English and Spanish. Testing of question wording, formatting and visual design, and other parts of a multi-language survey in all languages is critical.
- 4.6.15 Collection of a limited set of biomeasures or consent for administrative linkage is possible in self-administered modes, although consent rates are lower than in face-to-face studies. The range of the types of measures is limited, as neighborhood and environmental observations by an external observer are not possible through the self-administered mode itself.

5 Testing Strategies for Getting Questionnaires and Other Materials from One Mode to Another

Moving from one mode to another or to a combination of modes may yield significant changes to multiple features of the data collection instruments. Thus, key to these mode transitions is testing. Just as testing is important for the initial fielding of a survey instrument to understand whether accurate information will be collected, testing in surveys transitioning to new modes can also provide insight into the potential effect of new modes on data quality.

Anytime one transitions to a mode that changes communication channels (i.e., visual versus aural) or adds or takes away an interviewer or computerization, the changes in stimuli to the respondent are significant enough that testing will need to be conducted in the new mode(s). For example, surveys conducted primarily via telephone may rely on conversational norms that have no corollary in self-administered modes. Telephone surveys may also have complex routing and skips that are difficult or impossible to replicate on paper. In addition, some questionnaire features may be particularly prominent in self-administered modes, but not used in telephone administrations, such as the use of a grid in a self-administered mode for items that were administered individually in an interviewer-administered mode. Thus, even if question wording remains the same, changes in mode warrant additional questionnaire testing.

There are a variety of methods available for questionnaire development and testing (Presser, et al. 2004; Tourangeau, Maitland, Steiger, and Yan forthcoming). In our survey of organizations, most indicated that they performed questionnaire testing during the transition, primarily cognitive interviews, pilot tests and usability testing. Many organizations combined these strategies in order to evaluate the instrument and then the entire data collection protocol.

5.1 Expert Reviews

Many surveys that transitioned from telephone to self-administered modes report convening a panel of experts to help with this transition. These can be informal, such as when panels of experts are asked to review materials and provide feedback on the materials. Individual experts can also be asked to provide a list of problems they identified. In surveys that transition, expert reviews and panels contained both methodological and subject matter experts (e.g., Brick, Williams and Montaquila 2011; Wells, et al. 2018; Federal Highway Administration and Westat 2018; Ghandour, et al. 2018) or stakeholders broadly defined (e.g., Cantor, et al. 2009). How exactly these experts were used is not always described, but include: “evaluated various frame and mode options to supplement or replace the existing data collection methodology” (Wells, et al. 2018, p. 12), did “work to refine and revise selected content” (Ghandour, et al. 2018), or “were instrumental in shaping the design of the Pilot Study” (Brick, Williams and Montaquila 2011, p. 409).

Other surveys used a more formal expert review process. More formal expert reviews can be conducted when these experts use a standardized evaluation tool to evaluate the questionnaire. Three examples of such tools are the Question Appraisal System (QAS - Willis and Lessler 1999), Question Understanding System (QUAID – Graesser, et al. 2006), and the Survey Quality Predictor (SQP - Saris and Gallhofer 2007). For instance, the QAS was used to review the Residential Energy Consumption Survey (RECS) instrument (Murphy, Mayclin, Richards, and Roe 2016).

In general, there is little information available on the details of how expert panels are used in these transitions. Expert reviews can focus on many aspects of the design including whether the right constructs are measured, how individual measures will work, question order effects, respondent burden, navigation, or recruitment methods. Compared to other testing methods, they are quick, inexpensive, and easy to implement. They provide a good means to identify possible problems with a questionnaire (especially those related to retrieval and respondent burden problems linked to item-nonresponse and inaccurate reporting – Olson 2010), and thus can be very informative about what parts of the questionnaire should be prioritized for additional testing. Given these many uses, more research is needed on the most effective use of expert panels and expert review when transitioning a survey from telephone to self-administered or mixed modes, including the composition of these panels or experts (substantive; methodological; data users), the frequency with which the experts are engaged (monthly, quarterly, annually), when the experts are engaged in the transition process (before starting design decisions; after the study team has identified core content; etc.), the level of formal assistance from the experts (informal conversations; coding sheets), and more.

5.2 Cognitive Interviews

Cognitive interviews are a well-accepted and commonly used method of testing questionnaires, used to obtain qualitative information from potential respondents about the process they use when answering survey questions (Willis 2005, 2015). A typical cognitive test is conducted by a trained interviewer who solicits verbal reports from a respondent as they answer survey questions, using structured follow-up probes to gather information on specific steps in the response process (Willis 2015). For example, asking participants to tell what a

question is asking in their own words can help identify comprehension problems while asking them how easy or difficult it was to recall something can help assess recall challenges.

Many surveys transitioning to mixed modes conduct additional cognitive testing for new versions of questionnaires. For example, the Health Information National Trends Survey (HINTS) conducted three rounds of cognitive interviews for the 2007 phone administration of the instrument and three rounds of cognitive testing for the mail administration of the instrument (Cantor, et al. 2009). Because both modes were implemented that year, the computer-assisted telephone interview (CATI) cognitive interviews informed the design of the mail questionnaire, and the mail cognitive interviews focused on navigation issues, question formatting issues (e.g., indentation, font size), and other issues around visual layout. These cognitive interviews also asked respondents to react to the cover of the questionnaire.

One important consideration, especially for self-administered modes, is when and how the interviewer should interact with the participant during the interview. Think-aloud procedures and having the cognitive interviewer probe on a question-by-question basis ensures that participant thoughts are heard when they are occurring. However, real-time probing may also interrupt processing, leading to a participant experience and behaviors that are very different from how self-administered surveys are actually done (i.e., reading more closely than usual, paying more attention to instructions and definitions, etc.). Foregoing the think-aloud procedure and saving probes until the end of the interview (i.e., retrospective interviewing and probing) may mean some details are forgotten, but will keep the survey experience closer to field conditions and minimize interviewer impact on participants as they complete the questionnaire. It may also allow for more observation of usability issues (discussed below) such as navigation errors or difficulty registering responses. For even greater realism, the National Survey of

Veterans mailed cognitive interview participants a questionnaire and asked them to complete it and return the questionnaire to the survey organization; respondents were then called on the telephone to provide insights into their difficulties in completing the questionnaire (Westat 2010). Surveys often use concurrent or think aloud probes in initial rounds of cognitive interviews to explore understanding, while in later rounds, the strategy often shifts to retrospective probing in order to gain an understanding of how the entire instrument is performing (Willis 2005).

Cognitive interviews have also been used to test other important factors like the effects of visual design features in questionnaires, how well respondents navigate questionnaires, and how respondents process implementation materials (Dillman and Allen 1995; Sawyer and Dillman 2002; Dillman, Parsons, and Mahon-Haft 2004; for an overview of the extension of cognitive interviews to self-administered modes, see Dillman and Redline 2004). For instance, Martinez, Eggleston, Katz, and Morales (2018) used cognitive interviews to examine a series of mailings in the mixed-mode American Community Survey (ACS). To address the mixed-mode data collection of the ACS, “likely internet responders” and “likely paper responders” were recruited, based off of prior statistical analyses of ACS that identified demographic and other correlates. Cognitive interview participants evaluated five different packages of recruitment materials, reflecting the ACS mailing strategy, including letters, questionnaires, postcards, and envelopes. These interviews revealed not only what respondents paid attention to in the letters and envelopes, but also perceptions of the recruitment protocol as described in the letters (e.g., negative comments about having to wait three weeks after the first mailing for a paper questionnaire among those who did not have internet access).

Because cognitive interviewing is very labor intensive and burdensome for respondents, researchers often conduct a small number of interviews and focus the interviews on areas identified as particularly problematic rather than testing the entire questionnaire. **Surveys that are transitioning to self-administered or mixed modes often focus on areas where the instruments are substantially different across the old and new mode(s).** For instance, the 2016 American National Election Studies (ANES) reported that a “subset of questions ... in the post-election CAPI [computer-assisted personal interview] instrument” (DeBell, et al. 2018, p. 28) were included in cognitive interviews. The RECS conducted two rounds of cognitive interviews on selected items that were thought to have changed in meaning over time (e.g., questions about compact fluorescent lightbulbs) or would require changes in presentation across modes (Murphy, et al. 2016).

5.3 Web Probing

Transitions from interviewer-administered modes to self-administered modes are also occurring for cognitive testing methods. Recently, survey methodologists have used qualitative questions embedded in web surveys or “web probing” to ask follow up probe questions about respondents’ answers (Murphy, et al. 2016; Edgar and Scanlon 2017). For instance, Edgar and Scanlon (2017) provide examples from questions such as “How much have you spent on clothing in the past 3 months?,” which can be followed up with probes similar to those that would be used in interviewer administered cognitive interviews such as, “What types of clothing were you thinking of when you answered that question?”. The follow-up questions can be presented for all respondents or just for those providing specific answers, and can be

closed- or open-ended questions. Obviously, this method does not require a trained interviewer, but is limited to the probes that can be specified in advance.

This type of testing can be administered to larger numbers of respondents than are typically included in cognitive interviews, can be conducted very quickly and relatively cheaply, and can yield much more diverse participants than those typically used for cognitive interviews (i.e., recruit beyond those from the immediate geographic area of the cognitive lab) (Murphy et al. 2016). Scanlon (2018) used web probes among over 2000 respondents in a web panel to evaluate misunderstanding of a question on health insurance across age, education, and income subgroups. Respondents can be recruited from social media platforms or sources such as Amazon’s Mechanical Turk if testing is done outside the survey itself (Edgar and Scanlon 2017).

For surveys moving to a heavier reliance on web questionnaires, web probing may be a particularly useful method of evaluating whether the online instrument and questions are working as intended.

5.4 Usability testing

The use of cognitive interview methods for examining navigation issues and recruitment materials has increased with the increase in web surveys and is now often referred to as “usability testing.” The term “usability testing” comes from the more general website design and testing literature and refers to tests focused on a respondent’s ability to navigate a website and perform a task (Krug 2014). For web-based survey instruments, the task(s) include logging into the instrument, entering answers, navigating successfully through the instrument and submitting data. Usability tests might focus on the design of the log-in screen, scrolling, automated editing or warnings, and the survey submission procedure. For instance, Hunsecker (2018) used in

person and virtual (web-based) usability tests to evaluate problems with completing surveys and web panel enrollment forms online. These usability tests revealed traditional problems with question wording, and also problems with navigating the web forms and other question formatting issues. In paper questionnaires, navigation and skip patterns are a focus of usability testing. Given there is no interviewer to assist with these tasks in self-administered surveys, **usability tests are very important in surveys transitioning to self-administered modes and are best used as a complement to traditional cognitive interview methods that focus on the response process for questions.**

Increasingly, respondents are answering web surveys on mobile devices like phones or tablets. The limited screen size of smaller devices and alternative input methods (i.e., touch screens, scrolling, spin wheels, etc.) may have a substantial impact on the layout and formatting of instruments and respondents' ability to navigate through them. Indeed, **the most common remark from our survey of organizations transitioning regarding types of questions difficult to move to new modes was that grid or matrix questions became problematic for surveys that were likely to be completed on cell phones or mobile devices.** Usability testing across device types can help ensure that the design does not inadvertently introduce measurement error for subsets of device users. For example, in a test for the mixed-mode Consumer Expenditure Survey, Williams, et al. (2018) evaluated the usability of a web-based consumer expenditure diary for both desktop and mobile respondents. Respondents were allowed to select the device used to complete the web diary, with many not selecting a mobile device because they thought it would be difficult to use. As with a pilot test, the usability test also revealed differences in timeliness and the types of expenditures reported by respondents across devices.

Usability tests are especially important when transitioning to web modes for understanding aspects of technology that may not be uncovered when the questionnaire and its specifications are displayed on paper. For example, Omsted-Hawala, Nichols, and Myers (2018) identified technology related usability problems with the mobile version of the mixed-mode Decennial Census Test such as the mobile device keyboard covering up the answer categories and entry boxes and failure to default to a numeric mobile keyboard for fields requiring only numeric entry. These types of problems would not have been revealed in a paper-based test using questionnaire specification documents.

Usability testing may also be important when transitioning to self-administered surveys if the level of internet experience, literacy, or English language ability is expected to vary across respondents. That is, respondents who are not comfortable with computers or the internet or who have lower levels of literacy may have more trouble with web or paper instruments than those who use computers more regularly or have higher literacy levels, and thus should be included in usability tests.

Questionnaires that are transitioned to self-administered surveys and offered in multiple languages also require attention in usability tests. For instance, Olmsted-Hawala, Nichols, and Myers (2018) report that respondents who speak languages other than English may have browsers that automatically translate survey login pages or questionnaire forms, even if the survey organization has a translated version available. Furthermore, usability tests revealed that locating a toggle button for a questionnaire in Spanish or another language far away from entry fields poses problems with respondents finding this button; similarly, those with limited familiarity with the English language may have difficulty with even the initial task of entering a

URL into a web browser. Each of these aspects should be tested across multiple populations when transitioning a survey to mixed modes, and especially those that include the web.

5.5 Field Tests

Field tests, also known as pilot tests, are small-scale studies of the entire survey procedure, including implementation materials and processes and the questionnaire itself, and yield empirical data about the new design under real survey conditions (i.e., in the field, with actual target population members). Field tests can give realistic estimates of item nonresponse rates, response distributions, and skip errors in the questionnaire, although often they cannot reveal problems with question comprehension. In computerized modes, they can also provide paradata to help understand survey question timing and rates of answer changes. For implementation, field tests provide information about response rates, sample composition, timing, costs, staffing needs, staff communication and coordination, and the effectiveness of field monitoring systems, but are costly. Field tests in mixed-mode surveys also facilitate comparisons of responses across modes; if the modes include the web, responses can be examined across devices within web respondents as long as those paradata are collected. These comparisons allow the researcher to evaluate the impact of using multiple modes or of multiple response devices on estimates, data quality, timing, and costs. For instance, when transitioning the National Longitudinal Study of Adolescent to Adult Health (Add Health) from interviewer-administered to mixed mode, Biemer, et al. (2018b) conducted a small pilot test prior to a larger implementation to evaluate response to a web-only survey and the quality of sample members'

email addresses. This pilot test informed the design of the study and possible experiments for a larger scale implementation.

One strategy when moving from a single to mixed-mode survey is to try to maintain comparability to the existing modes. As such, field tests may involve simultaneous fielding of the new modes and the old mode to evaluate how responses change with the change in design. This is expensive, and as an alternative, some surveys compare a field test in the new self-administered or mixed modes with the most recent implementation of the survey in the interviewer-administered mode. For instance, the Panel Survey of Income Dynamics (PSID) compared the implementation of a web instrument in 2016 with the most recent telephone administration of the instrument in 2015 (McGonagle, Freedman, and Griffin 2017), focusing on differences in questionnaire and section length, important survey estimates, and the ability to code answers about work and occupations from narrative open-ended questions. Brick, Williams, and Montaquila (2011) compared response rates, household eligibility, and demographic characteristics for a 2009 pilot study for the National Household Education Surveys (NHES) with the most recent telephone administration in 2007. Link, et al. (2008) compared response rates, demographic characteristics, and costs for a mail pilot of the 2005 Behavioral Risk Factor Surveillance System (BRFSS) with the 2005 telephone survey being done in the same states at the same time.

5.6 Experiments

Experiments assign different versions of the same question or implementation feature to random subsets of the sample. Experiments can help researchers determine which changes in their design will matter and how much. **When transitioning from telephone to self-**

administered or mixed-mode surveys, many experiments are conducted as part of field tests (if the field-test sample is large enough), but they are also often conducted within production surveys. While production surveys aim to optimize quality and cost trade-offs for each mode, experimental research on its own or within a production survey aims to optimize equivalence of design elements that are not part of the experimental variation in order to isolate the effect of a specific feature on the outcomes of interest. Experiments allow researchers to quantify the effects of alternative versions of a questionnaire or implementation procedure, but a weakness of some experiments, especially for questionnaire design and measurement purposes, is that they sometimes do not reveal the underlying cause of the difference, leaving it unclear which of two versions is “better”. The findings from experimental research build up the design principles and theory underlying potential differences that may be observed in self-administered and mixed-mode surveys compared to telephone surveys, while the findings from production surveys are empirically based and decisions are based on subjective assessment of what is important, what works, and what is available in terms of the survey budget and resources.

One important decision to be made in mixed-mode experiments is whether to experimentally assign sample members to modes/devices or allow sample members to self-select these modes and devices. In theory, random assignment ensures that differences found across modes/devices is due to the modes/devices themselves and not self-selection, although in practice, differential nonresponse across modes/devices undermines the random assignment. For instance, the Pew Research Center experiments (e.g., Keeter, et al. 2015) randomized members of their ongoing American Trends Panel to phone or web, but both arms of the experiment had nonresponse, opening the door to compositional differences. Additionally, randomly assigning sample members to particular modes and devices may systematically exclude large portions of

the population (e.g., those without internet access; those without both desktop and mobile devices), making generalization more difficult. In most mixed-mode survey experiments for studies that transition from telephone to mixed modes, however, selection into modes occurs through the mixed-mode design itself. For instance, when transitioning the Gallup-Sharecare Well-Being Index survey from telephone to web and mail modes, Marken, Auter, and Marlar (2018) randomly assigned sample members to a mail only condition, a simultaneous web and mail condition, a sequential mail-web condition, and a sequential web-mail condition, allowing respondents to self-select mode in the various mixed-mode conditions.

5.7 Packages of Testing Strategies in Surveys that Transitioned

Many surveys that transitioned from interviewer-administered to self-administered or mixed modes used a package of testing strategies during this transition. For instance, the RECS faced a number of challenges in their transition from in-person to web and paper modes, including a short timeline for testing to determine the design that would be used in the production survey. As such, they adopted a multi-phase testing approach in which the best features of the testing phases were built into future tests and the final production design on a flow basis (Murphy, Biemer, and Berry 2018). Testing started with expert review of the questionnaire, in-person cognitive interviews, and online self-administered cognitive interviews (Murphy, et al. 2016). The RECS in-person cognitive interviews (2 rounds with 15 people each from three cities) focused on particularly challenging content, for example, questions about new technologies, revisions of outdated or previously problematic questions, and “mode sensitive” content. The online cognitive interviews were similarly focused, although they also included

some updates made based on the in-person cognitive interview findings. Problems and changes identified in early testing were addressed and retested in later testing (Murphy et al. 2016).

A series of field tests were then conducted to test the feasibility of collecting energy use data via self-administered modes and to experimentally test questionnaire length and initial mode assignment (Murphy et al. 2018). The first test focused on a subset of localities, showing that a 30 minute self-administered RECS survey was feasible for both web and mail modes with respect to budget, timing, and response rate; that most people preferred the mail mode; and that the web mode produced higher quality data at lower costs. The second field test, a national level test designed while the first field test was still in the field (using daily tracking results) and conducted alongside the 2015 RECS CAPI data collection, adopted the materials and strategies that worked in the first field test but also included experiments with incentives and mode to try to push respondents to the web. Four key metrics were participation rates, web response rates, respondent sample representativeness, and costs per completed case (Murphy et al. 2018). These were monitored on a daily basis and had the biggest influence on decision making as the testing progressed and on the final production design.

Similar to the RECS, the transition of the NHES surveys from telephone to mail also involved considerable testing that started with a comprehensive review, redesign, and three rounds of cognitive interviews and ended with two pilot field tests (Westat 2009; Montaquila, Brick, and Kim 2012; Montaquila et al. 2013). The cognitive interviews tested recruitment materials, screener questionnaires, and topical questionnaires, with early interview findings informing changes that were tested in later interviews (for design details and findings, see Westat 2009). Two pilot field tests were subsequently conducted, one in 2009 (n=11,800 – see Brick, Williams, and Montaquila 2011) and another in 2011 (n=60,000 – see Montaquila et al. 2013).

These pilot tests included experiments on prenotice letters, incentives, questionnaire design, postal delivery methods, Spanish language materials, and envelopes. Key outcomes from the field tests included the screener response rate, eligibility rate, topical response rate, overall response rate, number of eligible households required to get one screener completed with an eligible household, number of eligible addresses required to get one completed topical survey, respondent characteristics, and costs (Montaquila et al. 2012).

5.8 Tools to Evaluate Questionnaire Features

Assessing the impact of mode transitions on estimates and data quality is complex. It requires researchers to plan ahead, identifying outcomes or metrics that will be used in such assessments, and ensuring that the proper information is collected to analyze these outcomes. Advance planning for the types of information to capture and the analysis to use will facilitate evaluation of the impact of transitions on the quality of measurement. Any analysis of the changes of mode of data collection may be confounded by the nature of who responds via different modes, whether through self-selection into response mode or differential nonresponse across response modes. All analysis of mode of survey, and especially those for which mode is not experimentally manipulated, must wrestle with these confounding factors. Chapter 8 on Survey Estimation addresses these estimation issues in more detail.

5.8.1 Data to be Collected and Associated with a Response or a Question

Surveys that transition from interviewer-administered modes to self-administered modes may want to plan to capture certain data as part of the data collection effort for facilitating analysis either by the data collector or secondary analysts of the data. Survey

researchers who collect their own data may need to identify systems that can collect this information. Researchers who contract to another organization to administer the survey may want to include these items in the data collection contract. This list aims to be comprehensive, recognizing inherent challenges in measuring or capturing some of the information at different survey organizations and with different data collection instruments.

Mode of response. Where multiple modes will be used, an indicator of which mode (telephone, paper, web-based) each respondent used is crucial to permit analyses of outcomes across modes.

Device used for responding. For surveys conducted by web, information about response device type (desktop, laptop, tablet, smartphone) is needed to allow comparison of responses and data quality across devices. Device type can be collected simply by asking respondents for a self-report. To fully understand the nature of the responding device, researchers may want to collect as much information as feasible and practical about the responding device (e.g., iPhone 6SE, screen size, and screen resolution). Such detailed information may yield more insights than simply categorizing the devices into broad-based categories (e.g., tablet vs. smartphone), although most existing analyses simply focus on these broad-based categories. Device type information can be recorded as paradata in what is called a “user agent string” (i.e., a string of text that identifies information about the responding device like resolution, operating system, browser, etc.). For an overview of collecting device type via paradata, see Callegaro (2010). For longer surveys, capturing device type at multiple points in the survey may be needed to evaluate whether respondents switch devices partway through.

Question characteristics. Most surveys document the wording of questions. Researchers can refer to this documentation to understand question features like the question wording, whether it is open or closed, or how many response options there were. This documentation often does not capture other questionnaire design features that can impact the quality of measurement. In visual modes where the graphical display of the question can communicate meaning to respondents and in mixed-mode surveys in which the questionnaire design has been optimized for display depending upon the mode and device used, simply documenting question wording and response options is insufficient. Design features such as scale orientation (vertical vs. horizontal), the use of verbal analogs (end points only or all points, regardless of device used for data collection), placement in a grid versus presentation on separate screens and a number of other visual design features can also influence responses (for an overview see Dillman, et al. 2014). These features are best captured by retaining production copies of the paper questionnaires and screenshots of web questionnaires on desktop and mobile devices.

In our review of surveys that have transitioned to self-administered or mixed modes, we were almost always able to find copies of paper questionnaires, but almost never able to find documentation of how the survey appeared on the web when that mode was used. Thus, in general, **better documentation and dissemination of screen captures of web and mobile surveys is needed.** In the case of web especially, once the study is out of the field and a little time has passed, it may be difficult, if not impossible to reproduce the questionnaire how respondents saw it because of technological changes. Thus, **it is paramount that web and mobile screenshots be taken during the field period to provide accurate documentation.**

Paradata. To the extent that data are collected via computerized means, auxiliary data about the data collection process can inform post-data collection analysis. These auxiliary data may include keystroke data both for interviewers as well as self-administered respondents, use of “help” screens, and timing information for any particular screen, questionnaire section, or the entire interview (Kreuter 2013; Olson and Parkhurst 2013). Surveys that transition from a computerized interviewer-administered mode to a computerized self-administered mode can use this information to understand differences in questionnaire and section length, as well as other particular problems encountered by respondents during data collection.

Interviewer Information. If interviewer administration is retained as part of the mix of data collection modes, information about interviewers is needed. At minimum, this should include an anonymized interviewer identification number for each case so researchers can nest respondents within interviewers. Interviewer ID numbers allow investigators to uncover some of the error in interviewer-administered surveys by examining interviewer bias and variance in responses (Groves 2004; Fowler and Mangione 1989; Elliott and West 2015). When possible and when there is no danger of identifying individual interviewers, additional interviewer characteristics such as race, gender, age, overall interviewing experience (i.e., tenure), and experience within the specific study (i.e., within study interview count) also may yield insights into potential interviewer-related error in a mixed-mode context (e.g., Catania, et al. 1996; Krysan and Couper 2003; Olson and Peytchev 2007).

5.8.2 *Possible Analysis for the Evaluation of Mode Effects*

Chapter 8 provides an overview of types of analyses that focus on diagnosing nonresponse and measurement errors in mixed-mode surveys. Here, we suggest additional analyses that survey researchers can conduct to evaluate the quality of data collected in different modes.

Item nonresponse rates. Self-administered modes tend to lead to higher rates of item nonresponse (Nicolaas and Tipping 2006) as well as the loss of information concerning the nature of the item nonresponse (refusal vs. don't know). Mail questionnaires generally have slightly higher item nonresponse rates than web (e.g., see Survey Practice, Volume 5, Issue 2) and, in the case of paper instruments, item nonresponse rates may be higher due to noncompliance with skip patterns. These differences should be expected as a matter of course, but exceptionally high item nonresponse rates overall or to individual questions may indicate other problems with the questionnaire design in one or the other modes that should be further explored through testing. Chapter 4 discusses differences across modes in item nonresponse rates.

Response distributions. As described above, there are many reasons to expect differences in response distributions across modes, such as differences in social desirability due to interviewer presence, extreme positive responses in interviewer-administered modes, and differences due to automation or lack thereof. Sample composition differences may also lead to differences in response distributions across modes. Chapter 8 deals with these analyses in much more detail.

Open ended questions. Comparisons of responses to open-ended questions across modes can focus on either content of the responses or the quality of the responses, both of which require considerable data processing. With respect to content, researchers can examine whether the same substantive themes or ideas occur across different modes, a task that will require qualitative coding. With respect to quality, researchers can compare the amount of information collected, which can be operationalized as character or word counts, as a count of the number of themes (i.e., independent ideas that answer the question), or whether there was any elaboration (description or expansion on a theme) in a response. If audio recordings are available of interviewer-administered questions, researchers can also compare the response given by the respondent to the response recorded by the interviewer to assess interviewer accuracy in keying responses. Differences in responses to open-ended questions is described in Chapter 4.

Nondifferentiation. For items that appear in a series, a common measure of data quality is nondifferentiation, or the extent to which answers are the same (i.e., not varying) across the items. There are a number of different operationalizations of nondifferentiation (see Kim, et al. 2018 for an overview). The strictest is the straightlining rate, which is the percent of respondents who gave the exact same answer to all items in the series. Others such as the within-respondent standard deviation across the items in the series are less strict. Generally it is assumed that nondifferentiation is a form of satisficing or respondents shortcutting the response process; however, the content of a specific series of items one is assessing also influences nondifferentiation rates. For example, we might expect variation among a set of items about how often one does different recreational activities, but we would expect very little variation among a set of items about how often one engages in different criminal activities. In the latter case, most

respondents are expected to select “never” for most items with the resulting nondifferentiation representing high quality responses, not measurement error. Differences across modes and devices in nondifferentiation rates is discussed in Chapter 4.

Response Order Effects. Response order effects can be an indication of measurement error. For example, if responses to ordinal scalar items are more highly sensitive to scale order in one mode than another, this could be an indication that respondents in that mode are reacting less to the content of the response options themselves and more to more formal features like the presentation of the scale. If one mode is more prone to primacy (i.e., higher selection of items appearing first regardless of their content), it is often assumed the mode is more error prone (i.e., respondents are satisficing or misunderstanding the scale order). Likewise, large differences in endorsement of nominal items based on their position in the response options can be an indication of respondent confusion or misunderstanding of response options. These types of comparisons require experimental designs where response option order is varied in the same way (inverted, randomized, etc.) within each of the survey modes being used. Without such designs, it is impossible to differentiate the effects of content from position. Differences in primacy and recency effects across modes is discussed in Chapter 4.

Response time. Response time is often used as a proxy for measurement error (Yan and Olson 2013). Response times that are too fast may indicate problems with the administration or answering of the questions, such as that respondents did not carefully think about their answer. Similarly, response times that are too slow may indicate that respondents were confused or

distracted. Generally, differences in response times may be observed across computerized modes, but substantial differences may indicate a problem with questionnaire design.

Reliability. Reliability can be assessed several different ways, depending on the types of items and data at hand. Internal scale reliability (typically Cronbach's Coefficient Alpha) for a set of related items can be compared across modes. Previous studies comparing scale reliability across modes have found few differences (de Leeuw 1992; Borkan 2010). Additionally, increased scale reliability may, in some cases, reflect increases in correlated measurement error rather than improved measurement (Peytchev 2007). Another way to assess reliability over time is with test-retest or repeated measures designs such as those used by Cernat (2015), Chang and Krosnick (2009), Braunsberger, Wybenga, and Gates (2007), and the multitrait-multimethod experiments reported by Saris and Gallhofer (2007a,b). Yet a third is to examine the extent of random error in a measure under the rationale that less random error yields higher reliability (Klausch, et al. 2013). Assessments of these types generally find few differences in reliability between interviewer-administered modes and few differences between self-administered modes, but self-administered modes have higher reliability than interviewer-administered modes.

Validity. Validity can also be assessed in different ways, with each requiring different types of data. Concurrent validity can be assessed by predicting the relationship between two measures taken in the same survey that are theoretically related and thus should be highly correlated. Previous research has shown little difference in concurrent validity between interviewer-administered modes (Jackle, Roberts, and Lynn 2010), but a slight advantage for web panel responses over random digit dial (RDD) telephone responses (Chang and Krosnick 2009).

Validity can also be assessed by measuring the extent to which a time 1 attitude or behavior predicts a time 2 attitude or behavior that it should predict. For example, Chang and Krosnick (2009) show that pre-election candidate preferences were more strongly related to reported vote choice in web panel responses than in RDD telephone survey responses. The gold standard measure of validity is a record check study where self-reports can be compared to high quality records. In an early meta-analysis, de Leeuw (1992) found no difference in record check validity across modes.

5.9 *Summary and Takeaways*

- 5.9.1 There are a number of testing methods available to help with transitioning modes. Each yields a different type of information and thus is appropriate at different phases of the transition. Generally, extensive testing using multiple methods will be needed to make the transition as strong as possible; previous surveys that transitioned have used a combination of expert review, cognitive testing, usability tests, and field tests.
- 5.9.2 Evaluating the effect of a transition on measurement and data quality requires forethought and planning to ensure that the necessary data are collected to make desired comparisons. Among other things, this should include response mode, response device, question characteristics, paradata, and information about interviewers if they are utilized.
- 5.9.3 Such evaluations can examine item nonresponse rates, response distributions, quality of open-ended responses, nondifferentiation, response order effects, response time, reliability, and validity. Researchers need to ensure prior to data collection (whether a field test or production data collection) that they have the right design for the evaluations they choose to conduct.

6 Recruitment, Nonresponse, and Operational Issues

Response rates to household surveys in the US and around the world are falling, both for face-to-face surveys (e.g., Williams and Brick, 2018) and for telephone surveys (e.g., Lavrakas, et al. 2017, Appendix D). Transitioning to a mixed-mode data collection can help improve coverage and reduce nonresponse (de Leeuw 2005; Cornesse and Bosjnak 2018). Research from a number of different countries show that using multiple modes to contact sampled units can improve response rates and potentially reduce nonresponse bias because different types of respondents are more or less likely to respond to certain modes (Messer and Dillman 2011; Bandilla, Couper, and Kaczmirek 2014; Dillman, Smyth, and Christian 2014; Kappelhof 2015). In our convenience sample of surveys that have transitioned from interviewer-administered to self-administered modes, 12 of 22 organizations reported that declining response rates to the interviewer-administered survey was extremely important in their decision to transition to a self-administered or mixed-mode survey, and 10 organizations reported that anticipated response rates to the self-administered or mixed-mode surveys were extremely important in their decision to transition. Of those 17 organizations that reported on what actually occurred to response rates as part of the transition, 5 reported that the survey response rate decreased with the transition, 5 reported that the response rate stayed about the same with the transition, and 7 reported that the response rate increased.

Figure 6.1 displays response rates from a set of surveys conducted in the US that have examined transitioning from interviewer-administered modes to self-administered modes, focusing only on one-stage (i.e., no screening) surveys where the self-administered mode survey was conducted within two years of the most recent interviewer-administered survey to help with

comparability of the essential survey conditions for the two administrations. The figure orders the surveys by the year of the transition study and separates studies that examined concurrent mixed-mode designs, sequential mixed-mode designs, and single mode designs (either mail only or web only). Some of these comparisons are experimental (interviewer- and self-administered modes mounted at the same time) whereas others are observational (self-administered mounted at a different time, limited here to those with no more than two years between the interviewer- and self-administered surveys, or one mode used as a follow-up mode for another mode). The response rates are taken directly from the available reports or articles, and thus some are AAPOR Response Rates (RR1 and RR3 are common) whereas others are CASRO Response Rates (citations in appendix table 6.A). Many factors vary across the studies. **Yet patterns can be easily observed. In the one-stage surveys conducted between 2001 and 2012, response rates to the telephone mode tended to be higher or at about the same level as response rates to the self-administered or mixed modes. After about 2013, response rates to the self-administered or mixed modes tended to exceed those for the telephone mode.**

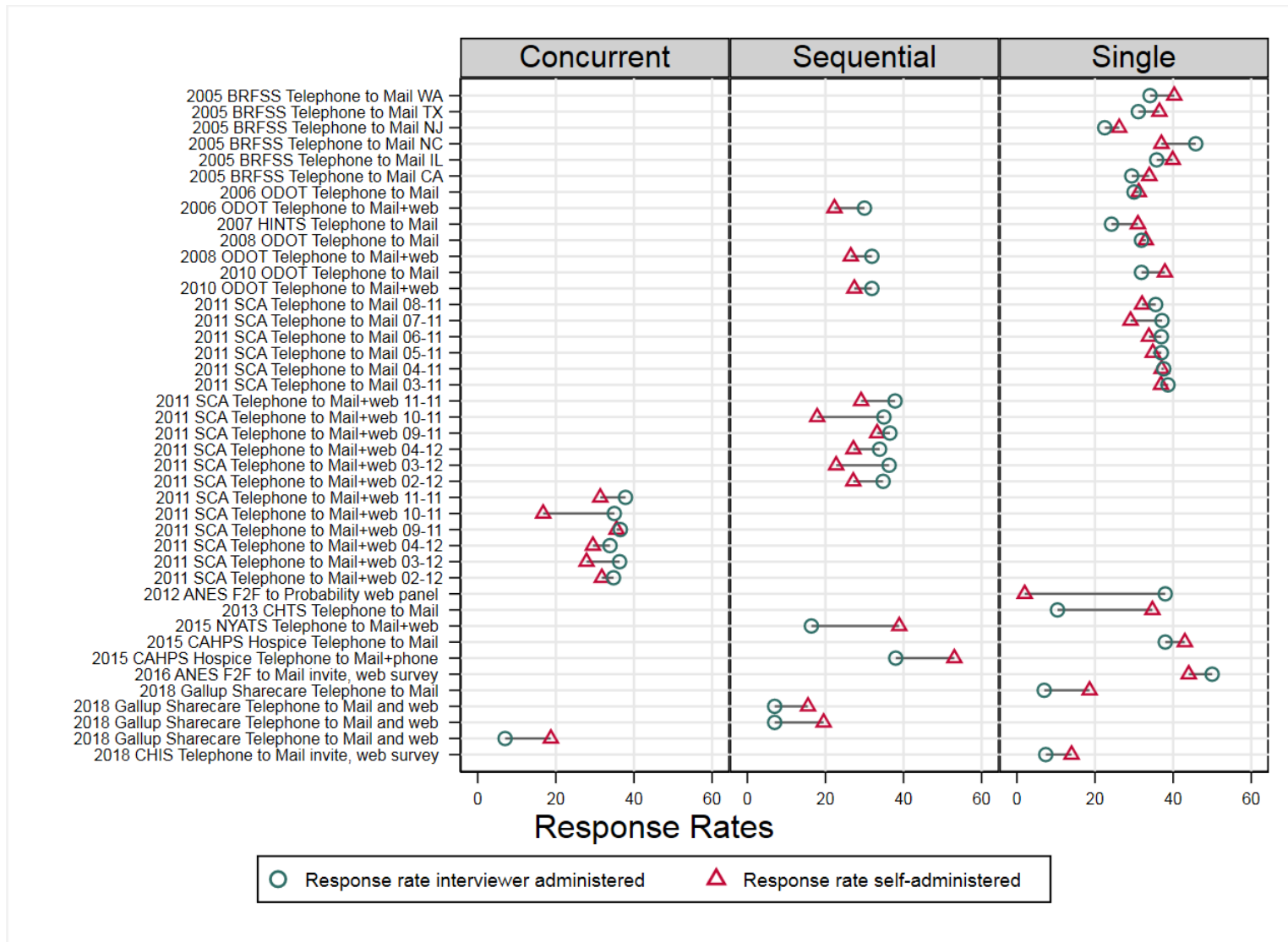


Figure 6.1: Response Rates for Surveys Conducted in Both Interviewer-Administered and Self-Administered or Mixed-Mode Data Collection Modes, Only US Surveys with Interviewer-Administered Mode Conducted within Two Years of Self-Administered Mode

Surveys were excluded from Figure 6.1 when the most recent interviewer-administered version was more than two years prior to the self-administered or mixed-mode implementation. Yet the response rates for the self-administered or administration of these studies are similar to those for the self-administered modes displayed in Figure 6.1. For instance, the response rates (AAPOR RR3) in the pilot study for the 2015 Residential Energy Consumption Survey (RECS) ranged from 35.4% for the web-only condition to 43.9% for the concurrent mail and web condition that offered a bonus incentive for participating online (Biemer, et al. 2018). The weighted response rate for the 2016 National Survey of Children’s Health (NSCH) was 40.7% (Ghandour, et al. 2018). Response rates for cross-sectional international surveys that transitioned to self-administered or mixed modes were similar, ranging from about 18% (e.g., Hoebel, et al. 2014; Bosa, Gagnon, and Caron 2017; Mauz, et al. 2018) to 28% (Klausch, Hox, and Schouten 2015, web) to about 50% (Klausch, Hox, and Schouten 2015, mail).

As can be seen in Figure 6.1, **almost all of the single-stage surveys that transitioned from telephone to self-administered or mixed modes use a single mode or a sequential mode design rather than a concurrent mode design to collect responses.** Concurrent mode designs offer the sampled individual a choice of modes in the initial survey request, for example, information to log into a web survey or return paper questionnaire from the initial recruitment request. Sequential mode designs offer an initial mode (e.g., web) alone followed by a different mode (e.g., mail) for the nonrespondents. This lack of concurrent mixed-mode studies may be because previous reviews and meta-analyses found that mail-only surveys have higher response rates than concurrent web and mail mixed-mode surveys (Shih and Fan 2007; Manfreda, et al. 2008; Medway and Fulton 2012; Dykema, et al. 2013). Notably, these meta-analyses are limited to studies conducted before 2011 (Medway and Fulton 2012), and generally before 2007 (Shih

and Fan 2007; Manfreda, et al. 2008; Dykema, et al. 2013) and thus the findings may not be applicable today.

Response rates for surveys that transitioned to a two-stage screening questionnaire plus main or topical questionnaire are more complicated to compare with response rates from a prior interviewer-administered mode. Some of the interviewer-administered modes do not report response rates for within-household selection or screening step separately from the overall survey response rates, and some of the self-administered modes do not report an overall response rate or a main response rate. Figure 6.2 displays available screener (black), main (blue), and overall (red) response rates for surveys that transitioned to self-administration or mixed modes with two-stages of contact in the interviewer-administered modes (circles) and self-administered modes (triangles). In these studies, overall response rates (displayed as the red circle for the interviewer administered survey and red triangle for the self-administered survey) are generally similar in the four surveys where these rates are available, with the interviewer-administered mode overall response rate slightly higher than the self-administered mode overall response rate. Screener response rates (black) and main survey (blue) response rates, conditional on completing the screener, vary.



Figure 6.2: Response Rates for Surveys Conducted in Both Interviewer-Administered and Self-Administered or Mixed-Mode Data Collection Modes, Only US Surveys with Interviewer-Administered Mode Conducted within Two Years of Self-Administered Mode

We now examine different design features and design decisions related to field operations and nonresponse follow up made by survey organizations when transitioning from telephone to self-administered or mixed modes. The goal of this chapter is not to review all of the literature examining mixed-mode surveys; rather, we focus on particular surveys that have been transitioned from interviewer-administration to self-administration or mixed modes for recruitment and/or data collection. In the next two sections, we will emphasize primarily cross-sectional surveys. The task of recruiting previous wave participants in a longitudinal survey is different from that of recruiting a fresh-cross section and as such, we will discuss longitudinal surveys separately.

6.1 Modes of Contact in Self-Administered and Mixed-Mode Surveys

One major difference that arises as surveys transition from telephone to self-administered and mixed-mode surveys is that the mode for initially contacting the sampled household and the mode of data collection may differ. In telephone-administered surveys, the contact and interview mode are usually the same. Although sampled cases may receive a pre-notification letter in the mail, in order to respond, an interviewer will typically contact the sample member to screen for eligibility and concurrently complete the survey by phone. In self-administered and mixed-mode surveys, a number of different recruitment methods are available and researchers must determine which will be the most effective in eliciting participation among the target population given the contact information researchers have for these sample members. That is, when transitioning from a telephone-administered method to a self-administered method, researchers must carefully consider the contact information available for all sample members and develop a recruitment method accordingly.

The decision of how to combine recruitment and administration modes is critical when transitioning surveys from telephone to self-administered or mixed modes. Contacting respondents in multiple modes has the potential to increase response rates because this diversification increases the chance that sample cases will receive and attend to the request for participation (Dillman, Smyth, and Christian 2014). However, which modes are offered and the order in which they are offered can have important implications for response rates. Modes can be mixed concurrently (that is both options are presented to respondents at the same time) or offered separately in sequence. When offered in sequence, surveys often start with the least expensive (i.e. web) mode followed by the more expensive options of mail or interviewer-

administered modes for nonresponse follow up. This particular strategy of web followed by mail is often referred to as a “web-push” approach (Dillman, Smyth, and Christian 2014).

Table 6.1 displays the new mode of contact and mode of administration of the survey for surveys that transitioned from telephone to self-administered or mixed mode. This table also includes whether the survey instrument was offered in a single mode or as a mixed-mode survey, including whether the modes were administered concurrently or sequentially. **The most common recruitment mode among surveys that have transitioned to self-administered or mixed modes is mail.** This method of recruitment is typically conducted with a mailed letter. Some surveys have transitioned to a fully mail-based recruitment and interview mode. **Other surveys use mailed invitations including a URL and a unique access code for sampled addresses to complete the survey online.** These mail-based invitations to complete the survey online also allow researchers to include other relevant information in the mailing such as brochures, FAQs and even a pre-paid incentive. Follow-up reminders may also include mailed paper screeners and/or topical questionnaires. The mailed web invitation method requires the respondents to type the URL into a website browser, and then to enter a unique username and/or access code to access the survey.

Still other surveys are able to take advantage of email addresses available on the sample frame. An email recruitment message with an embedded URL and access code is used to invite sample members to the survey, so that the respondent only needs to click a link to proceed to the survey. The use of email to recruit sample members is limited to those studies using a special population selected from a list containing email addresses (e.g., students; employees), from probability or non-probability web panels, for studies that used a screener survey to collect an email address, or in longitudinal surveys where an email address has been obtained at a prior

wave. For instance, the Penn State Harrisburg Lion Poll (2019) recently transitioned from telephone to an online nonprobability panel, recruiting survey participants through an emailed link to the web survey. Email can also be used to thank respondents for completing a survey. For example, upon completing the screener, a random subset of the National Household Education Survey (NHES) web respondents were asked to also provide an email address for the knowledgeable respondent for the topical survey to which thank you messages would be sent (McPhee, et al. 2018).

Table 6.1 Examples of Modes of Contact and Modes of Administration for Surveys that Transitioned or are Transitioning to Self-Administered or Mixed Modes

Modes of Contact and Administration	Example Surveys
Contact Mode: Mailed letter	
Administration mode: Mail survey	2005 Behavioral Risk Factor Surveillance System pilot; 2006-2014 ODOT surveys; 2007 Health Information National Trends Survey; 2011 Field Test - Survey of Consumer Attitudes; CAHPS Hospice Survey; Coastal Household Telephone Survey; Dutch Crime Victimization Survey mode experiment; Gallup Sharecare Well-Being Surveys; National Survey of Fishing, Hunting, and Wildlife-Associated Recreation; Racial and Ethnic Approaches to Community Health (REACH) U.S. Risk Factor Survey, Phase 1-3
Administration mode: Web survey	2015 Residential Energy Consumption Survey National Pilot study; 2016 American National Election Studies Time Series Study; 2018 California Health Interview Survey Push-to-web pilot; Canada National Travel Survey pilot; Dutch Crime Victimization Survey mode experiment; National Immunization Survey;
Administration mode: Concurrent mail and web survey	2011 Field Test -Survey of Consumer Attitudes; 2015 Residential Energy Consumption Survey National Pilot study; National Longitudinal Survey of Adolescent to Adult Health Wave V pilot
Administration mode: Sequential mail survey followed by web survey	2015 New York Adult Tobacco Survey
Administration mode: Sequential mail followed by telephone	CAHPS Hospice Survey; National Household Education Survey: 2009 Pilot Study; Racial and Ethnic Approaches to Community Health (REACH) U.S. Risk Factor Survey
Administration mode: Sequential web survey followed by mail survey	2006-2014 ODOT surveys; 2011 Field Test -Survey of Consumer Attitudes; 2015 New York Adult Tobacco Survey; 2015 Residential Energy Consumption Survey National Pilot study; 2016 National Household Education Survey; 2016 National Survey of Children’s Health
Administration mode: Sequential web followed by telephone	2018 California Health Interview Survey Push-to-web pilot
Administration mode: Sequential web followed by mail followed by telephone	German Health Update 2.0 (GEDA) pilot study; 2017 National Survey of College Graduates

Table 6.1 Examples of Modes of Contact and Modes of Administration for Surveys that Transitioned or are Transitioning to Self-Administered or Mixed Modes

Modes of Contact and Administration	Example Surveys
Contact mode: Mailed screener	
Administration mode: Telephone topical Survey	2013-2014 California Health Interview Survey ABS pilot; Wisconsin Family Health Survey;
Administration mode: Mail topical survey	2016 National Household Education Survey; National Household Education Survey: 2011 Field Test; National Survey of Veterans;
Administration mode: Web topical survey	2017 National Household Travel Survey
Contact mode: Mailed letter, Multiple modes screener and topical survey	
Administration mode: Sequential: Mail topical survey followed by phone	National Household Education Survey: 2009 Pilot Study
Administration mode: Sequential: Web screener and/or topical survey followed by mail	2016 National Household Education Survey; National Longitudinal Survey of Adolescent to Adult Health Wave V
Contact mode: Email	
Administration mode: Web survey	2015 Canada Election Study; American National Election Studies 2012 Time Series Study; American Trends Panel; Penn State Harrisburg Lion Poll; Rutgers-Eagleton Poll 2019;
Administration mode: Sequential web survey followed by telephone and face-to-face	University of Michigan 2015 Campus Climate Survey

Table 6.1 Examples of Modes of Contact and Modes of Administration for Surveys that Transitioned or are Transitioning to Self-Administered or Mixed Modes

Modes of Contact and Administration	Example Surveys
Contact mode: Telephone	
Administration mode: Concurrent phone and web	2005 Health Information National Trends Survey (HINTS)
Administration mode: Sequential telephone followed by mail	Racial and Ethnic Approaches to Community Health (REACH) U.S. Risk Factor Survey

Some studies have used text messaging for recruitment or reminders. Because U.S. law and EU regulations require all text message senders to have explicit permission to send the sample member a text prior to doing so, text messaging is difficult to use if such consent has not already been obtained. **Text message invitations and reminders are thus rarely practical for one-time surveys, but can be particularly useful for panel and longitudinal surveys where consent can be obtained.** McGeeney and Yan (2016) examined the effectiveness of using SMS messages in addition to email messages using the Pew American Trends Panel, finding that these two modes of recruitment boosted initial response to a web survey in 2015, although the overall response rate for those who responded to the survey did not differ by the end of the field period. The 2017 National Household Travel Survey (NHTS) asked households that completed a screening questionnaire to provide an email address or phone number. These email addresses and phone numbers were used to either email, text, or send automated phone messages through Interactive Voice Response (IVR) to the sample units to remind them about completing the travel log (Westat 2018).

6.2 Modes of Response in Self-Administered and Mixed-Mode Survey Transitions

Table 6.1 also contains the mode of administration for self-administered and mixed-mode surveys. **Early mixed-mode studies that transitioned from telephone often still included telephone as one of the contact modes and modes of data collection.** For instance, the 2005 Health Information National Trends Survey (HINTS) contacted households by telephone and offered them the option of completing the survey online - this approach only yielded 95 web

respondents and reduced the overall extended interview response rate from 65.4% for the telephone only group to 57.0% for the group provided a mode choice (Cantor, et al. 2005).

Many initial mixed-mode surveys used mail as a method of gaining telephone numbers for cellphone only households and/or households where the address could not be linked to a telephone number through a reverse directory telephone match. For instance, Allison, Stevenson, and Kniss (2014) sent a one-page mail questionnaire requesting a telephone number to an address-based sample of Wisconsin households that could not be reverse directory list-matched to a telephone number for the 2012 Wisconsin Family Health Survey, with 43% of unmatched households returning the questionnaire, 91.6% of which had a valid telephone number to be called for a telephone interview. This approach was also used in two pilot studies for the California Health Interview Survey (CHIS) (Jans, et al. 2013; Kali and Flores Cervantes 2016). Here, the CHIS matched an address-based sample with listed telephone numbers in selected counties who were called via the regular CHIS telephone approach. For portions of the sample that could not be matched to a telephone number, households were mailed a screening survey to request a telephone number (Kali and Flores Cervantes 2016). Only 15% of households returned the form with a telephone number, yielding a 9% completion rate among the unmatched ABS sample. However, this is higher than the main CHIS landline and cell phone sample completion rates (4.1% landline; 5.2% cell phone) (see also Jans, et al. 2013). Pilot work for the National Crime Victimization Survey tested two approaches in one metropolitan area – sending screeners to obtain telephone numbers only for unmatched households (about 40% of the sample) versus to all households (Brick, et al. 2013). In both approaches, about one-third of unmatched households completed the mail screener (47% of matched addresses returned the screener), and just under 75% of returned screeners contained a telephone number.

Other mixed-mode studies use telephone as a nonresponse follow-up data collection mode in addition to mail for addresses that are linked to telephone numbers through reverse directory lookup or other information on the sample frame. In the 2009 National Household Education Surveys pilot, matched addresses with telephone numbers (about 57% of the sample) were randomly selected for nonresponse follow-up with telephone or an additional mail attempt. Following up the mail survey with telephone calls yielded lower screener response rates (34.4%) than staying with mail alone (49.3%; Brick, Williams and Montaquila 2011). Additionally, the Racial and Ethnic Approaches to Community Health across the US Risk Factor Survey (REACH US) randomly assigned addresses matched to a telephone number to be initially contacted in a telephone mode and then nonrespondents followed up with a mailed paper questionnaire (the phone-first approach), or to be initially contacted with a mail questionnaire and then nonrespondents followed up with telephone (the mail-first approach) (Amaya, et al. 2015; see also Murphy, Harter, and Xia 2010; LeClere, et al. 2012). Following up mail nonrespondents with telephone had a higher screener response rate (48.7%) and higher interview completion rate (79.8%) than the following up telephone nonrespondents with mail (screener: 44.8%; interview: 70.8%). Finally, in a survey of students about sexual misconduct, Axinn, Wagner, Couper, and Crawford (2018) used an invitation email request to student email addresses, obtaining a 54.0% response rate. Interviewers then followed up with nonrespondents either on the telephone to encourage an online response or in-person, bringing tablet computers so that sampled members could complete the survey online at the time of face-to-face contact, increasing the overall response rate by 13 percentage points.

However, many surveys that transitioned from telephone to self-administered surveys abandoned telephone and used only mail for contact and data collection mode. For

example, the National Household Education Survey (NHES) was an early adopter of a mail survey as an approach for transitioning from RDD to self-administered surveys (e.g., Montaquila, et al. 2013). NHES achieved response rates for a mail-only design that equaled or exceeded the most recent telephone administration of the surveys. Brick Andrews, and Mathiowetz (2016) report on a one-stage recruitment of a rare population (participants in recreational saltwater fishing in specific geographic areas) in which a single-stage mail survey yielded response rates across four states that averaged 34.7%, over three times higher than the telephone survey mounted at the same time in the same states (10.4%). A subsample of nonrespondents to the initial survey were re-sent the mail survey with an increased incentive of \$5, yielding a combined weighted response rate of 64% across the two phases of data collection.

More recent mixed-mode studies use mail to recruit sampled individuals, but use only web as a data collection mode. For example, in Canada, the National Travel Survey transitioned to a mail recruitment of adults in households who are asked to go online and report information on domestic and international trips that they have taken for either personal or business-related reasons in a web survey (Bosa, Gagnon, and Caron 2017). In the US, the 2017 NHTS asks sampled households to complete a screener either online or via a mailed paper questionnaire. Households that complete the screener are mailed paper logs for recording their travel activity on a sampled day, and then asked to enter it online into a web instrument, or to complete it over the telephone (Federal Highway Administration and Westat 2018). The web component of the 2016 American National Election Studies (2018) mailed US households an invitation to complete a screening questionnaire online and have a randomly selected adult US citizen living in the household then complete the online questionnaire.

Other studies use a mailed contact letter to recruit respondents to complete either a mail or web questionnaire. In these designs, typically, a mailed recruitment letter to a web survey is followed in later contact attempts with a mailed paper questionnaire (Dillman, Smyth, and Christian 2014). For instance, Marlar, et al. (2017) use a mail screener survey to identify those with fishing activity (and a small subset of those without fishing activity), who are then sent a letter asking them to go online to complete the topical survey. Nonrespondents were followed up with a mail survey. The NSCH sent a mailed invitation to sampled addresses containing a URL, username, and password, plus a \$2 or \$5 cash incentive for logging on the web survey (US Census Bureau 2018; Ghandour, et al. 2018). Nonrespondents received repeated mailings about logging into the website to complete the screening instrument and the main questionnaire; remaining nonrespondents were sent a paper screening questionnaire and, if eligible, were mailed a paper topical questionnaire. The 2016 NHES added a web component to the data collection (McPhee, et al. 2018). Households were mailed a cover letter containing a URL and login information for the web survey; after completing the web screener, screener respondents who were the selected topical respondent immediately continued into the appropriate web survey. Households with an eligible person in the household who was not the screener respondent were mailed a topical web package containing a letter identifying the appropriate respondent. Nonrespondents to the web survey screener were sent a paper screening questionnaire; nonrespondents to the topical survey were followed up by a paper questionnaire for remaining nonrespondents.

Although early meta-analyses found that concurrent web-mail surveys had lower response rates than mail-only surveys, more recent experiments have found few consistent differences in response rates across single, concurrent, and sequential designs. For example,

several recent studies have found no notable difference in response rates between single mode web-only or mail-only surveys and concurrent web and mail surveys (Mathews, et al. 2012; Steele, et al. 2016; Marken, Auter, and Marlar 2018; Biemer, et al. 2018). Other recent studies that have compared sequential modes (web+mail or mail+web) with single mode studies have found either no difference in response rates between these two designs (Weaver, Beebe, and Rockwood 2019) or higher response rates for the sequential modes than the single modes (McMaster, et al. 2017; Biemer, et al. 2018; Millar, et al. 2018). Finally, other studies have found similar or higher response rates for concurrent web+mail design than a sequential mixed-mode design (web+mail) (Lesser, et al. 2016; Bucks, Fulford, and Couper 2018).

Some of the difference in response rates for web surveys in recent years may be due to shifts in the proportion of adults who have access to the internet. In 2000, only about half of the US adult population had access to the internet (Pew Research Center 2019). This grew - about 90% of the US adult population in 2019 has access to the internet, and access is almost universal for adults under age 50. We speculate that this shift in both internet coverage and internet familiarity may change the “best practices” recommendations for how to combine and sequence modes. Clearly, more research and a systematic meta-analysis of recent studies comparing response rates across these combinations of modes of data collection is needed.

6.3 Longitudinal Surveys and Transitions to Self-Administered Modes

Longitudinal surveys of adults in the United States and throughout the world are often conducted with an initial face-to-face recruitment for the first wave of data collection, and transition to alternative, less expensive modes, for follow-up waves of data collection (de Leeuw 2005; Schoeni, Stafford, McGonagle, and Andreski 2013). Traditionally, the alternative less

expensive mode of data collection has been telephone. For example, the Current Population Survey in the United States starts with a face-to-face recruitment of sampled addresses, follows up for the next three months with about 85% of the interviews conducted on telephone, returns for a face-to-face interview for the fifth wave, and then returns to telephone interviews (Bureau of Labor Statistics, 2018). Here, we will focus exclusively on longitudinal surveys that select a probability sample of individuals with the goal of following these individuals over time on a common set of measures rather than online panel surveys that allow clients to purchase administration of items with varying survey topics or content.

It is becoming increasingly common for longitudinal household surveys to transition from an interviewer-administered mode to a self-administered mode or a mixed-mode data collection strategy for at least some of the follow-up waves (Dillman 2009). The National Longitudinal Study of Adolescent to Adult Health (commonly known as Add Health) used face-to-face interviews for the first four waves of data collection, moving to web and paper survey instruments for the fifth wave of data collection in 2016-2018, with face-to-face and telephone follow-up and in-person administration of biomarker collection (Harris 2018). The Health and Retirement Survey (HRS) uses mail and web surveys during years in between the face-to-face data collection efforts to collect additional information from respondents on a wide variety of topics (Health and Retirement Survey 2019). The Panel Study of Income Dynamics (PSID) also uses a mixed-mode questionnaire during non-interview years; in 2014, a web data collection was used to collect information about experiences that the PSID sample member had as a child (McGonagle, Freedman, and Griffin 2017) and in 2016 a web followed by mail survey was used to collect information on topics including well-being, personality traits, and literacy and numeracy skills (Freedman 2017). Understanding Society, the UK Household Longitudinal

Study, used a face-to-face recruitment during the first wave of the survey, with primarily face-to-face interviews through wave six. In Wave 7, nonresponding households to at least two prior waves were provided the opportunity to complete a web survey, with a face-to-face follow up. In Wave 8, the web-first group was expanded to 40% of the sample during the survey production year (Bianchi, Biggignandi, and Lynn 2017; Carpenter 2018). The Canadian Labour Force Survey (LFS) uses a face-to-face or telephone recruitment for the first wave of data collection. Starting in 2015, the Canadian LFS started offering a web survey for data collection in the second through sixth months of data collection (Francis and Laflamme 2015).

Other longitudinal surveys start with self-administered modes and use more expensive interviewer-administered modes for nonresponse follow-up or in an attempt to tailor to a respondent's reported preferences. For instance, the High School Longitudinal Survey of 2009 collected self-administered electronic questionnaires from ninth grade students using an in-school administration, sequential web to telephone questionnaires from parents of those students (with mailed shortened questionnaires for nonresponse follow-up), and concurrent telephone or web questionnaires for teachers, school administrators, and school counselors (Ingels, et al. 2011). The 2017 National Survey of College Graduates (NSCG), collected by the US Census Bureau, is a longitudinal survey of adults holding at least a bachelor's degree, sampled from the American Community Survey (SESTAT 2018; OMB 2017). Newly sampled persons are initially contacted via mail to complete a web survey, and nonrespondents are followed up first with a mail questionnaire and then a telephone interview. All longitudinal cases are provided information about completing the interview via web; some longitudinal cases are also provided information about completing the questionnaire via mail or telephone, with nonrespondents followed up with both web login and mailed information. Similarly, Monitoring

the Future (MTF) is examining the use of mail recruitment to a web survey and email recruitment to a web survey for the longitudinal follow-up of survey respondents who initially complete a survey in the classroom (Patrick, et al. 2018).

6.4 Adaptive/Responsive Designs

Adaptive and responsive designs (Groves and Heeringa 2006) can be used to attempt to reduce nonresponse bias or survey costs by deliberately attempting to tailor data collection methods to the “optimal” method for individual sampled cases or groups of cases. The goal of an adaptive or responsive design is to maximize response rates, reduce costs, and make it more likely that the sample will adequately represent the target population. In practice, a common approach to adaptive design is to design an upfront differential strategy that will target specific data collection strategies to subsamples to gain cooperation. Self-administered and mixed modes of data collection are among the strategies that are considered in adaptive and responsive designs.

Surveys that are transitioning to mixed-mode or self-administered surveys from telephone surveys can plan to use adaptive mixed-mode strategies before data collection begins. **Planning for active monitoring of multiple metrics during data collection is important when transitioning to self-administered or mixed modes. This planning is especially important when using experiments to identify the “best” mode or combination and sequencing of modes going forward and there is no prior data to use to conduct initial data analyses to understand the impact of these decisions.** For instance, one might tailor mixed-mode strategies to certain subgroups based off of information on the sampling frame, even without existing data about the potential benefit of these decisions. The 2015 National Census Test used a mixed-mode design with various mail strategies. The majority of the strategies started with mailed

letters containing a URL and login information for accessing the web questionnaire, followed by mailed questionnaires to nonrespondents (a web-push design). However, in areas with low internet penetration (Phelan 2016), identified from geographic information on the Census Master Address File, sampled addresses were offered a choice between mail and web from the initial mailing.

In an example of a survey that transitioned from face-to-face to a self-administered mixed-mode format, Murphy, Biemer, and Berry (2018) used adaptive and responsive design approaches to monitor data collection during a mixed-mode experiment for the RECS pilot. A number of nonresponse-related metrics were identified as being important to monitor for each mode condition, including completion rates, the percent of completed questionnaires submitted via the web, metrics of relative cost, and important key estimates, including housing unit type and current heating fuel estimate versus the national benchmark, among others. Because the survey had preidentified these metrics and actively monitored them during the field period, these metrics could be used to compare the yield from the self-administered modes to the face-to-face administered main RECS survey. When the face-to-face RECS fell behind the self-administered modes, nonresponse follow-up for the face-to-face main RECS study was conducted using the self-administered mixed modes from the pilot study. A similar approach was used for monitoring the adaptive design in the mixed-mode National Longitudinal Survey of Adolescent to Adult Health (Add Health) Wave V data collection (Murphy, et al. 2019).

Surveys transitioning from telephone to self-administered or mixed modes can also use adaptive measures during data collection to attempt to “optimize” the use of different modes. **In particular, limited data collection resources can be judiciously assigned to where they are most likely needed via more expensive modes (i.e. interviewer-administered) or other data**

collection methods (e.g., incentives). For repeated cross-sectional surveys or longitudinal surveys that have previously had a mix of modes in data collection but want to increase the proportion of self-administered modes, statistical analysis and simulation using the existing data can be used to plan interventions in a new round of data collection. For example, Coffey and colleagues (2013, 2015; Finamore, et al. 2015) used interventions in the National Survey of College Graduates (NSCG), a sequential mixed-mode survey, to improve the representativeness of the sample. The potential interventions were preidentified through simulating the impact of different decisions on different metrics (e.g., R-indicators, costs) on a previous round of data collection (Coffey, et al. 2013). In the 2013 NSCG, the adaptive design deliberately changed the mix of modes during the field period to improve representation of the sample and control costs. In particular, they increased telephone follow-up to cases who were under-represented during data collection, including Black and Hispanic sampled cases who have a Bachelor's degree, while reducing or eliminating telephone follow-up and increasing web follow-up for cases that were over-represented (Whites with a Bachelor's Degree). This strategy resulted in a more representative sample, without increasing costs or negatively affecting response rates.

The Dutch Labor Force Survey also used an adaptive design to sequence the use of web, telephone, and face-to-face modes of data collection, as well as the use of additional call attempts in the face-to-face mode (Schouten, Peytchev, and Wagner 2018). Through use of a rich frame (the Dutch Population Register and the Dutch Tax Board Register), modes and the sequence of modes could be optimized across subgroups of the population to make the most efficient use of these modes while controlling costs and reducing potential nonresponse bias.

6.5 Designing Contact Attempts

When switching from an interviewer-administered to a mixed-mode survey, the entire recruitment protocol must change. Rather than call attempts administered through a telephone call scheduler, in self-administered modes and mixed-mode surveys that contain self-administered components, recruitment comes via mailings sent to a household or, for select studies, emails sent to the sampled individual. If there are interviewer-administered components, these may be attempted later in the data collection field period to reduce costs.

Table 6.2 contains a summary of the number and type of contact attempts used across a non-exhaustive list of surveys that have transitioned from telephone to self-administered or mixed modes without a screener, excluding surveys that used exclusively telephone to recruit sampled individuals to a survey and then offered other modes of data collection during the recruitment. **There is no single contact protocol used for self-administered or mixed-mode surveys.** The number of contact attempts (sent via mail primarily; telephone, email and text message rarely) ranged from 1 to 14. About half of the studies in table 6.2 included an advance letter or advance postcard; the rest contained a questionnaire or login information directly from the first mailing. Follow-up materials tended to include at least one reminder postcard in almost all of the studies, at least one replacement paper questionnaire in studies that used a mail questionnaire, and at least one letter after the initial containing the survey URL and login information in studies that used a web questionnaire. Those studies that switched modes in a sequential mixed-mode web/mail design tended to do so at the second or the third contact attempt to the household. Those studies that switched at the third contact attempt did so after sending a reminder postcard to the household to complete the survey in the initial mode.

Table 6.2. Number and Type of Contact Attempts in Example Surveys That Have Transitioned from Telephone to Self-Administered or Mixed Mode

Survey Name	# contact attempts	Type of contact attempts	Source
CAHPS Hospice Survey	2	Mail: (1) Invitation Letter and questionnaire; (2) Paper Questionnaire	Parast, et al. (2018)
2005 Behavioral Risk Factor Surveillance System pilot	3	Mail/Telephone: (1) Invitation Letter and questionnaire; (2) Telephone follow-up (5 attempts) (1) Invitation Letter and questionnaire; (2) Postcard reminder; (3) Replacement questionnaire	Battaglia, et al. (2008); Link, et al. (2008)
Coastal Household Telephone Survey	3	(1) Invitation letter and questionnaire; (2) Postcard reminder / Telephone call reminders; (3) Replacement questionnaire	Brick, Andrews, and Mathiowetz (2016)
German Health Update 2.0 (GEDA) pilot study	3	Concurrent web/mail/CATI: (1) Invitation letter and questionnaire and URL and log-in information plus CATI survey return form; (2) Letter with URL and login information; (3) Reminder Letter with URL and login information Sequential web/mail/CATI: (1) Invitation letter with URL and login information; (2) Paper questionnaire and URL information; (3) Reminder letter with URL and login code and CATI survey form	Mauz, et al. (2018); Hoebel, et al. (2014)
Gallup Sharecare Wellbeing Index	3	Mail only: (1) Invitation Letter and questionnaire; (2) Postcard reminder; (3) Postcard reminder Concurrent web/mail: (1) Invitation letter and questionnaire and URL and log-in information; (2) Letter with URL and login information; (3) Postcard reminder Sequential web/mail: (1) Invitation letter with URL and login information; (2) Paper questionnaire and URL; (3) Postcard reminder Sequential mail/web: (1) Invitation Letter and questionnaire; (2) Letter with URL and login information; (3) Postcard reminder	Marken, Auter, and Marlar (2018)
Canada National Travel Survey pilot	3	(1) Invitation letter with URL and login information; (2) Letter with URL and login information; (3) Reminder letter with URL and login information	Bosa, Gagnon, Caron (2017)
2015 New York Adult Tobacco Survey	4	Web/mail sequential: (1) Advance letter; (2) Letter with URL and login information; (3) Reminder postcard with URL; (4) Paper questionnaire and URL Mail/web sequential : (1) Advance letter; (2) Paper questionnaire; (3) Reminder postcard; (4) Paper questionnaire and URL	Brown, et al. (2018)

Table 6.2. Number and Type of Contact Attempts in Example Surveys That Have Transitioned from Telephone to Self-Administered or Mixed Mode

Survey Name	# contact attempts	Type of contact attempts	Source
Dutch Crime Victimization Survey	4	Mail to F2F: (1) Invitation Letter and questionnaire; (2) Paper questionnaire; (3) Replacement questionnaire; (4) Face-to-face attempt	Klausch, Hox, and Schouten (2015)
American Crime Victimization Survey Field Test	4	Web to F2F: (1) Invitation letter with URL and login information; (2) Letter with URL and login information; (3) Reminder Letter with URL and login information; (4) Face-to-face attempt	Williams, Edwards, Giambo, and Kena (2018)
2018 California Health Interview Survey pilot	4	(1) Invitation letter with URL and login information; (2) Postcard reminder with URL and login information; (3) Reminder letter with URL and login information; (4) Telephone follow-up	Wells, et al. (2018)
2006-2014 ODOT surveys	5	Mail: (1) Advance letter; (2) Paper questionnaire; (3) Postcard reminder; (4) Replacement questionnaire; (5) Replacement Questionnaire	Lesser, et al. (2016)
2007 Health Information National Trends Survey	5	Web/Mail sequential: (1) Advance letter; (2) Letter with URL and login information; (3) Reminder postcard with URL; (4) Replacement questionnaire; (5) URL Letter and Questionnaire	Cantor, et al. (2009)
Survey of Consumer Attitudes	5	Mail: (1) Advance letter; (2) Paper questionnaire; (3) Postcard reminder; (4) Replacement questionnaire; (5) Postcard reminder	Elkasabi, et al. (2014); Survey of Consumers (2012)
National Immunization Survey	5	Mail/Web concurrent: (1) Advance letter; (2) Paper questionnaire and URL information; (3) Postcard reminder; (4) Paper questionnaire and URL; (5) Postcard reminder	
2015 Residential Energy Consumption Survey National Pilot study	7	Web/Mail sequential: (1) Advance letter; (2) Letter with URL and login information; (3) Reminder postcard with URL; (4) Reminder Postcard with URL; (5) URL Letter and Questionnaire	
		(1) Advance postcard; (1) Letter with URL and login information; (3) Reminder postcard with URL; (4) Reminder letter with URL and login information; (5) Reminder postcard with URL and login information	Skalland , et al. (2017)
		(1) Advance postcard; (2) Letter with URL and login information; (3) Reminder postcard; (4) Paper questionnaire and URL; (5) Postcard reminder; (6) Reminder letter; (7) Short questionnaire	Biemer, et al. (2018)

Table 6.2. Number and Type of Contact Attempts in Example Surveys That Have Transitioned from Telephone to Self-Administered or Mixed Mode

Survey Name	# contact attempts	Type of contact attempts	Source
National Survey of College Graduates	14	(1) Advance letter; (2) Letter with URL and login information; (3) Reminder postcard; (4) Reminder letter with URL and login information; (5) Reminder email; (6) Reminder letter with URL and paper questionnaire; (7) Reminder postcard; (8) Telephone reminder; (9) Reminder letter with telephone information; (10) Telephone calls; (11) Reminder letter with URL; (12) Reminder letter with URL and paper questionnaire	Coffey (2016); National Academies of Sciences, Engineering, and Medicine (2018)

In general, many of the mail-based protocols reflect the recommendations made by Dillman, Smyth, and Christian (2014, p. 373). These protocols either consist of 5 contact attempts, including an advance mailing, questionnaire, reminder (letter or postcard), replacement questionnaire, and final reminder, or 4 contact attempts of a full questionnaire packet, reminder (postcard or letter), replacement questionnaire, and final reminder. Surveys that start with a paper questionnaire generally use at least two or three mailings with a complete paper questionnaire. Surveys that include a web questionnaire generally include the URL and login information at each mailing after the advance letter (Bosa, Gagnon, and Caron 2017; American National Election Studies 2018; Mauz, et al. 2018). Thus, the surveys that include web as one of the modes generally include the full information for participating in the web survey (URL, login information) at more mailings than the mail surveys that include the paper questionnaire at only a subset of the mailings. This makes sense from a cost perspective (paper questionnaires are more expensive to print and mail) and potentially from an error perspective (web questionnaires require more effort for the sampled individual to login to and complete than the mail survey).

In mixed-mode surveys that combine both web and mail sequentially, nonresponding households receive the paper questionnaire in the second (Marken, Auter, and Marlar 2018; Mauz, et al. 2018), third (Elkasabi, et al. 2014; Han, et al. 2010; Biemer, et al. 2017; Ghandour,

et al. 2018), or fourth mailing (Lesser, et al. 2016; Brown, et al. 2018) to the household. In a second type of design, a mailed recruitment letter to a web survey is sent to the household with nonrespondents followed up with an interviewer-administered mode (Klausch, Hox and Schouten 2015; Federal Highway Administration and Westat 2018; Wells, et al. 2018).

Table 6.3 contains a list of surveys that contain a screener, which use up to 9 contact attempts when combining the total number mailings for both the screener and topical surveys.

Surveys that include a separate screening questionnaire from the topical or main questionnaire have very similar mailing protocols. Almost all use either 4 or 5 mailings at each stage, with almost all of these studies requesting a response (e.g., including a screener questionnaire, URL and login information, form for telephone number) in the initial mailing rather than starting with an advance letter.

Table 6.3. Number and Type of Contact Attempts in Surveys That Have Transitioned from Telephone to Self-Administered or Mixed Modes

Survey Name	# of contact attempts	Type of contact attempts	Source
Wisconsin Family Health Survey	Screener: 3 Topical: Not specified	Screener: (1) Invitation letter and form requesting telephone number; (2) Reminder postcard; (3) Cover letter and form requesting telephone number Topical: (1) Telephone to addresses that returned form with telephone number	Allison, Stevenson, and Kniss (2014)
National Household Education Survey: 2009 Pilot Study	Screener: 4 Topical: 5	Screener: (1) Screener questionnaire and letter; (2) Reminder postcard; (3) Reminder screener questionnaire and letter or telephone reminder; (4) Reminder screener questionnaire and letter or telephone reminder Topical: (1) Topical questionnaire and letter; (2) Reminder postcard; (3) Reminder topical questionnaire and letter; (4) Reminder topical questionnaire and letter; (5) Telephone follow-up	Brick Williams, Montaquila (2011)
National Survey of Veterans	Screener: 4 Topical: 4	Screener: (1) Advance letter; (2) Screener survey and letter; (3) Postcard reminder; (4) Reminder survey Topical: Web: (1) Letter with URL and login information; (2) Reminder postcard; (3) Paper questionnaire and letter	Han, et al. (2010)

		with URL; (4) Paper questionnaire and letter with URL and telephone call in information	
National Household Education Survey: 2011 Field Test	Screener: 4 Topical: 4	Mail: (1) Paper questionnaire and letter; (2) Reminder postcard; (3) Paper questionnaire; (4) Paper questionnaire and telephone call in information Screener: (1) Screener questionnaire and letter; (2) Reminder postcard; (3) Reminder screener questionnaire and letter; (4) Reminder screener questionnaire and letter	Montaquila, et al. (2013)
2013 California Health Interview Survey ABS pilot	Screener: 4 Topical: Not specified	Topical: (1) Topical questionnaire and letter; (2) Reminder postcard; (3) Reminder topical questionnaire and letter; (4) Reminder topical questionnaire and letter Screener: (1) Invitation letter and form requesting telephone number; (2) Reminder postcard; (3) Cover letter and form requesting telephone number; (4) Cover letter and form requesting telephone number	Jans, et al. (2013); California Health Interview Survey (2016)
2017 National Household Travel Survey	Screener: 4 Topical: 5 Telephone: 7 days	Topical: (1) Telephone to addresses who returned telephone number request and those who matched directory listings Screener: (1) Invitation letter and paper questionnaire; (2) Reminder postcard; (3) Letter and paper questionnaire; (4) Reminder postcard with URL and PIN for online completion	Federal Highway Administration and Westat (2018)
2016 National Survey of Children's Health	Screener: 5 Topical: 4	Topical – phone: Interviewers attempted calls for 7 days after assigned travel day Screener: (1) Invitation letter with URL and login information; (2) Reminder letter with URL and login information; (3) Reminder letter with URL and login information; low web received paper screener; (4) Reminder letter with paper screener; (5) Reminder letter with paper screener Topical for web nonrespondents: (1) Paper Questionnaire; (2) Follow-up paper questionnaire; (3) Follow-up paper questionnaire; (4) Follow-up paper questionnaire	Ghandour, et al. (2018)

6.6 Incentives

A ubiquitous finding across surveys and across modes of data collection is that prepaid incentives raise response rates compared to no incentives, and that prepaid incentives are more

successful at encouraging response than promised incentives (Singer and Ye, 2013; Mercer, et al. 2015). As such, surveys that transition from telephone to self-administered or mixed modes often use incentives as part of the recruitment protocol. Most of the studies reported on in our survey of organizations that have transitioned use incentives (20 of 24 answering). About half of respondents reported that the level of incentives used in each mode did not change, other than to account for inflation as time passed. In general, the changes reported were modest in size, though one respondent reported that the shift to internet data collection brought enough cost savings to offer gift cards when no budget had previously been available for incentives. One organization shifting from RDD to web reported a shift in incentives from \$5 for cellphone respondents to variable incentives for all respondents ranging from \$5 to \$20. Another is offering a bonus incentive for respondents who voluntarily shift from paper to web. Both pre-paid and promised incentives were reported.

In mailed invitations to a mail or web survey, pre-paid incentives are highly effective in increasing participation. Table 6.4 contains an overview of incentives that have been offered in surveys that have transitioned to self-administered or mixed modes of data collection. Many of these studies included experimental comparisons of incentive levels versus a \$0 condition (excluded from this table); surveys that transitioned but did not mention an incentive level are excluded from this table.

Table 6.4: Summary of monetary incentive levels and example studies using the incentive amount

Incentive Amount	Example Studies
Prepaid	
Amount not reported	Brick, Andrews, and Mathiowetz (2016); Breton, et al. (2017)
\$1	Skalland, et al. (2017); Andrews, Brick, and Mathiowetz (2013); Williams, Edwards, Giambo, and Kena (2018)
\$2	Brick Williams, Montaquila (2011); Cantor, et al. (2009); Montaquila, et al. (2013); Allison, Stevenson, and Kniss (2014); Jans, et al. (2013); Ghandour, et al. (2018); Federal Highway Administration and Westat (2018); Wells, et al. (2018); Jackson, McPhee, and Lavrakas (2019); Williams, Edwards, Giambo, and Kena (2018)
\$5	Montaquila, et al. (2013); Elkasabi, et al. (2014); Murphy, Harter, and Xia (2010); LeClere, et al. (2012); Ghandour, et al. (2018); Federal Highway Administration and Westat (2018); Amaya, et al. (2015); Brown, et al. (2018)
\$10	Jackson, McPhee, and Lavrakas (2019)
\$20	American National Election Studies (2018)
\$30	National Academies of Sciences, Engineering and Medicine (2018)
Promised	
\$5	Cantor, et al. (2005); Brick Williams, Montaquila (2011); Montaquila, et al. (2013)
\$10	Biemer, et al. (2017); Montaquila, et al. (2013)
\$15	Cantor, et al. (2005); Brick Williams, Montaquila (2011); Montaquila, et al. (2013)
\$20	Allison, Stevenson, and Kniss (2014); Biemer, et al. (2017); Montaquila, et al. (2013); Federal Highway Administration and Westat (2018)
Promised >\$20	American National Election Studies (2015); American National Election Studies (2018); Harris (2019)

Looking across surveys, prepaid incentives of \$2 and \$5 are common. Promised incentives are less commonly used, but when used, tend to be larger in value than prepaid incentives. In mixed-mode surveys, a combination of prepaid and promised incentives can be effective in pushing respondents to a new mode. For instance, the proportion of respondents who complete via a web instrument in a web+mail survey can be increased when a small prepaid incentive is followed by a larger promised incentive paid to those who respond by web (Biemer et al. 2018).

One concern often voiced for federal surveys that are transitioning from telephone to self-administered modes on the use of incentives is potential restrictions from the Office of

Management and Budget (OMB) on the use of incentives. OMB has allowed the use of incentives in Federal data collections, albeit on a limited basis. Guideline 2.3.1 of the Standards and Guidelines for Statistical Surveys (OMB 2006) notes, “Although incentives are not typically used in Federal surveys, agencies may consider use of respondent incentives if they believe incentives would be necessary to use for a particular survey in order to achieve data of sufficient quality for their intended uses(s).” Typically, requests for incentive use must be approved based on specific justification for their use as part of the overall survey docket, and ideally with explicit plans for evaluation of their effectiveness. As such, incentive experiments are often a part of federal surveys that have transitioned from interviewer-administered to self-administered modes, with amounts as represented in the Table 6.4 above and a control condition of \$0 for comparison.

In addition to the question of whether or not to use pre-paid or post-paid monetary incentives, some studies that have transitioned to a self-administered or mixed mode have examined (1) how to distribute monetary incentives, and (2) whether to use a non-monetary incentive. Monetary incentives can be distributed via debit cards, plastic or electronic gift cards, cash, or checks. Debit cards and checks incur costs only when they are cashed, which can result in significant cost savings to the survey organization. For example, in the 2015 mixed-mode NSCG, only 35% of recipients used the debit card (Vasquez 2019). Similarly, in the mail component of Phase III of the Agricultural Resource Management Survey, prepaid \$20 ATM cards were cashed by 39% of recipients (Beckler, Ott, and Horvath 2005). With self-administered and mixed-mode surveys, some types of incentives may be more appropriate for specific modes. For example, electronically delivered incentives may make sense for web surveys where sample members are only contacted via email, whereas incentives such as cash, debit cards, or checks may be more appropriate for those who are contacted via mailed letters.

Non-monetary incentives can also be delivered in a mixed-mode survey, although with more limited effectiveness. For example, the web and mail 2018 National Sample Survey of Registered Nurses used lanyards and pens, incentives that were thought to be salient to the target population (US Census Bureau 2017, FR-2017-13292). The 2014 NHES-Feasibility Test included a Department of Education magnet in the screener questionnaire. There was no statistical benefit to including this non-monetary incentive on response rates or eligibility rates (McQuiggan, et al. 2015).

Survey practitioners also need to decide which sample cases will receive an incentive. One option is to target incentives or to provide differential incentives to groups that are least likely to respond or are demographically different. For example, Jackson, McPhee, and Lavrakas (2019) used a tailored incentive in the sequential mixed-mode web and mail 2016 NHES, targeting higher incentive levels to those estimated to have lower response propensity. This targeted incentive lowered response rates relative to a uniform incentive to all respondents and did not improve the representativeness of the survey relative to population benchmarks. In contrast, the mixed-mode 2015 NSCG used a targeted incentive for those who are less likely to participate and contribute significantly to the final estimates through having a large weight (National Academies of Sciences, Engineering and Medicine 2018). This targeted incentive (and other targeted interventions) improved representativeness of the final respondent pool (Thieme and Reist 2017). More work is needed in self-administered and mixed-mode surveys on optimal allocation of resources for incentives.

6.7 Tracking Contacts in All Modes

One challenge to surveys when implementing a mixed-mode survey design is keeping track of the contact attempts via recorded paradata for each sampled case in each mode. When using a mixed-mode survey, especially with an adaptive design strategy during which interventions occur, it is important to consider what kinds of measures will be used before, during, and after data collection to effectively evaluate response rates, representativeness, and data quality. Some surveys may simply want to keep track of what was done for contacting and gaining cooperation with each sampled unit, and plan for analysis of the data after the data collection period is over. Other surveys may want to produce estimates of interest and quality measures regularly during data collection to help with data monitoring and to measure the impact of interventions. As such, **having data collection systems that effectively track what contacts cases have received and ensure interventions are properly employed are critically important. Having systems that talk to each other across multiple modes and also permit real-time analysis of data collection may be challenging or require significant infrastructure development at survey organizations.**

Challenges in managing mixed-mode data collection systems may be especially present for smaller survey organizations. Smaller organizations may use off-the-shelf web or telephone survey software systems that do not easily permit managing the number and types of mailings and contact attempts across modes, especially for non-computerized modes (e.g., mail). Additionally, off-the-shelf software systems are limited in the number and types of analyses that can be done regularly. **As such, survey organizations may manage and evaluate the mailings and web-based contacts in different files, using Excel, SPSS, SAS, or other spreadsheet-style programs for analysis and reporting.** For instance, Murphy, Biemer, and Berry (2018) report using SAS and Excel to create daily reports for monitoring the RECS's mixed-mode data

collection. These reports include response rates plotted on graphs throughout data collection and other field metrics, permitting evaluation of the effectiveness of different data collection strategies (see also Kreuter and Olson 2013). Although there are few available examples of using off-the-shelf software for mixed-mode surveys, lessons may be drawn from interviewer-administered surveys. For example, Kirgis and Lepkowski (2013) report using SAS and Excel to create visualizations for monitoring the in-person National Survey of Family Growth. Jans, Sirkis, and Morgan (2013) use SAS to create quality-control charts for the National Health Interview Survey.

Alternatively, some organizations may build in-house mixed-mode data collection systems, requiring substantial commitment of resources, planning, and extensive use of field managers, researchers, and IT professionals. For example, the NSCG required restructuring the entire paradata system to jointly manage and monitor mail, web, and telephone contact attempts, rather than manage each mode separately (Reist 2014). Other large survey organizations developed in-house sample and data management systems that track movement of cases through different modes and types of contact attempts, each requiring multiple years of planning and integration (Cheung and Maher 2015, Wernimont and Snowden 2015, Edwards, Maitland, and Connor 2017, Bonhomme 2018). Krzyzanowski, Qin, Robinson, and Sikes (2018) report on building extensive custom overlays to an existing off-the-shelf software system to manage a web and face-to-face mixed-mode survey. More research is needed on how to “best” design these systems, or how to integrate analyses with the field management systems (Schouten, Peytchev, and Wagner 2018, p. 103).

6.8 Sample Composition

One important question for switching to a self-administered or mixed-mode data collection is how well these data collection efforts reflect the characteristics of the target population. Telephone surveys systematically miss certain individuals who cannot be contacted, refuse to participate, or do not speak the language of the survey. Transitioning to a mail or web survey adds other potential causes of nonresponse, including low literacy, lack of internet access, and low familiarity with computers (e.g., Brick, Williams, and Montaquila 2011). Our goal in this section is not to review all of the literature evaluating nonresponse bias on self-administered or mixed-mode surveys. Rather, we look across a set of surveys we have identified that have transitioned from telephone to self-administered and mixed modes. These surveys are inconsistent in whether and how nonresponse bias is evaluated. In many instances, we lack information about nonresponse bias on these estimates prior to the transition to a self-administered mode. As such, we identify trends across these studies in demographic characteristics of who is over- or underrepresented in the self-administered or mixed-mode surveys, but cannot easily comment on whether these biases are better or worse than that for the same telephone surveys.

We focus first on demographic variables commonly used in weighting schemes – age, sex, and race. First, **mail surveys of the general population tend to underrepresent younger adults and overrepresent older adults** (Battaglia, et al. 2008; Han, et al. 2010; Klausch, Hox and Schouten 2015; Lesser, et al. 2016; NHES 2016; Mauz, et al. 2018), similar to recent unweighted telephone surveys (Keeter, et al. 2017). The degree to which age distributions in the respondent pool differ from that of the population differs somewhat across modes, where some (but not all) web or web with mail follow-up surveys yielded a higher proportion of younger adults or more representative sample based on age (Klausch, Hox and Schouten 2015; Biemer, et

al. 2017; Marken, et al. 2018; Wells, et al. 2018). Second, **there are few consistent patterns across studies in whether men or women are more likely to participate in certain self-administered modes of data collection.** In some studies, men are overrepresented in self-administered modes (Han, et al. 2010; Lesser, et al. 2016; Winneg, Ben-Porath, and Jamieson 2017; McPhee, et al. 2018), in other studies women are overrepresented (DeBell, et al. 2017; Breidt, et al. 2018), and in other studies, there is no difference (Klausch, Hox and Schouten 2015). Third, **racial/ethnic minorities are underrepresented in self-administered and mixed-mode surveys,** either when looking at the race/ethnicity of the respondent (Battaglia, et al. 2008; Link, et al. 2008; Han, et al. 2010; Brick, Williams, and Montaquila 2011; Kali and Flores-Cervantes 2016; DeBell, et al. 2017; Winneg, Ben-Porath and Jamieson 2017; Breidt, et al. 2018; Wells, et al. 2018) or in areas with higher proportions of racial/ethnic minorities (Cantor, et al. 2005; Cantor, et al. 2009; McPhee, et al. 2018). This is similar to underrepresentation of minorities in telephone surveys (Link, et al. 2008; Keeter, et al. 2017).

Next, we examine socioeconomic variables of education and income. **More highly educated individuals are systematically overrepresented in self-administered and mixed-mode surveys** (Battaglia, et al. 2008; Link, et al. 2008; Brick, Williams, and Montaquila, 2011; Lesser, et al. 2016; DeBell, et al. 2016; NHES 2016; Marken, et al. 2018; but see Wells, et al. 2018; Breidt, et al. 2018), similar to telephone surveys (Link, et al. 2008; Keeter, et al. 2017). **The representation across income levels is more variable, partially due to substantial variation in how income is operationalized across studies.** Some surveys show a higher proportion of higher income households participating in a mail or web survey than the population (Link et al. 2008; Lesser, et al. 2016; Marken, et al. 2018; McPhee, et al. 2018; Wells, et al. 2018), others show a greater representation of lower income households (Brick, Williams, and

Montaquila 2011; Amaya, et al. 2015; Kali and Flores-Cervantes 2016; Breidt, et al. 2016), some show simply discrepancies in the income distribution (Klausch, Hox, and Schouten 2015; DeBell, et al. 2017; Biemer, et al. 2017). Education levels may also affect web uptake rates (e.g., Lesser, et al. 2016; Steele, et al. 2016). For example, in general population surveys including the American Community Survey and Decennial Census Tests, the proportion of the sample that participants via web is typically between 30 and 40 percent (Baumgardner 2018; Bentley 2019). This compares to approximately 80 percent in surveys with a more highly educated population (e.g., the National Survey of College Graduates) (Finamore 2019).

6.9 Unique Issues in Transitioning Surveys from Interviewer-Administered Modes to Self-Administered Modes

Collection of Additional Information. Sample surveys are increasingly collecting biomeasures (e.g., height, weight, saliva, blood), consent to link to administrative records, and other physical environmental samples (e.g., dust, air quality, soil), or geocoded measurements, in addition to asking survey questions. In our sample of surveys that transitioned to self-administered or mixed modes, 5 of 24 organizations reported collecting information in addition to survey data. These additional requests for information included blood samples, consent for linkages to administrative data, and geographical information. Chapter 4 discusses collecting biomeasures and consent to link in more detail. This is an area in self-administered and mixed-mode surveys that needs more research.

Nonresponse due to Language Difficulties. As noted in Chapter 3, surveys transitioning from telephone to self-administered or mixed-mode surveys should plan strategies for successful recruitment of non-English speakers and readers. Self-administered surveys systematically underrepresent racial and ethnic minorities, including those that speak languages other than English (e.g., Brick, et al. 2012; Wells, et al. 2018). **Surveys that have successfully recruited Spanish speaking respondents have translated the survey materials into Spanish and included those Spanish-language materials (letters, questionnaires, other survey information) in mailings from the very first contact attempt** (e.g., Brick, et al. 2012; Jans, et al. 2013; Amaya, et al. 2015; Blake, et al. 2016; Skalland, et al. 2017; Ghandour, et al. 2018; McPhee, et al. 2018). Web surveys that also permit the respondent to “toggle” between English and Spanish-language questionnaires facilitate representation of Spanish-speakers in a questionnaire (e.g., Kennedy, et al. 2016; ANES 2018; Ghandour, et al. 2018). Offering a language-specific telephone line for non-English speaking sample members is less successful (e.g., Cantor, et al. 2009; Wells, et al. 2018).

Eligibility Rates. **Even when the target population for a study does not change, the observed eligibility rates may differ across modes due to differential nonresponse or other error sources.** This may lead to differential sample yield from what was observed in the telephone survey, differential estimates of coverage and/or eligibility depending on the mode or combinations of modes, and may require different numbers of contact attempts depending on the mode selected for data collection. Early studies that transitioned telephone to mail questionnaires had a goal of adequately covering the cell phone population. For example, Link, et al. (2008), in examining the potential for a mail-based BRFSS, found that 6.5% of the mail survey respondents

self-reported being cell-only and 1% had no telephone at all, aligning well with benchmark estimates from the National Health Interview Survey of 6.7% and 1.7%, respectively.

Additionally, the proportion of “valid” elements on the sample frame may vary across modes, either due to quality of the sampling frame, differences in how materials are delivered to those sampled units, or differences in the time domain for estimating the potential eligibility rate. For example, the NSCH estimated that 11% of the addresses would be non-residential or undeliverable, but found that 16% of addresses were confirmed to be undeliverable or nonresidential, with another 5% estimated to be undeliverable or nonresidential (US Census Bureau 2018). In the NHES, McPhee and Zuckerberg (2018) report that about 10% of the addresses in a sample are designated as undeliverable as addressed at some point during the data collection period, but that about 15% of these potentially undeliverable addresses actually return a completed mail questionnaire.

In surveys that target particular subgroups, screening eligibility rates can differ across modes of data collection and sample frames. For example, the National Survey of Veterans found that a mail-based screening instrument yielded better coverage of the target population of veterans (59.6%) than a web-based screening instrument (46.5%), and that including an informative paper insert in the mail-based group increased the effective coverage rate even further (66.1%) (Han, et al. 2010). These differences could not be attributed solely to different response rates across the groups. The mail-based 2011 NHES found that 32% of households had eligible children, slightly under the 35% estimated from the American Community Survey (ACS) (Montaquila, et al. 2013). This eligibility rate is similar to the 2009 NHES pilot study that found that 31% of households had children in eligible age ranges, compared to 35% in the ACS, but that children aged 1 year old and under were substantially undercovered (Brick, et al. 2011).

When the NHES added a web component to the existing mail survey, eligibility rates and response rates differed across the modes of data collection (McPhee, et al. 2018). The Coastal Household Telephone Survey yielded around a 10% eligibility rate for general population households that engage in recreational saltwater fishing; a mail-based version had two frames – a license-based frame (which was included to screen more effectively) and an ABS sample that was matched to the license frame. As anticipated, eligibility was higher in a fishing license-based frame (37.2%), followed by an ABS sample that could be matched to the license frame (21.9%), followed by an ABS sample that could not be matched to the license frame (6.6%) (Andrews, Brick, and Mathiowetz 2013). The mail survey yielded a much higher estimate of fishing prevalence than the telephone survey. Similarly, an evaluation of the National Survey of Fishing, Hunting, and Wildlife-Associated Recreation found much higher incidence rates of fishing, hunting, and wildlife-watching in a mail-based approach than a face-to-face approach to the survey, a difference that could not be attributed solely to differences in screening decisions (Breidt, et al. 2018). More research is needed to understand how decisions made in each mode affect eligibility rates for surveys of different topics.

6.10 Summary and Takeaways

6.10.1 In surveys conducted in the early 2000 through about 2012, telephone survey response rates tended to be higher or at about the same level as the self-administered or mixed-mode surveys. After about 2013, the self-administered or mixed-mode surveys generally had response rates that exceeded the same survey's telephone survey response rates. Recent experiments have tested single, concurrent, and sequential mixed-mode designs and found few differences in response rates.

- 6.10.2 The most common recruitment mode among surveys that have transitioned to self-administered or mixed modes is mail.
- 6.10.3 Telephone is still included as one of the modes of data collection, either as a primary data collection mode or a follow-up mode. Mail can be used to obtain telephone numbers for households where an address cannot be linked to a telephone number through a reverse directory telephone match, and as a follow-up mode for nonrespondents where a telephone number can be linked to a sampled address.
- 6.10.4 Mixed-mode studies mail a URL and login information for sampled addresses to complete the survey online. These surveys often include a mailed paper questionnaire in follow-up mailings, sometimes referred to as a web-push design.
- 6.10.5 Use of email to recruit sample members is limited to those studies using a special population selected from a list containing email addresses (e.g., students; employees), from probability or non-probability web panels, for studies that used a screener survey to collect an email address, or in longitudinal surveys where an email address has been obtained at a prior wave.
- 6.10.6 Text message invitations and reminders are rarely practical for one-time surveys, but can be particularly useful for panel and longitudinal surveys where texting consent can be obtained.
- 6.10.7 Longitudinal household surveys now often include a self-administered mode or a mixed-mode data collection strategy for at least some of the follow-up waves.
- 6.10.8 Planning for active monitoring of multiple metrics during data collection is especially important when using experiments to identify the “best” mode or combination and sequencing of modes going forward and there is no prior data to use to conduct initial

data analyses to understand the impact of these decisions. Having data collection systems that effectively track contact modes and attempts, and ensure interventions are properly employed, is critically important, although this may be challenging or require significant infrastructure development at survey organizations.

- 6.10.9 Repeated cross-sectional surveys or longitudinal surveys that have previously had a mix of modes in data collection can use statistical analysis and simulation of existing data to plan interventions in a new round of data collection.
- 6.10.10 There is no single mailing protocol used for self-administered or mixed-mode surveys, although surveys that include a separate screening questionnaire and topical or main questionnaire use very similar mailing protocols.
- 6.10.11 Prepaid incentives of \$2 and \$5 are common incentive levels. Promised incentives are less commonly used, with the monetary levels of promised incentive levels much higher than prepaid incentive levels.
- 6.10.12 Self-administered surveys tend to underrepresent younger adults and racial/ethnic minorities and overrepresent older adults and adults with higher levels of education. There is less consistency in the quality of representation across gender and income categories.
- 6.10.13 Surveys that have successfully recruited Spanish speaking respondents have translated the survey materials into Spanish and included those Spanish-language materials (letters, questionnaires, other survey information) in mailings from the very first contact attempt.
- 6.10.14 Even when the target population for a study does not change, the observed eligibility rates may differ across modes due to differential nonresponse or other error sources.

Appendix Table 6.A: Citations for Response Rates for Surveys Conducted in Both Telephone and Self-Administered or Mixed-Mode Data Collection Modes

Survey Name	Year		Source
	Interviewer-administered	Self-Admin / Mixed Mode	
2005 Behavioral Risk Factor Surveillance System six-state pilot	2005	2005	Battaglia, et al. (2008); Link, et al. (2008)
2005 Health Information National Trends Survey (HINTS)	2005	2005	Cantor, et al. (2005)
2006-2014 ODOT surveys	2006-2008	2006-2014	Lesser, et al. (2016)
2007 Health Information National Trends Survey (HINTS)	2007	2007	Cantor, et al. (2009)
National Household Education Survey: 2009 Pilot Study	2007	2009	Brick Williams, Montaquila (2011)
National Survey of Veterans	2001	2009	Han, et al. (2010)
Dutch Crime Victimization Survey	2011	2011	Klausch, Hox, and Schouten (2015)
National Household Education Survey: 2011 Field Test	2007	2011	Montaquila, et al. (2013)
Survey of Consumer Attitudes	2011	2011	Elkasabi, et al. (2014); Survey of Consumers (2012)
Racial and Ethnic Approaches to Community Health (REACH) U.S. Risk Factor Survey, Phase 3	2011	2011	LeClere, et al. (2012)
German Health Update 2.0 (GEDA) pilot study	2012	2012	Mauz, et al. (2018); Hoebel, et al. (2014)
Wisconsin Family Health Survey	2011	2012	Allison, Stevenson, and Kniss (2014)
2013 California Health Interview Survey ABS pilot	2013-2014	2012	Jans, et al. (2013); California Health Interview Survey (2016)
American National Election Studies 2012 Time Series Study	2012	2012	American National Election Studies (2015)
Coastal Household Telephone Survey	2013	2013	Brick, Andrews, and Mathiowetz (2016)
2013-2014 California Health Interview Survey ABS pilot	2013-2014	2013-2014	Kali and Flores Cervantes (2016)
2015 Residential Energy Consumption Survey National Pilot study	2009	2015	Biemer, et al. (2017); Residential Energy Consumption Survey (RECS) 2009 Technical Documentation Summary (2013)
2015 Canada Election Study	2015	2015	Breton, et al. (2017)
CAHPS Hospice Survey	2015	2015	Parast, et al. (2018)
2015 New York Adult Tobacco Survey	2015	2015	Brown, et al. (2018)
2016 American National Election Studies Time Series Study	2016	2016	American National Election Studies (2018)
National Travel Survey pilot	2016	2016	Bosa, Gagnon, Caron (2017); Statistics Canada (2018)
National Immunization Survey	2016	2016	Skalland, et al. (2017); CDC (2017)
National Survey of Fishing, Hunting, and Wildlife-Associated Recreation	2016	2016	Breidt, et al. (2018)

Survey Name	Year		Source
	Interviewer-administered	Self-Admin / Mixed Mode	
2016 National Survey of Children's Health	2011-2012	2016-2017	Ghandour, et al. (2018)
2017 National Household Travel Survey	2009	2017	US Department of Transportation (2011); Federal Highway Administration and Westat (2018)
Gallup Sharecare Well-Being Surveys	2017	2018	Marken (2018)
2018 California Health Interview Survey Push-to-web pilot	2017	2018	Wells, et al. (2018)

7 Data Preparation, Processing and Management

7.1 Introduction

A review of the current literature regarding transitions from single mode data collection efforts to mixed-mode data collection provides little experimental or empirical data with respect to how such transitions affect data processing. As a result, this section of the report will provide little experimental or empirically-based guidance; rather, our intent is to raise issues that could affect data quality, timing, and costs. We will draw on current examples of data preparation and processing used for mixed-mode surveys, but cannot evaluate alternative methods for doing data processing where none exist. We list factors to be considered as surveys move away from a single mode data collection to mixed-mode data collections.

As has been already noted (see Chapter 4), part of the challenge of mixed-mode data collection efforts that span varying levels of technology is whether and to what extent that technology is utilized in the capture of the data, when possible. For example, consider a mixed-mode data collection that utilizes both Internet web response and mail questionnaires, such as the current design for the U.S. American Community Survey (ACS). While range checks, validation, and data edits can be incorporated into the web-based instrument, these are not feasible in the paper format. Because these processes cannot be incorporated into both modes, designers must decide whether to maintain quality checks in the technology-assisted instruments and to what degree. This is a key data management issue—whether or not to take advantage of technology that could potentially improve data quality at the cost of varying data quality across the different modes. For example, does one integrate a range check for the web-based data collection, when such an option does not exist for the paper version?

In our convenience sample of organizations that have transitioned a survey from telephone to self-administered or mixed modes, eight of the respondents said that data editing for their project varies by mode; 11 said it did not. Open-ended responses about data editing provide no consensus about this aspect of the surveys. Respondents said to “be meticulous” and make sure multiple people examine the data, build edits into and thoroughly test the data capture process, save original files from the discrete mode sources and review each step of the process.

As a field, we have not uniformly identified, researched, and addressed basic philosophical questions about data processing in a mixed-mode survey environment. Is the goal of data processing in a mixed-mode data collection environment to have the resulting data blind to data collection mode or to preserve those differences that may arise from the mode of data collection? Should those differences be noted and addressed with the same rigor as response rates and sampling error? Another philosophy of data processing and management is to try to ultimately achieve comparable levels of quality across modes, and using the best methods within each mode to achieve it. A construct for quality in this case could mean comparable levels of consistency, completeness, and coherence. The concept of ‘comparable’ quality with mixed-mode design is open for discussion and research.

Regardless of the mode by which data are collected, once captured, the data often move through a process involving multiple stages before it is utilized for analysis. Drawing on the Generic Statistical Business Process Model developed by the United National Economic Commission for Europe (UNECE 2013), we organize this section of the report according to the following processes:

- Data integration
- Classification and coding

- Review and validation
- Editing and imputation
- Calculation of weights
- Finalization of data files

To illustrate the data preparation and processing, we expand on the ACS model of data preparation and processing (U.S. Census Bureau 2014, page 105, Figure 10-1). Figure 7.1 depicts the overall flow of data as they pass from data collection operations through data preparation and processing.

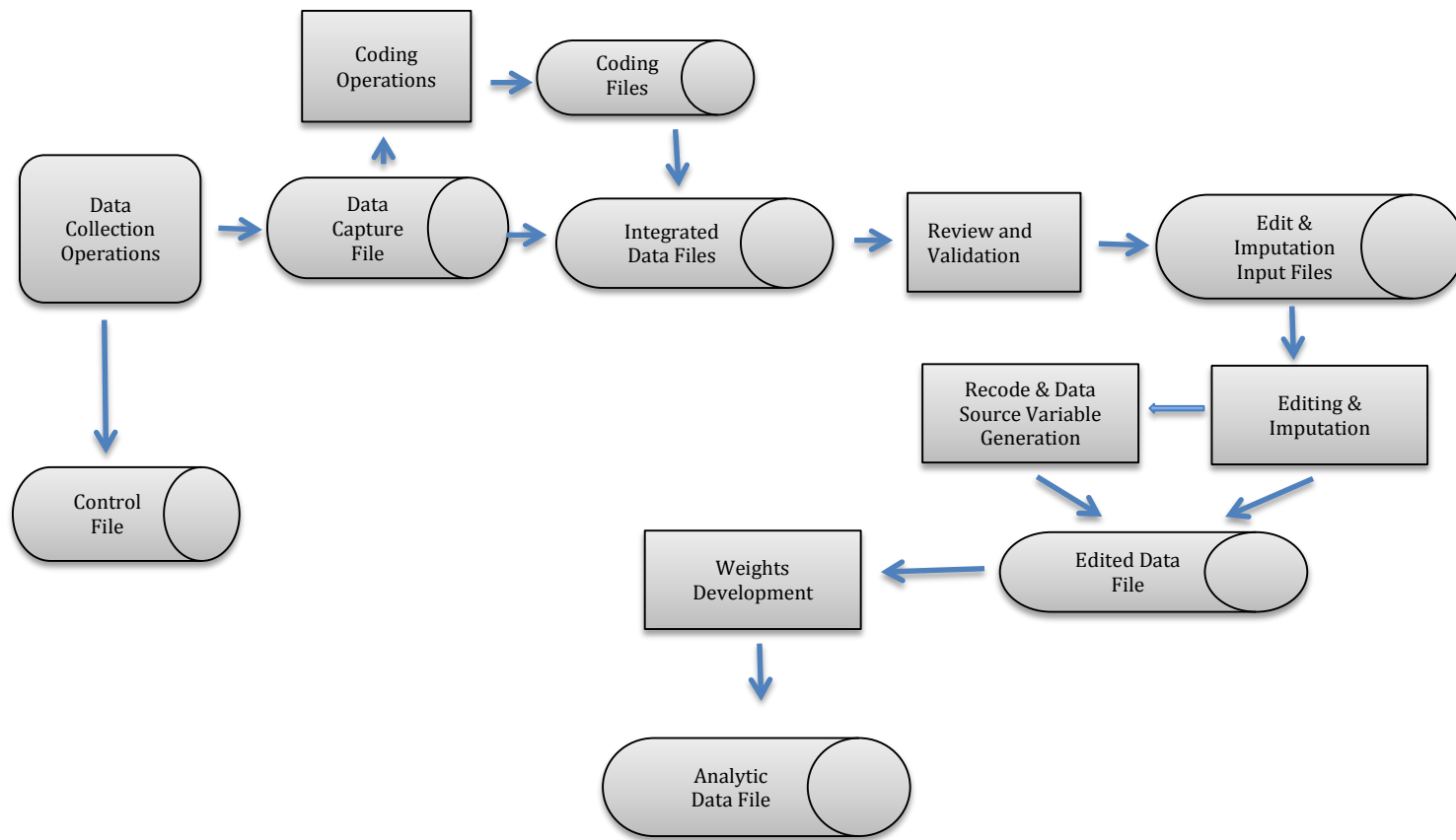


Figure 7.1: Data Processing Flow Chart, adapted from U.S. Census Bureau 2014, page 105, Figure 10-1

7.2 The Importance of Transparency

Before discussing the stages of processing data, the importance of transparency must be noted with respect to mixed-mode data collections. Transparency – as we use it here – refers specifically to “the availability of documentation for a given estimate or data set that identifies the data sources and potential error associated with the methods of data collection, processing and estimation” (Martinez 2018, p. 2). **That transparency extends not just to the identification of a data source (e.g., survey mode; administrative or survey data or other source) but also to the identification of alterations to data, such that analysts can make informed decisions as to the pooling of data across modes as has been the practice for decisions to use raw or edited/imputed data.** For example, every case in a mixed-mode data collection effort should identify the mode in which it was collected or, for non-survey sources of data - including auxiliary and administrative data - the source should be named for the case. Similarly, variable flags should indicate if the analyst is utilizing respondent reported data (sometimes called raw data), or edited data or imputed data that also identifies the mechanism by which the data were edited or imputed (e.g., logical edited, hot deck imputation). We note that clear and transparent documentation of some survey data processing steps are not currently part of the Disclosure Elements in the AAPOR Code of Ethics (e.g., data entry; data edits; mode for each case), but others are (e.g., weight calculations; deduplication efforts; validation). With a move to mixed-mode data collections wherein processing rules may vary across modes, we recommend that survey organizations also include any survey processing, coding, editing, imputation, and finalizing of data sets in technical documentation, and especially how these decisions vary across modes, for clear and transparent documentation.

7.3 A Note on Data Quality Control

Throughout this chapter, we refer to manual and automated processes. Regardless of the processes and the original mode of data collection, best practices with respect to data processing involve the integration of quality control measures, such as independent verification and validation. These quality control measures include but are not limited to the use of double entry or scanning for paper questionnaires, independent coding for open-ended items, and the monitoring of interviewer-based data collection efforts, including reinterviews. We do not address these issues in the text that follows, since we believe that such quality control measures are independent of the issues related to mode of data collection and should be integrated regardless of how the data were originally captured.

7.4 Data Capture and Integration

The first step in data processing is to first capture data in electronic form (we will call this *data capture*), regardless of mode of data collection. In computer-assisted interviewer-administered modes (e.g., computer-assisted telephone interviews [CATI], computer-assisted personal interviews [CAPI]), the interviewer enters answers to questions directly into a survey instrument that is then transmitted to a central data system. In web surveys, the respondent enters answers directly into a web-based instrument. In paper surveys, the respondent enters answers into a paper form that is then transformed into electronic data through data entry. Later in the processing of survey data, other data (e.g., administrative records) may be integrated with those data. Data integration also occurs when data from multiple modes are combined into one analytic

data file. Although we begin with data integration, integration of data may occur at any point in the processing and may, in fact, be iterative.

Given varying levels and use of technology in mixed-mode data collections, key initial decisions involving data capture need to be made. For mixed-mode designs involving fully automated data capture via data collection software (e.g., CATI, CAPI and Web), it may be possible to utilize similar or identical systems in the data capture. Data processes, flows, and management become more complex when the modes involved in the data capture incorporate different levels of automation (e.g., web and paper) or different software platforms that used for different modes (e.g., the CATI software is different from the Web software).

As one considers the data entry phase for mixed-mode data collection efforts, the first decision made by survey organizations is whether the processing will involve a single unified data entry system for all computerized modes and data entry of paper forms or multiple systems that vary across modes of data collection. The use of a single integrated system may simplify decision-making as well as allow for consistent rule setting and coding with respect to data collection and entry. This may come, however, at the cost of reduced flexibility in system and user interface design – for instance, data entry of paper questionnaires into a web survey software may not permit the types of invalid responses often seen in mail questionnaires (e.g., marking two response categories), thus losing potential information about the quality of data in one mode. Additionally, smaller survey organizations may only have access to one type of survey software for data collection and need to retrofit it to data entry for a mail survey or use other off-the-shelf data entry (e.g., Epi Info) or spreadsheet (e.g., Excel; SPSS) software for data entry of mail surveys that is not used for other modes of data collection. In large organizations that build in-house systems, different IT teams are often used to develop and program systems

for data capture in different modes. As such, integration of technology systems within survey organizations also may not be straightforward.

A second consideration for any mixed-mode collection that includes paper instruments will be how data are captured and processed. Paper forms may require initial form reviews, manual data entry and re-entry processes, or may rely principally on the use of scanning technology. While raw data from computer-assisted information collections is stored and archived, whether paper forms and/or their images are also archived varies. The U.S. Decennial Census, for example, has preserved all scanned images of 20th century census forms, but largely for recordkeeping and genealogical purposes. In contrast, it is not the practice of most recurring federal surveys to scan or store images of data collected via paper forms; rather, only the electronic capture of useful data from these forms is stored and the forms themselves are securely discarded.

In addition to capturing the response data, mixed-mode data collection should capture sufficient paradata about the collection device used and other aspects of the response process to inform processing, editing and analysis of the data. **At a minimum, the mode used to complete the survey should be reported on public use data files. Additionally, for mixed-mode surveys that include a web component, paradata that permit the analyst to identify the device that is used (e.g., laptop/desktop, mobile phone, mobile tablet) are also important to capture and report in a public use data file for future analysis.**

Following data capture, a key decision is the determination of which cases qualify as “complete.” Rules and practices for such determinations are typically project-specific or driven by an organization’s standards and processes. With the introduction of mixed modes, what constitutes a “complete” case may or may not vary across modes. Clearly, self-reported answers

on paper questionnaires complicate the decision rules and assumptions driving how multi-mode data should or can be integrated; responses on paper forms may have higher rates of missing or inconsistent data than web surveys (e.g., Dillman, 2012). Regardless of the modes involved in the data collection process, *a priori* assumptions and decisions rules should inform whether data collected are sufficient to be considered complete. Sampled units (households, people, establishments) that do not meet the criteria for a completed interview can either be treated as partial interviews or as noninterviews. Treatment as partial interviews complicates the remaining stages of processing (as well as analysis); treatment as noninterviews implies no further data processing of such cases. **Rules that define complete, partial, and insufficient, and out of scope cases should be reported, including whether these rules differ across modes, and made available to analysts through the survey's technical documentation.**

Steps to deduplicate cases in mixed-mode studies are often necessary. Duplicates may occur in single-mode data collection efforts (e.g., a sampled case completes two mail surveys), and may be even more likely in mixed-mode data collection efforts. For example, when one mode is used sequentially for nonrespondents, a sampled case may participate in a survey using the first mode offered while the survey organization is processing the materials to request respondents to participate via the second mode (e.g., a sampled case completes both a web survey and a follow-up mail questionnaire). After the universe of acceptable interviews is determined, cases should be reviewed to ensure that multiple responses were not received from the same sample unit.

In both study types, the sequencing of information requests is often staged relatively proximate in time. Duplicates, therefore, may arise from responses to both the original and subsequent request for data (e.g., original and replacement mailings; original Internet return and

subsequent paper questionnaire). In the case of household or establishment surveys, it is possible to have different members or employees respond to each request, by different modes, with different data levels of completeness. For example, one household member may participate via the Internet-based form, and before that form is received, another may respond to the subsequent mail request. Whether different members of sampled units have different levels of understanding or knowledge to report about a topic and that interacts with mode preferences or accessibility for survey requests is open for study, especially in establishment studies.

Surveys use different decision rules for deduplication of completed survey responses. For example, the 2016 National Household Education Surveys (NHES) had options to participate both through mail questionnaires and web instruments (McPhee, et al. 2018). Deduplication rules included: (1) Keep the first screener that was sufficiently completed for within-household selection, unless there was a second paper screener sent in the same week that had more information; (2) If duplicate mail questionnaires were received, keep the questionnaire that had more information completed; and (3) Keep the questionnaire of any mode that was more complete. For the topical surveys, there were also rules about keeping the first questionnaire if duplicates were received that had identical numbers of items completed, and decision rules about what to do if both paper and web screeners were completed with different topical surveys completed. Similarly, the National Survey of Children's Health (NSCH, US Census Bureau 2018) prioritized a completed questionnaire in any mode, but selected the completed web questionnaires if both web and mail questionnaires were returned and completely filled out.

Once again, **rules about completed interviews and deduplication will need to be established *a priori* as to which record to use, especially if the data from the two modes of data collection differ.** Approaches to choosing which record to use could be based on objective

measures of data quality, e.g., percentage of complete items against those items that should have been answered, a preference for a specific mode of response that provides higher quality data, or the date of the response. These simple rule-based measures may be inadequate for complex topics and studies, with multiple respondents or with multiple sub-questionnaires.

7.5 Classification and coding

Respondents provide write-in responses to either open-ended questions or when a response set has pre-coded response options but also allows a respondent to write-in a response not listed in the closed set (e.g., “other, please specify” questions). Regardless of the mode of data capture, once these items are stored in an electronic format, the raw text needs to be recoded according to a reduced or prescribed list of options facilitate statistical analysis. Coding can be completed as an automated process, i.e. via a series of computer programs to assign codes, coded by hand, or a combination of the two processes. For example, the coding of occupation has for years been completed manually; current developments include the use of automated-assisted coding systems for standardized coding such as occupation and industry (e.g., Russ, et al. 2016) and the use of automation for categorization of open-ended text (e.g., Schonlau and Couper, 2016).

The decision to use an automated system for text processing, a human-based system, or a hybrid will be determined not so much by the mode of data collection, but rather the nature of the data to be coded. The amount and quality of recoding possible, however, may vary somewhat by mode. The key issue may be that an interviewer, trained in gathering information sufficient for coding, influences the quality of responses and is not present across all modes of collection. Of course, interviewers are not free from making errors in recording answers, with error rates for

interviewers recording open-ended questions up to 24 % or higher (Rustemeyer 1977; Fowler and Mangione 1990; Lepkowski, et al. 1995; Mitchell, et al. 2008; Strobl 2008; Smyth and Olson *forthcoming*). Nevertheless, surveys that developed coding rules based on trained interviewers may need to develop separate coding rules for self-administered modes if the amount of information differs, or write different questions specifying exactly the material needed to facilitate similar coding strategies. New rules or processes may be especially needed for automated coding programs. For example, the ACS anticipated that new rules may be needed when coding web survey responses to open-ended questions (e.g., occupation) because respondents can use punctuation marks (e.g., “:”) that are not used by interviewers and are not data entered in a mail survey (US Census Bureau 2014). As such, an additional cleaning step for web text may be needed to facilitate use of existing automated coding rules. Apart from instances where the automated coding has been integrated with an electronic data capture (CATI, CAPI, Web-assisted), surveys that are transitioning from telephone to self-administered or mixed modes should try to develop data capture processes that have little to no impact on classification and coding processes across modes, or anticipate where those rules may need to be modified.

7.6 Review and Validate

The UNECE identifies validation as the process by which “potential problems, errors and discrepancies such as outliers, item-nonresponse, and miscoding” occur as a separate step, distinct from data editing (UNECE 2013, p. 18). For example, this may be the stage at which range checks that were integrated into an automated data collection instrument are reapplied to an integrated, mixed-mode data file in order to identify cases, collected using less automation or without these built-in range checks, fail the expected set of plausible values.

When the survey industry adopted automated data collection methods (CAPI, CATI, Web-based), the use of technology in a questionnaire moved the identification of data outliers and inconsistencies forward from a largely post-processing activity into the data collection instrument itself. Logical, range, and consistency checks and correct routing of skip patterns are often now integrated into the questionnaire design process itself. These checks can be simple (e.g., age cannot exceed 115 years old) or complex. For example, in the US National Household Travel Survey (NHTS), respondents report trips that they took during a day by mode of transportation. If a person reports taking a trip in a car, logic checks built into the web questionnaire automatically identify whether the person in the household was also reported as being a driver in a different part of the questionnaire (Federal Highway Administration and Westat 2018). In a mixed-mode survey where mail questionnaires are part of the data collection process in addition to computerized questionnaires, the data collected on paper forms are likely to have more inconsistent responses, data outside normal ranges, skip and logical errors than those collected via the computerized instruments when the computerized instruments also include validation and edit checks.

How the various quality checks are articulated and implemented in mixed-mode data collection efforts reside with people who have different roles in stages of the survey process: the researcher, the questionnaire designer, the programmer/data manager, and the analyst. Work formerly completed by a survey designer in a single mode may now be spread across a many staff with differing responsibilities. As such, mixed-mode data collections require more deliberate coordination. Data management may need to be more centralized in order to identify data processes (e.g., range checks) across modes and stages (data collection or data processing).

This would increase the likelihood that data anomalies are avoided, or flagged and addressed in a consistent, coherent, and efficient manner.

7.7 Editing and Imputation

Making inferences from a sample to a population requires robust statistical methods. Before those processes begin, any remaining data quality issues such as missing information and inconsistencies must be fully addressed. Where data are considered missing, incorrect, or unreliable, data providers may choose to replace the missing or incorrect data through the application of one more replacement techniques. The General Statistical Business Process Model identifies the following substeps (UNECE 2019, p. 21):

- “Determining whether to add or change data;
- Selecting the method to be used;
- Adding/changing data values;
- Writing the new data values back to the dataset and flagging them as changed;
- Producing metadata on the editing and imputation process.”

At the editing and imputation stage, data collected from multiple modes creates new challenges. As discussed more completely in the questionnaire design chapter, item nonresponse rates vary between self- and interviewer-administered modes. For example, Griffin reports missing data rates for income on the 2006 ACS varied from a low of 7.7% among respondents who completed the survey by mail compared to a 20.2% missing data rate for the same item among those who responded via computer-assisted personal interview (Griffin 2009). In the 2015 National Content Test for the 2020 Decennial Census, the item nonresponse rates for the phone administration of a relationship question in the household roster were about 1.4%,

compared to between 0.5% on the web and about 0.6-0.8% in the mail questionnaire (Seem and Coombs 2017). In part, the differences in missing data rates may be a function of the differences in who responds by mode. In addition, the presence or absence of an interviewer may impact the fullness of a response, patterns of missing data and other measures of data quality. Finally, different prompts to the respondent in different modes (e.g., “this item does not have a valid response”) may also affect item missing rates and thus imputation needs across modes.

In the spirit of transparency, we recommend that both single and mixed-mode data collection efforts develop a data-source taxonomy, documenting how different variables are created and/or the source from which they derive. For example, one data source taxonomy is simply the mode and/or device from which each interview or variable is collected in a public or restricted use data file. For instance, the mixed-mode National Survey of College Graduates (NSCG) public use file contains a variable SRVMODE that indicates whether the questionnaire was completed by mail, telephone, web, or a “telephone interview using [the] web instrument” (SESTAT 2018). The ACS contains a variable called RESPMODE that documents whether the data came from mail, CATI/CAPI (combined), or Internet modes (IPUMS USA n.d.).

A second taxonomy could document which variables are comprised of raw data as reported by the respondent, data that was algorithm or machine corrected (e.g., items must add up, logical relationships between variables must hold), clerical/analyst corrected data (logical edits or assignment based on other data as reported by the respondent), imputed data (for missing data), calculated or modeled data (beyond imputed, for example, household income as percent of the poverty rate), disclosure-protected data for which the data fields can be revealed (e.g., top coded; collapsed into categories). Where possible, including flags for edits and imputations are

conducted is also a useful set of information for analysts, especially if different modes of data collection lead to different data values due to editing and imputation differences for the same variable across modes. We note that some surveys regularly release information about edit and imputation flags and software code for creation of variables – for example, the National Household Education Surveys (NHES) includes imputation flags as part of the public use data files, as well as SAS code for created variables in the technical documentation and data user’s manual (McPhee, et al. 2018). These types of taxonomies would document metadata about how the variables are created and that may be relevant in analysis for users outside the survey organization that collected the data.

Similarly, information concerning variations in the implementation of a data collection instrument across modes (e.g., randomization of response options for computer-assisted instruments that cannot be implemented for paper instruments) should be preserved for data analysts and considered in the data processing, editing and imputation processes. For example, in the 2016 NHES, there were a number of items that were not asked in the web instrument based on skip patterns that were present in web, but not in the mail questionnaire (e.g., child grade and date of birth confirmed based on screener data in the web survey, but reported by the knowledgeable adult in the paper questionnaire; MCPhee, et al. 2018). As a result, edits were needed to combine the paper and web data into a single analytic data set for further editing and processing. In the NSCH, questions that were the filter question in a skip pattern ordered the “no” and “yes” options in a different order for the web and mail questionnaires, and thus the data-entered mail questionnaire data had different values corresponding to “no” and “yes” than the web questionnaire data. As such, the mail questionnaire data had to be edited to have the

“no” and “yes” responses have the same numeric values as the web questionnaire data (US Census Bureau 2018).

Analysts may want to consider using mode of data collection as well as other auxiliary data to inform data editing and imputation procedures. In response to the Task Force’s request for advice on best practices for data editing in mixed-mode surveys, one respondent suggested that editing be specific to the mode of data collection; attempting to simply import a single editing approach to save time and cost would likely achieve neither goal. However, no empirical data was provided to support that assertion. To date, we know of only a few studies which account for data collection mode in imputation (Slud 2015; Wang, et al. 2017), with mixed results across these two studies about the effect of the use of mode as an additional class variable in the imputation models.

Additionally, different modes may yield different imputation challenges. One common problem that occurs in mail questionnaires, but not in other modes, is that respondents select more than one response on an item that requested only one answer. Survey organizations may have to make different decisions about how to deal with these responses depending on mode of data collection. For example, in the 2015 National Content Test for the 2020 US Decennial Census, respondents who selected every single checkbox in a race item were designated as “invalid” responses (and was the same code for those who wrote in invalid responses of “Martian” or “human”) (Mathews, et al. 2017a). In addition, in the mail Health Information National Trends Survey (HINTS-FDA) survey, on most of the items, a standard hot-deck imputation could be used to address missing data. On four items, multiple respondents selected more than one answer. Rather than select one of the answers provided by the respondents as the “correct” answer, all of the answers were set to missing and then imputed with a single answer

(Blake, et al. 2016). Documenting how these types of inconsistencies occurred across modes, as well as different edit and imputation decisions, is an important step for transparency in processing across modes in a mixed-mode survey.

7.8 Weighting

While new standards and processes may need to be constructed for imputation methods to offset quality differences induced mixed modes (e.g., widely different item response rates, correlations between missingness and mode or between mode and level and patterns of poor quality), there may also be times when changing the mode of data collection from telephone to self-administered or mixed-mode also changes the weighting approach. One issue that may vary is the determination of eligibility (as a function of the presence or not of an interviewer). The means by which to arrive at the determination of eligibility may vary by mode, where interviewer administered modes may have fewer cases with unknown eligibility than self-administered modes. At least one survey has taken that into account for weighting purposes, using a model to determine a predicted eligibility among those cases with unknown eligibility in a self-administered, but not interviewer-administered mode. Specifically, in the 2015 Residential Energy Consumption Survey (RECS), the eligibility adjustment (used in the determination of the final analytic weights) was calculated “differently for the in-person and web/mail cases. Unlike the in-person cases where eligibility was determined by field interviewers, the eligibility of the web/mail cases was determined by a propensity model based on survey responses and contact mailing status” (Energy Information Administration, 2018, p. 17).

Another key issue to address in the development of weights is whether to provide a single-integrated set of weights or separate weights that would facilitate analysis of subsets of the

data by the mode of responses. In part, that decision may be informed by the design of the multiple mode data collection effort. If analysts will be pursuing the comparison of data to determine “mode” effects, adjustment for the different population groups who respond across the various modes may motivate the necessity of creating separate weights for each mode. For example, the 2016 American National Election Studies (ANES) Time Series data collection releases six weights – two for the full sample (making inference for only the pre-election survey data or both the pre-election and post-election data), two for the face-to-face mode alone and two for the web mode alone (again, separate weights for pre-election and post-election data in each mode) (American National Election Studies 2018).

7.9 Finalizing Data Files

Most critical in the development of the final data files is the inclusion of metadata such as a variable to identify the mode (and device, if applicable) for the analyst as well as flags that capture variations in details of the data collection (e.g., randomization) for individual questions. To the extent that differential effort was exerted across different modes of data collection, paradata documenting the data collection process is also important to include in the final data files.

Documentation should also include any notes for potentially inconsistent responses that may appear when data are collected in one mode, but not in other modes. For example, in the Health and Retirement 2017 Consumption and Activities Mail Survey, the documentation for the data set notes (Health and Retirement Survey 2017, p. 9): “The data will have some inconsistency. Respondents did not always follow the correct skip patterns and in most instances

their “incorrect” answers were preserved. Likewise, there were times when respondents interpreted questions involving percentages in different ways.”

7.10 Special Considerations: Longitudinal Data

Longitudinal data raise particular issues with respect to data processing within a mixed-mode environment. These issues include but are not limited to timelines to facilitate subsequent data collection efforts as well as decisions related to the use (or not) of prior round data, i.e. dependent interviewing. All multi-wave data collection efforts must determine to what extent data will be processed and utilized in subsequent waves; multi-mode collections introduce more, and earlier stage considerations.

- Will sample units be allowed to choose different modes of data collection between waves? The extent to which respondents’ modes vary may induce arbitrary differences in data quality within and between waves.
- Will explicit considerations be made for how dependent interviewing is implemented?
- How will implementation be modified for data collection modes involving less technology-dependent modes?
- What data will be presented to the respondent for verification, used to control skip patterns, and utilized to resolve discrepancies across items or between waves? How do different mode support this process?

Clearly, prioritization of those elements required for subsequent data collection –and processed to a similar degree regardless of mode of data collection – will be important to maintaining comparable data collection timelines and quality across all modes of data collection.

7.11 *Summary and Takeaways*

As we noted at the onset of this chapter, there is scant empirical literature with respect to best practices in the processing of data collected via multiple modes. However, in our review, we have seen evidence or advice that suggest the following:

- 7.11.1 Conduct explicit planning of the data processing, editing, and imputation as part of the design efforts, so as to integrate such activities across the stages of data collection, with particular attention given to how to address data quality of modes involving varying levels of technology. Because design features may vary between modes, designers will have to reconsider what, where, and how data processes occur in the overall survey design to assure comparable quality between modes. Standard processes may be enhanced, diminished or moved to different phases. These would include logical, range, and consistency checks, design features that may rely on technology (e.g, randomization), and the use of skip-patterns.
- 7.11.2 Provide transparency and documentation so as to facilitate informed decision-making by data analysts, including but not limited to the mode and device used in the data collection, information on variation in administration across modes (e.g., randomization vs. nonrandomization), and detailed flags providing information about the nature of the edited and imputed data.
- 7.11.3 Engage in *a priori* decision-making that addresses the nature of a complete case, rules concerning deduplication, the use of single vs. parallel review, clerical and/or automatic

editing, and imputation (that is, treating the data as a single entity or parallel processing that recognizes the variations in mode of data collection).

7.11.4 Regardless of data integration, preserve the original data capture files, by mode of data collection.

8 Survey Estimation

Accounting for mode effects is an important and difficult issue in mixed-mode survey estimation. When surveys transition from telephone to self-administered or mixed modes, estimates may no longer be comparable over time and or across subgroups because of changes in multiple sources of survey errors. First, if different subgroups participate at different rates in different modes or different subgroups are covered differentially in sample frames, then there will be mode-specific errors of nonobservation, sometimes called selection effects, which may change survey estimates. Second, measurement errors also may change across modes (de Leeuw, Hox, and Dillman 2008). As such, a change in survey modes may create time trends for surveys in which estimates change due to both selection (combined nonresponse and coverage errors) and measurement effects, rather than true changes. Survey estimation in a mixed-mode survey context may adjust for (1) differential selection effects, (2) differential measurement errors, or (3) both selection and measurement.

Survey practitioners who transition surveys from telephone to self-administered or mixed modes should be aware of these multiple approaches in mixed-mode survey estimation. For example, the University of Michigan's Surveys of Consumers is evaluating methods for combining data from telephone, web, and mail modes (Curtin 2019). In this survey, the telephone survey data include longer and more detailed responses to open-ended questions compared than the self-administered modes. Yet the web-mail method is operationally less complicated and has higher response rates for the same survey budget. As such, this survey is currently examining methods to combine data from both interviewer- and self-administered modes instead of replacing the telephone data collection altogether, following methods that have

been developed and tested in the combining survey literature (Elliott, Raghunathan, and Schenker 2018).

Even in carefully conducted surveys, measurements are imperfect and measurement error can vary substantially across survey modes (which we will call *nonignorable differential measurement error* in this chapter). Statistical models can be used to account for mode effects in survey estimates and measures of uncertainty. How to do this will depend on the survey setting and in particular, on the availability of either an external “gold standard” measurement or side-by-side mode comparisons. Existing mixed-mode survey inference methods include adjustments for both differential selection errors and nonignorable differential measurement errors (Hox, De Leeuw, and Klausch 2015, 2017; Tourangeau 2017; De Leeuw, Suzer-Gurtekin, and Hox 2018; Suzer-Gurtekin, Valliant, Heeringa, and De Leeuw 2018). Any single survey or estimation method will have idiosyncrasies and address particular scientific questions of interest. The purpose of this chapter is not to outline a specific recipe for every possible idiosyncrasy. Rather, this chapter outlines some basics for diagnosis and estimation requirements in a mixed-mode survey context from existing literature. As a result, this chapter is slightly more technical than the other chapters of this report.

When transitioning from a telephone to a self-administered or mixed-mode survey, designers aim to (1) minimize mean squared error (MSE) of the self-administered survey estimates, independent of existing telephone survey estimates or (2) minimize the MSE of the self-administered survey estimates with respect to existing telephone survey estimates. The evaluation of goal (1) (prioritize the quality of the self-administered survey estimates) requires validation data to compute the MSE while in evaluation of goal (2) (prioritize comparability to the telephone survey estimates), this reference quantity is specifically substituted with the

telephone survey estimates. Although the MSE term includes both bias and variance terms, the variance term is largely a function of sample size and sample design. Therefore, we focus on evaluating the bias components of survey estimates in this chapter.

8.1 Null Hypotheses to Test in Telephone vs. Self-Administered Data Comparisons

Consider $\hat{\theta}_{self-admin}$ and $\hat{\theta}_{phone}$ are survey estimators for an estimand, θ , collected via a self-administered survey and telephone survey, respectively. These estimators include adjustments for unequal selection probabilities, nonresponse, and undercoverage for a target survey population, usually accomplished through weighting. When designing a self-administered or mixed-mode survey to minimize the MSE of estimates compared to the existing telephone survey estimates, the hypotheses are:

- (1) Evaluate the difference in the overall estimate between interviewer and self-administration:

$$H_0: \theta_{phone} - \theta_{self-admin} = 0 \text{ vs. } H_A: \theta_{phone} - \theta_{self-admin} \neq 0, \text{ where } \theta_{phone} \text{ and } \theta_{self-admin} \text{ are estimands under telephone and self-administered methods.}$$

- (2) Evaluate the difference in the change estimate between the interviewer and self-administration:

$$H_0: (\theta_{phone,t=1} - \theta_{phone,t=2}) - (\theta_{self-admin,t=1} - \theta_{self-admin,t=2}) = 0 \text{ vs. } H_A: (\theta_{phone,t=1} - \theta_{phone,t=2}) - (\theta_{self-admin,t=1} - \theta_{self-admin,t=2}) \neq 0, \text{ where } \theta_{phone} \text{ and } \theta_{self-admin} \text{ are survey estimands under telephone and self-administered methods and } t \text{ indexes time.}$$

When we formulate the mode comparison research question as in null hypotheses (1) and (2), the interest shifts from identifying of the overall bias in the estimates from each mode to comparing the relative bias from each mode (Peytchev, Ridenhour, and Krotki 2010, Elkasabi et al. 2014). In addition to evaluating relative overall bias, it is possible to decompose relative overall bias to each survey error source: coverage, nonresponse and measurement. This type of decomposition will help to further refine the design decision making.

8.2 Assumptions Made by Single Mode and Mixed-Mode Surveys about Mode-Specific Biases

Under the total survey error framework (Groves 1989), mode effects are typically classified as part of observational measurement errors. In single-mode telephone surveys, measurement errors are generally assumed to be smaller (that is, statistically ignorable) than the errors of nonobservation (e.g., coverage, nonresponse). As such, methods for estimation and inference from telephone surveys have largely ignored measurement errors, including creating weights and estimating appropriate design-based standard errors (e.g., Lavrakas, et al. 2017). In a single-mode telephone survey, the assumption that measurement errors are negligible arises from two sources. First, careful attention to survey design and execution, including editing of inconsistent or invalid responses, explicitly aims to reduce observational measurement errors. Second, with only one mode, any mode-specific biases are difficult to measure and are constant across the mode of data collection (but see Elliott and West 2015 for discussion of variable measurement errors related to interviewers in telephone surveys).

Surveys that transition to self-administered or mixed modes, however, may find this assumption of negligibility of measurement errors across modes to be problematic. In particular, problems for surveys that have transitioned from telephone to self-administered or mixed modes

may arise when differential measurement errors across modes of data collection are not ignorable. Recent experiences with mixed-mode surveys and with surveys that switched modes have found differential errors across modes (see Chapter 4). For example, the Pew Research Center, in transitioning from RDD surveys to the online American Trends Panel, found that telephone respondents were more likely than web respondents to report being satisfied with their family and social life, that certain social groups experience more discrimination, and that they talk with their neighbors and rate their health as excellent, for example (Pew Research Center 2015).

Thus, a major challenge for surveys that want to account for potential mode effects in estimation, referred to as a “mode effect,” is a difficult-to-quantify bias. Each variable collected in the initial telephone survey and the transitioned self-administered or mixed-mode survey yields a different bias term. In a single mode survey, such as a telephone survey, mail-only survey, or web-only survey, then the bias due to measurement is almost always completely unknown and cannot be adjusted for at the estimation stage. In surveys conducted in multiple modes, differences in responses between modes can be used to detect the presence of mode effects but do not provide a direct way to quantify the biases without extra information providing a “true” or “gold standard” value or other specially planned data collections.

Surveys that are transitioning from telephone to self-administered or mixed-mode surveys should consider what the design will be for the survey into the future. That is, will the future study design be a single mode (only web, only mail), a mix of web and mail modes, or a mix of web, mail, and telephone methods of data collection? A survey that is transitioning to a single mode (e.g., mail only) may select a weighting and estimation method that is different from a survey that is retaining telephone as a component of data collection. Furthermore, a survey that

generally assumes differences in measurement errors across modes are ignorable (that is, do not affect inferences from that survey) may draw on estimation strategies that come from a dual-frame estimation approach (Groves and Lepkowski 1985; Lepkowski and Groves 1986). Surveys that account for nonignorable differential measurement error may need to use other methods (Elliott et al. 2018). Additionally, some surveys may have data available to test mode effects (e.g., Jackle, Roberts, and Lynn 2010; Klausch, Schouten, and Hox 2017; Klausch, Schouten, Buelens, and van den Brakel 2017) through experimental comparison of parallel surveys or reinterviews, whereas other studies may require use of data as collected in two or more modes, necessitating statistical methods that follow observational data evaluation approaches (e.g., Suzer-Gurtekin, Valliant, Heeringa and de Leeuw 2018).

8.3 Diagnosing and Adjusting for Measurement and Selection Errors in Mixed-Mode Surveys

Statistical mode comparisons generally focus on disentangling the effect of measurement errors from selection errors (nonresponse and coverage) on the overall relative differences in estimates across modes. Multiple review papers cover examples of these analytical approaches (Hox, de Leeuw, and Klausch 2015; Klausch, Hox, and Schouten 2015; Hox, de Leeuw, and Klausch 2017; Tourangeau, 2017; de Leeuw, Suzer-Gurtekin, and Hox 2018; Suzer-Gurtekin et al., 2018). We use the Hox, de Leeuw, and Klausch (2017) terminology to distinguish between the steps in mixed-mode survey inference: (1) Design, (2) Diagnosis, and (3) Adjustment. This chapter specifically focuses on the diagnosis and adjustment steps. Specific data may be needed to facilitate the diagnosis and adjustment steps, and as such, should be considered at the design stage.

Diagnosing whether mode differences are due to differential measurement errors or differential selection errors requires data on both modes for the survey items of interest. Differential survey error comparisons are more straightforward when studies include designs to isolate and test specific errors. To evaluate and diagnose differential sources of selection and measurement errors in mixed-mode surveys, data need to be gathered through (1) “gold standard” or administrative data record systems, (2) parallel surveys, conducted in different modes on different respondents, sometimes called “benchmark” surveys (3) repeated measurements on the same respondents, in different modes, or (4) statistical modeling and analysis approaches.

8.3.1 Diagnosing using “Gold Standard” or Administrative Data

External administrative data systems that contain a “gold standard” measurement for important survey items of interest are ideal, but rare (Hox, de Leeuw, and Klausch 2017). For example, in examining a sequential mixed-mode experiment (telephone to mail vs. mail to telephone), Sakshaug, Cernat, and Raghunathan (2019) have data from driver history records maintained by the Michigan State Department of Motor Vehicles, including number and type of traffic offenses and number and type of accidents, as well as demographic variables. They also have one survey question (“have you ever had a traffic accident”) that matches the administrative record variables. They use the record data to identify *nonresponse bias*, calculating the difference in the estimate for the respondent pool from the full sample estimate, as estimated only using the record data. To evaluate *measurement error bias* with the record data, they calculate the difference in the estimate between the reports from the survey data and the estimate as calculated from administrative record data on respondents alone. Using administrative record

data, Sakshaug, Cernat, and Raghunathan (2019) conclude that using a mixed-mode design reduces nonresponse bias on the demographic variables, no matter the order of the mail and telephone modes, but that following up a mail survey with a telephone survey reduced nonresponse bias on the key driving history estimates. They also found that the two mode sequences yielded similar reductions in measurement error on the traffic accident question.

8.3.2 *Diagnosing using Parallel Surveys*

In a parallel survey approach to diagnosing differential selection and measurement errors, identical surveys are mounted at the same time, but in different modes of data collection with all of the components needed for a “real world” version of that survey. For example, a telephone survey with a sample drawn from an RDD telephone frame is implemented at the same time as a mail survey with a sample drawn from an address-based frame. The use of parallel surveys also undergirds some of the adjustment methods discussed below. That is, decisions about whether and how to address differential measurement errors requires having measurement of these errors in multiple modes at the same time.

One method of diagnosing the joint influences of selection and measurement errors involves (1) examining differences in estimates in one mode that contains both covered and noncovered units to diagnose coverage error, (2) examining difference in nonresponse-adjusted weighted and unweighted (or selection weighted) estimates to diagnose nonresponse error, and (3) examining differences in nonresponse-weighted estimates across modes to diagnose measurement errors. For example, Peytchev, Ridenhour, and Krotki (2010) examine potential differences across coverage, nonresponse, and measurement in the 2007 Health Information National Trends Survey (HINTS) which was conducted with simultaneous landline telephone

and address-based mail designs. To diagnose the effect of *coverage error* of the landline frame versus an address-based frame, Peytchev, et al. (2010) focused only on the mail survey respondents, examining differences in estimates for those with a landline telephone only (as reported in the mail survey) compared to the full sample estimate. To diagnose the effect of *nonresponse error* in both modes of data collection, the authors subset the mail data to only landline respondents (the telephone survey was collected only on landline respondents) and examine differences in an estimate that is weighted only for selection probabilities compared to the same estimate that is adjusted to account for nonresponse. Notably, the weighting adjustment was different across modes. The telephone survey included an adjustment for nonresponse and subsampling to the telephone screener, an interview nonresponse adjustment accounting for cross-classification of age, gender, and listed number status, and poststratification to demographic characteristic population totals (which are unspecified). In the mail survey, there was no screener; the completed questionnaire nonresponse adjustment was conducted separately for the whole household and nonresponse for members of the household, using census region, whether the household was in a high or low minority area, and the proportion of adults in the household who completed the survey; poststratification to demographic population totals was also done in this mode (again, unspecified). Finally, *measurement error* was examined by taking the difference between the poststratified mean for the telephone survey and the landline-only poststratified mean for the mail survey, assuming that the nonresponse adjustments and poststratification were effective in removing differential selection biases across the two modes. Peytchev, et al. (2010) found that the magnitude of each of the error sources differed across estimates, stating “for some estimates the ABS design may be preferable, but more so because of

less measurement error bias, while for other estimates the RDD design may be preferable due to lower nonresponse bias” (pp. 132-133).

One additional assumption that can be made when comparing parallel surveys in the absence of a gold standard is that one survey is the “benchmark” survey, sometimes called a “reference” survey (Klausch, Schouten, and Hox 2017). In a benchmark survey, one mode is considered to be “preferred” or “optimal” in terms of one error source, and thus the residual differences are due to a second error source. Survey practitioners select which mode they want to use for benchmarking, which may be challenging since each mode has advantages and disadvantages. For instance, some early evaluations assumed that the benchmark mode is the single-mode survey (e.g., interviewer-administered) and that the mixed-mode survey is the comparison survey. These studies assumed away any differential selection effects or changes in the measurement context between the ‘benchmark’ and other survey (e.g., Vannieuwenhuyze, Loosveldt, and Molenberghs 2010, 2012). Kolenikov and Kennedy (2014) explicitly note that the benchmark or reference mode may be the mode in which a survey has traditionally been administered (e.g., telephone), emphasizing the importance of mitigating differences in time trends when comparing a new mode to the old mode.

Others have used a benchmark approach to evaluate selection versus measurement errors. For example, Sakshaug, Cernat and Raghunathan (2019) estimate nonresponse bias by calculating the difference between weighted and unweighted estimates for estimates where record data are not available, similar to Peytchev, et al. (2010). In this study, for the mixed-mode context, the authors calculate weights by estimating response propensity models for respondents to the first mode (mail; telephone) and respondents to the full sequence of mixed modes (mail to telephone; telephone to mail), using the same set of administrative data as predictor variables in

these four models. To evaluate measurement error for items without a “true” value on the records, they take two approaches. First, they assign one mode as the benchmark mode – here, the nonresponse-adjusted estimate obtained from the mail respondents prior to the mode switch. Second, they take a directional measurement error hypothesis for sensitive items, assuming that the mode or mode combination in which the estimate for the reported sensitive item is the highest is the “benchmark,” thus allowing the “best” mode to vary across mode combinations. Here, four of the “benchmark” modes came from the nonresponse-adjusted estimates calculated on the mail survey prior to switching to telephone, three from the telephone survey prior to switching to the mail mode, two from the mail-telephone sequential mixed-mode design, and none from the telephone-mail combination. Notably, they reached different conclusions about measurement error under different assumptions for the “benchmark” mode.

In 2015, the Panel Study of Income Dynamics (PSID) experimented with adding an internet option to their existing telephone methodology (McGonagle, Freedman, Griffin, and Dascola 2017). One goal was to disentangle mode effects from true changes across panel waves, so they could evaluate data quality differences between modes and any potential breaks in series. To do this, they created a benchmark, or an expected baseline for expected change between waves, by comparing the consistency of reports between the 2013 and 2015 telephone surveys. The 2015 telephone survey only consisted of those respondents that were eligible to respond by web in 2016, thereby making the universes comparable when they next examined consistency between the 2015 telephone respondents and the 2016 web respondents. Without using this method, true changes that occurred between 2015 and 2016, the change in the reference period covered in the question, or mode effects were confounded. By having a benchmark for change, “true” mode effects could be measured.

The Hospital Consumer Assessments of Healthcare Providers and Systems (HCAHPS 2008) Survey publish mode-adjusted data using a reference survey as well. The survey is used to compare hospitals on different patient care-related outcomes. Different hospitals serve different types of patients and use different modes to administer the HCAHPS survey. Thus, comparisons across hospitals confound administration mode (both measurement and nonresponse errors), patient composition, and true differences. Using an experiment in which patients within hospitals were randomized to administration modes, a multistep regression-based method is used to adjust the estimates from the telephone, mixed mode (mail survey is followed up by telephone mode), and the IVR modes relative to the mail mode (considered the reference or benchmark mode) (HCAHPS 2008; Elliott, et al. 2009). A variety of demographic and other patient characteristics are used to account statistically for the compositional differences across hospitals when predicting the measurement outcomes of interest. The measurement-related mode adjustments focused on positivity bias (“top box” responses) in the interviewer-administered modes, using a multilevel linear regression framework to derive regression coefficients for telephone, mixed-mode, and IVR modes of administration compared to mail, after adjusting for compositional differences in patient characteristics. Using this regression-based method, Elliott et al. (2009) computed a statistic-specific mode adjustment for eight items of interest, and used this model to obtain predicted values for each hospital as if they had the same composition of patients and administered in the same mode. This adjustment factor thus reduced the positive ratings for hospitals that used interviewer-administered modes.

8.3.3 *Diagnosing using Reinterview Surveys*

Mixed-mode reinterview studies are an alternative method for diagnosing differential measurement error across modes (Klausch, Schouten, Buelens and Van Den Brakel 2017; Klausch, Schouten, and Hox 2017). This approach requires additional data collection efforts, and thus additional resources. For example, in a sequential mixed-mode survey, some respondents are interviewed in the first mode (e.g., web) and some respondents are interviewed in the second mode (e.g., mail). In a reinterview, all or a subset of the respondents to the first mode (e.g., web) are reinterviewed using questions from the initial survey in the second mode (e.g., mail). One of the modes used is designated as the “benchmark” mode, thus serving as a measure of the “true” value for the respondents who participated in both the initial mode and the follow-up reinterview mode. As such, there are multiple measures for each respondent. Klausch, Schouten, and Hox (2017) expand this idea to account for potential selection errors in one mode being optimal (e.g., face-to-face), but measurement errors in a second mode being optimal (e.g., web). Survey practitioners could consider whether the additional cost is justified based on the stakes for nonignorable differential measurement error.

These reinterview surveys can then be used to also adjust for measurement error. How exactly to do this adjustment is still under investigation. For example, Klausch, et al. (2017) examined six alternative estimators, finding variation in the quality of the estimates (root mean square error) across the different estimators and assumptions about which mode is the “benchmark” mode. When examining the Dutch Crime Victimization Survey, Klausch, Schouten and Hox (2017) use multiple imputation to impute potential responses in each of two modes (face-to-face and web) for each of two time periods (initial interview and reinterview).

8.4 Analytic Approaches to Diagnose and Adjust for Selection and/or Measurement Error

Surveys that transition from telephone to self-administered or mixed modes of data collection may want to adjust for differences in selection and measurement errors, in addition to diagnosing these differences. In situations where the diagnostic methods used for mode comparisons present no evidence against ignorable differential measurement error, the usual survey adjustment methods apply to the new mode(s) of data collection: 1) adjustment for unequal selection probabilities, 2) sample-based nonresponse adjustment, and 3) calibration approaches such as poststratification and raking (Valliant and Dever, 2018). Adjustments for nonignorable differential measurement error are more burdensome (Peytchev, Ridenhour and Krotki, 2010; Schouten et al., 2013). A critical assumption underlying adjustment methods for differential measurement error is that the respondent *could* have provided an answer in a different mode, but we only observe one of those modes in the data collection effort. This assumption draws on the causal inference literature, using the idea of a “potential outcome” (the plausibility of a response in a different mode that was not observed) to generate plausible counterfactual values for the responses.

The major analytic approaches to statistically account for differential measurement errors across modes are regression models, propensity score adjustments and imputation. Each of these methods relies on auxiliary or reference data, as discussed above. All of the adjustment methods assume that auxiliary or reference data proxy for factors influencing responses due to mode differences (e.g. age is a proxy for someone’s cognitive ability and by older age, as a result of declined memory capacity, recency effect could be larger in an auditory mode compared to a visual mode). In surveys with compositional differences in these proxies between two modes,

adjustments are needed to draw conclusions about the relative measurement differences across modes.

Many studies simultaneously examine multiple methods for adjustment, evaluating whether conclusions about mode differences vary with the adjustment method. Different methods often do not produce differences in conclusions, although some may be more suited for different problems than others. For example, Kolenikov and Kennedy (2014) examine three methods for assessing and adjusting for differential measurement error in a longitudinal survey with a face-to-face recruitment, with either telephone or web follow-up in wave 2: regression adjustment, multiple imputation using chained equations, and multiple imputation based on implied utilities. Suzer-Gurtekin, Valliant, Heeringa and de Leeuw (2018) examine three different methods to assess and adjust differential measurement errors in web-mail surveys: propensity score adjustment (Camillo and D'Attoma 2011; Lugtig et al. 2011), response mode calibration (Buelens and van den Brakel 2011, 2015), and multiple imputation (Suzer-Gurtekin 2013), concluding that there were no large differences in measurement error between web and mail surveys across these methods. Hox, de Leeuw and Klausch (2017) examine four methods: a multigroup latent variable model, inverse propensity score weighting, multiple imputation, and a reference survey approach for adjustment, also finding few differences. The context of transitioning from a telephone survey to a mixed-mode survey is different for every study, and as such, prior findings of no differences may not apply.

We discuss a limited subset of adjustment methods that may be easy for survey organizations to implement. We recommend that survey organizations explore and test multiple methods, including methods beyond what we discuss here. Despite the available methods for mode comparisons (and possible adjustments) based on the observational data, in an ideal

situation, planning for data collection should take the mode comparison analysis into account through parallel surveys or reinterviews, if at all possible. Of course, this extra design has a cost implication and should be considered in light of overall total survey error.

8.4.1 Response Mode Calibration

One method for adjusting for variations in the proportions of respondents who participate in each mode of a mixed-mode survey is to predetermine what proportion of the final estimate should come from each mode. If the observed proportion of responses from each mode differs from the predetermined proportion, then the sample is calibrated (essentially poststratified) to align with these proportions. This method assumes that the primary estimate of interest is one of change over time, and as such, adjusting for potential selection effects across modes is unnecessary. That is, the point-in-time estimate will still have biases related to each mode, but the bias in the change estimate will “cancel” out, under the assumption of constant measurement bias with each mode over time. For example, Suzer-Gurtekin, et al. (2018) use a mode calibration in three countries for the World Values Survey to fix the proportion of web and mail responses at 50% each. For two of the three countries, about 80% of the responses were obtained via mail; the third country aligned closely to this distribution. Here, fixed mode proportions are incorporated into the standard post-survey weighting adjustment scheme as an additional population control total. One could imagine also including a fixed proportion for participating in each mode as a control total in a raking adjustment.

Mode calibration assumes a constant mode specific measurement bias over time. Yet mode calibration can add bias to the total estimator if it does not completely correct for differential nonresponse. If the assumption of no selection effects is too strong to be upheld, covariates related to the selection effect can be incorporated into population adjustment methods,

such as the generalized regression estimator (GREG) (Beulens and Van den Brakel 2011; see appendix 7.A for technical details). Buelens and Van den Brakel (2011) propose two alternative ways to check for whether differential nonresponse has been taken into account. First, if the survey has variables which are not subject to mode effects and have both survey and population values (such as those from a population register), then the analyst can compare the adjusted estimates and population estimates to detect whether nonresponse adjustments completely correct for the differential nonresponse. The second approach, in the absence of population-level true values, is to use different calibration levels and compare whether the estimates change (e.g., 20% from web/80% from mail vs. 30% from web/70% from mail vs. 40% from web/60% from mail and so on). This adjustment method does not adjust for the bias in the total estimator, it calibrates the total measurement error to be equal to offset the difference in the difference estimator.

In a slightly different approach, the U.S. American Community Survey (ACS) is a sequential mixed-mode survey that incorporates mode specific estimation steps in its methodology, accounting for mode-specific differential selection probabilities. The survey invitation starts with a request to complete an online response option, and nonrespondents are asked to complete the questionnaire either online or through mail questionnaires. A probability subsample of the nonrespondents to the self-administered response mode invitation are followed up by in-person visits. The ACS estimation method combines the web, the mail, and the in-person responses by weighting their proportions to the overall sample. As a part of the weighting procedure, ACS weighting and estimation incorporates a Mode Bias Factor (US Census Bureau 2014). Because ACS nonrespondents are generally nonrespondents to the in-person follow-up, ACS assumes that remaining nonrespondents are more like in-person respondents than

respondents to the web and mail modes. However, adjusting only the weights for the households that participated via the in-person follow-up would increase the variation of the weights substantially. Thus, ACS uses the Mode Bias Factor to “spread out” the contribution of the nonresponse adjustments from the in-person mode data across the web and mail respondents and across cells defined by housing tenure, month of data collection, and marital status of the respondent. The same weight is applied and used throughout the estimation process for all analyses produced using ACS data, and is thus not outcome specific. In addition, this method does not include a differential measurement error adjustment. The Mode Bias Factor is only one part of a complicated multi-part weighting adjustment that includes multiple steps beyond selection probabilities to post-stratify and further calibrate the weights (US Census Bureau 2014).

8.4.2 Regression-based Methods

A common statistical method for accounting for multiple factors simultaneously that may be confounders when examining a key independent variable is regression, including OLS linear regression for continuous outcomes and logistic regression or a linear probability model for binary outcomes. In one of the simplest adjustment methods to implement, an important survey variable is predicted as the dependent variable and survey mode is the key independent variable. Because there may be differential selection of individuals into the modes, covariates are included in the model as potential confounders of the measurement effect of mode on the outcome. In a linear model for a continuous outcome (or linear probability model for a binary outcome), the estimated coefficient on the “mode” indicator (e.g., mail=1, telephone=0) is then used to obtain a predicted value under the model, subtracting off the relevant “mode effect” for the new mode (attributed to be measurement differences) (Elliott et al. 2009; Kolenikov and Kennedy 2014). In

a logistic model, the average predicted probability of observing the outcome is estimated, assuming that all individuals participated in the reference mode.

Kolenikov and Kennedy (2014) use this regression-based approach to adjust for potential differences in measurement error across telephone and web modes in wave 2 of the Portraits of American Life Study. Here, a sequential mixed-mode design is used – a large random subset was assigned to the mode, with nonrespondents followed up with a telephone interview, and a smaller random subset was assigned to start in telephone, with a web follow-up. In estimating mode effects, Kolenikov and Kennedy (2014) only focus on the initial telephone responses in the telephone-web condition. They report evidence of socially desirable reporting on a small subset of items in the telephone survey compared to the web survey, which the linear and logistic regressions account for, but that the standard errors of the adjusted estimates from the linear regression appear to be substantially underestimated compared to the multiple imputation using chained equations .

Jäckle, Roberts, and Lynn (2010) used ordinal logistic regression models to control for nonrandom selection of modes and evaluate measurement differences in answering ordinal attitudinal items in the European Social Survey. Although the data are obtained from randomized experiments in their application, the differences due to nonresponse in the socio-demographic characteristics of respondents are controlled analytically as covariates in the regression models. They define mode effects as the differences in the mean or predicted response distributions between modes after controlling for selected socio-demographic variables. Using two alternative regression models, the authors showed that varying ordinal logit model assumptions yielded different conclusions about mode effects. In particular, Jackle, Roberts, and Lynn (2010) examined how conclusions about mode-specific differences in response distributions differed

when the effect of mode on shifting from one category to the next (from strongly agree to agree, for example) was constrained to be the same for all of the categories of the outcome (the proportional odds assumption) versus allowing it to vary over some of the categories of the outcome (partial proportional odds). In follow-up research, Lynn, et al. (2011) use these models to test specific mode effect hypotheses based on social and cognitive theories. This line of work helps to improve both the mixed-mode design principles and set the modeling assumptions in adjustment models.

8.4.3 Propensity Score Adjustments

In the context of mixed-mode survey data analysis, propensity score methods rely on using a logistic regression model to predict the probability of being observed in a focal mode, conditional on a set of covariates, and then using this logistic regression model to obtain predicted probabilities of being observed in the focal mode for all respondents. This method does not assume an underlying measurement error, and the measurement-based mode effects are implicitly defined as the difference in the average systematic reporting errors between modes for a given matching group. These predicted probabilities can be used in a number of ways.

First, propensity scores are used for propensity stratification, in which a discrete number of groups with similar mode observation/mode choice probability scores are defined and contain individuals who participated in each of the alternative modes of data collection. Mode effects as defined by the differences in means within each of the propensity strata. For example, Camillo and D'Attoma (2011) use propensity score stratification to evaluate differences in measurement error across a survey of Italian university graduates conducted by web, with nonrespondents followed up by telephone. Predicted propensities to be observed in the web portion of the survey were divided into four propensity strata; mean differences were calculated, finding very few

differences between the telephone and web respondents using the propensity stratification method.

Second, propensity scores can be used in matching methods to find those respondents who participated in one mode (e.g., telephone) who are statistically similar on the distribution of the covariates that were used to create the propensity score as those in the second mode (e.g., web). For example, Suzer-Gurtekin, et al. (2018) use a propensity model with demographic variables (e.g., age, sex, education, race/ethnicity) and characteristics of the household (e.g., marital status, urbanicity, number of people in the household) to predict responding by mail versus web in the Finland, Norway, and Italy samples of the World Values Survey. They use the predicted propensity scores to find the “nearest neighbor” (a respondent in the other mode with a similar mode propensity score), finding no differences across the two modes for self-reported health.

Finally, inverse propensity scores can be used as a weighting factor, analogous to a sample-based nonresponse adjustment. For example, Fessler, Kasy, and Lindler (2018) use an inverse propensity weight to evaluate measurement differences in measures of income inequality between telephone and in-person interviews in the second wave of the Austrian European Union Statistics on Income and Living Conditions survey, where respondents were initially contacted by telephone and followed-up face-to-face. They fit a mode propensity model with covariates including the gender, employment status, age, number of people in the household, income, telephone characteristics, marital and cohabitation status, and education, and area level characteristics. They find that the in-person interviews have a wider income distribution than the telephone interviews.

8.4.4 Multiple Imputation

The use of multiple imputation treats responses that are not obtained in one mode, but could have been, as a missing data problem (Powers, Mishra, and Young 2005; Peytchev 2012). In particular, multiple imputation predicts the counterfactual outcome of the response that would have been obtained in a particular mode, but wasn't, conditioned on a set of covariates (assuming ignorable differential measurement errors) for each of the modes using (often) a series of regression models. In this approach, imputed values are obtained for each respondent for each of the modes in question. There are many approaches for multiply imputing data; however, sequential regression imputation or multiple imputation using chained equations are the most common imputation methods used in the mode effects literature. Kolenikov and Kennedy (2014), Suzer-Gurtekin, et al. (2018), Hox, de Leeuw and Klausch (2017), Klausch, et al. (2017), and Suzer-Gurtekin (2013) all use multiple imputation to account for measurement and selection effects in mixed-mode surveys. (See Kolenikov and Kennedy 2014 for an alternative multiple imputation method using utility functions in logistic regressions to multiple imputation using chained equations for adjusting mode effects.) The multiple imputation approach to diagnose and adjust for mode effects is frequently used in the mode effects research literature, but is less common in production studies, to our knowledge.

8.5 Summary and Takeaways

8.5.1 Mode effects diagnosis and adjustment methods are available, but the literature on applications of these methods to actual data and production surveys is scarce. The practicality of each approach and the validity of the assumptions of these methods needs to be evaluated empirically on multiple surveys in the future.

- 8.5.2 Survey analysts can make design decisions to either optimize estimates in terms of mean squared error (MSE), considering only the new mode, or can attempt to make the estimates most comparable to those from the previous mode.
- 8.5.3 Some approaches require mounting new data collections in two or more modes at the same time; others use statistical techniques to adjust away potential confounding order effects. All of the approaches rely on a set of assumptions for estimation and adjustment.
- 8.5.4 Some surveys may want to calibrate information collected via a new mode to information collected in the old mode, or combine these modes together. In this case, using methods to diagnose differential selection and measurement errors is needed. Other surveys may want to simply break a time series and attribute all of the differences in estimates to the package of mode features (both selection and measurement) that vary between the old and new data collection approaches.

Appendix 8.A: Technical Appendix on GREG estimation

Buelens and Van den Brakel (2011) extended the classical GREG estimator for the mean of some Y variable assuming a linear association between the variable of interest and a subset of covariates. We will refer to this model as Y-response model.

$$y_j = \mu(X_j, \beta^{(\mu)}) + \varepsilon_j \quad (1)$$

where:

$j = 1, 2, 3, \dots, N$ indexes individual population persons,

μ_j depends on X_j , a vector of covariates for person j ,

$\varepsilon_j \stackrel{iid}{\sim} (0, \sigma^2)$.

Following the usual calibration notation, the corresponding GREG estimator for the total of Y , T , is

$$\hat{T}_{yr} = \sum_{j \in S} w_j y_j, \text{ where:} \quad (2)$$

$$w_j = \frac{1}{\pi_j} \left(1 + \frac{x_j^T}{\text{var}_M(y_j)} \left(\sum_S \left(\frac{x_j x_j'}{\text{var}_M(y_j) \pi_j} \right)^{-1} (T_X - \hat{T}_{X\pi}) \right) \right), \text{ where:}$$

π_j : the probability that unit j is included in sample S , $T_X = \sum_{j \in U} x_j$, $\hat{T}_{X\pi}$: the π estimator of T_X .

When the population size, N , is known the GREG estimator for the mean is

$$\bar{Y}_r = \frac{\hat{T}_{yr}}{N} \quad (3)$$

When mode specific systematic reporting error is introduced into the Y-response model:

$$y_j = \mu(X_j, \beta^{(\mu)}) + R_{Mj} B_M + \varepsilon_j, \text{ where:}$$

$j = 1, 2, 3, \dots, N$ indexes individual population persons,

$m = 1, 2, 3, \dots, M$ denotes survey response mode,

μ_j depends on X_j , a vector of covariates for person j ,

B_M = reporting error for person j who responds by mode m ,

$$R_{mj} = \begin{cases} 1 & \text{if population unit } j \text{ responds in } m \text{ mode} \\ 0 & \text{if otherwise} \end{cases}$$

$$\varepsilon_j \underset{\sim}{iid} (0, \sigma^2).$$

The classical GREG estimator for a total can be extended to:

$$\hat{T}_{yr} = \sum_{j \in S} w_j (\mu(X_j, \beta^{(\mu)}) + R_{Mj} B_M + \varepsilon_j)$$

Then the expectation of \hat{T}_{yr} with respect to the sampling, E_S , and Y-response model, E_M , is:

$$E_S E_M (\hat{T}_{yr}) = T_Y + \sum_M B_M E_S [\hat{T}_M] \quad (4)$$

As the expectation with respect to sampling and Y-response model shows in equation (5) unless $B_{Mj} = 0$, \hat{T}_{yr} is not an unbiased estimator. Buelens and Van den Brakel (2011) focuses on the estimation of change over time. The other component in the bias is $E_S [\hat{T}_M]$ and due to nonrandom assignment of modes and possible design variations, $E_S [\hat{T}_M]$ is not expected to be a constant over time, although B_M is defined as a constant and that could be a plausible assumption in some cases. Therefore the real change over time is confounded with the differences in total measurement error which is formulated by $\sum_M B_M E_S [\hat{T}_M]$. When we consider two time periods, time 1 and time 2, and the change between the time 1 and time 2, the expectation of the change will follow:

$$E_S E_M (\hat{T}_{yr}^1 - \hat{T}_{yr}^2) = T_Y^1 - T_Y^2 + \sum_M B_M (E_S [\hat{T}_M^1] - E_S [\hat{T}_M^2]) \quad (6)$$

If $E_S [\hat{T}_M^1] - E_S [\hat{T}_M^2]$ is equal to zero, the change estimator $\hat{T}_{yr}^1 - \hat{T}_{yr}^2$ will be unbiased. Buelens and Van den Brakel (2011) substitute \hat{T}_M^1 and \hat{T}_M^2 with constants, Y_M , to offset the effect of mode effects in the change estimator. When the time superscript is ignored:

$$\hat{T}_M = \sum_S w_j R_{Mj} = Y_M \quad (7)$$

Y_M are chosen arbitrarily and treated as population controls. Setting the \hat{T}_M^1 and \hat{T}_M^2 as constants yields an unbiased change estimator. Alternatively, response propensities can be used to estimate the population mode response proportions. The condition of (8) is achieved by including the response mode indicator in the GREG weighting model and mode calibrated GREG estimator is:

$$\hat{T}_{yr}^* = \sum_{j \in S} w_j^M y_j.$$

9 Costs

Survey costs are one of the primary drivers for surveys to transition from telephone to self-administered or mixed-mode surveys. Out of the 19 respondents to our survey of organizations that transitioned from telephone to self-administered or mixed mode, 9 (47%) reported that costs of the interviewer-administered survey were an extremely or very important factor in deciding to transition the survey from telephone to self-administered or mixed-mode and 8 of 20 (40%) reported that the costs of the self-administered or mixed-mode survey were extremely or very important.

In this section, we discuss the types of costs that may be incurred during a redesign effort. Where possible, we provide data about costs. Many surveys are currently pursuing or examining the possibility of a transition to a new mode, so current cost data is somewhat sparse. We also provide examples of how to model costs for different combinations of modes in order to identify when it may be cost effective to combine mail and web modes

9.1 Factors that Might Contribute to Changing Costs

As discussed in the preceding chapters, **surveys that transition from telephone to self-administered or mixed modes will need to consider a variety of factors that may affect costs when redesigning each step of the data collection process. This may include redesign-specific costs for sample design, sample management, contact materials, data entry, and other costs**, as listed in Table 9.1.

Table 9.1. Design features that may affect costs when transitioning from telephone to self-administered or mixed-mode surveys

New or different sample frame
New or different sample design
Identifying potential experiments needed to inform redesign
Developing materials for potential experiments needed to inform redesign
Questionnaire design and development, including for multiple modes
Pretesting new or revised questions, including for multiple modes
Developing or purchasing questionnaire administration software, including for multiple modes
Developing or purchasing sample management software, including for multiple modes
Developing new or revised protocols for sample release and/or field management, including decision rules for release to multiple modes
Use, administration, and management of incentives to be used in redesign, including sequential use at different mailings
Developing contact materials, including letters, brochures, emails, webpages, and other materials, including for multiple modes
Developing recruitment protocols, including sequencing of modes (if applicable) or responsive or adaptive design features of data collection
Developing within-household selection methods, including one- or two-stage selection of sampled eligible respondents
Developing field management and reporting protocols
Developing revised materials for review by Institutional Review Board
Developing and administering new or revised field monitoring reports, including number of completes and response rates by modes
Printing new recruitment letters, brochures, and other enclosures
Labor costs for formatting and layout of mail questionnaires, as needed
Printing mail questionnaires, as needed
Labor costs for programming web survey software, as needed
Evaluation of consistency of questionnaire and recruitment materials across modes, as needed
Postage to send new contact and other materials
Postage for returned mail surveys, as needed
Data entry for returned mail surveys, as needed
Labor costs for editing survey responses and dealing with missing data
Labor costs for other post-survey processing, including weighting and imputation
Labor costs for creating analytic data set, including disclosure review procedures
Costs of maintaining help line about new and old design
Labor costs for analyzing data from new and old design
Labor costs for reporting about new and old design
Costs for hosting materials about redesign on dissemination forum

Reports of costs for the telephone-based designs versus the new designs may include only some of these individual survey components, along with the other costs associated with telephone surveys, such as the costs of training and monitoring interviewers. When examining cost differentials across modes, then, it is important to understand what cost components are being reported on and what infrastructure is assumed to be in place versus those included in the cost for the data collection (Olson, Wagner, and Anderson 2018). It is also important to understand whether the study is reporting on only variable costs, only fixed costs, or both fixed and variable costs. For instance, studies that assert that web surveys have no costs per completed interview overlook the cost of the web survey software itself, programming and testing the instrument, possible use of incentives, costs for purchasing the sample, costs for developing recruitment materials, costs for recruitment itself, and so on.

Overall, actual data on costs for surveys before and after transitioning from telephone to self-administered or mixed modes are sparse, although cost data from mode comparisons are more common, especially in the epidemiology literature. Costs are generally reported for isolated components rather than across the entire study. Furthermore, costs for a redesign may be folded into the costs for the production survey itself, and thus difficult to disentangle the cost for the redesign from the cost for the survey. This is especially likely to be true for surveys conducted by smaller survey organizations. Finally, organizations that do not conduct a bridge survey may only have costs of the survey administration from previous years, and not the concurrent cost of the existing protocol compared to the new protocol.

9.2 Differential Costs Between Modes

9.2.1 Costs of New Mode versus Old Mode

Little data is available that compares costs directly for a previously administered telephone survey to a newly administered self-administered or mixed-mode survey. According to our survey, 13 out of 23 organizations reported that the surveys were redesigned in an attempt to reduce total survey costs, but 6 organizations reported that costs were not part of the decision process. Thirteen of the survey respondents indicated that both the total survey costs and cost per completed interview were reduced compared to the interviewer-administered mode.

Additionally, some organizations in our data collection effort are currently testing methods for transitioning from interviewer-administered to other modes, so cost data are not readily available.

The little data that we do have available from existing published studies suggests that the goal of reduced costs in the new self-administered or mixed-mode design is achieved, although it depends on how costs are measured. In perhaps the largest example we have available, in 2012, the American National Election Studies (ANES) conducted simultaneous face-to-face interviews and web surveys. The costs per complete in the face-to-face mode were over \$2000 per case; with reduced precision due to a clustered design, the effective sample size is reduced, raising costs per effective complete to \$2800 (Jackman 2015). The 2012 ANES web survey was conducted with the Gfk panel, at a total cost of \$240 per complete overall, or \$460 per effective complete. In 2016, a general population sample was used to recruit the internet component of the ANES (Debell, et al. 2017). Mailing and incentive costs alone for the letters to recruit individuals to the web survey, excluding the cost of printing the letters, programming the web instrument, or any other labor or design costs, for a pilot test for the internet version of the 2016 ANES ranged

between \$146 per complete to \$201 per complete, depending on whether the incentive structure was front-loaded or escalating, and how the letters were addressed.

In a smaller scale study, Lien (2015) compared costs for a tobacco survey in Oklahoma that had traditionally been conducted by telephone to a new design that started with a web survey and a telephone follow-up. The two surveys were fielded simultaneously. The total fixed costs for infrastructure, set-up, survey monitoring, and back-end processing across eight months of experimental survey administrations of both mode combinations were lower for the phone-only mode, at \$39,250 compared to \$44,250 for the web with phone follow-up (web+phone) design. It is not clear whether the differences in set-up were because the telephone survey had been conducted in the past, and thus had amortized costs across years. Yet the costs per sampled case and the costs per complete were lower for the web+phone mode, largely due to reduced interviewer time. In particular, the costs per sampled case dropped from \$6.25/sampled case for the phone-only mode to \$5.66/sampled case for the web+phone mode due to a decrease in the calling time for the interviewer. The costs per complete also fell from \$20.03/complete in the phone-only mode to \$17.65/complete in the web+phone mode.

Link, et al. (2008) compared costs for the traditional telephone-administered Behavioral Risk Factor Surveillance System (BRFSS) to a mail version of this survey. They found that the mail survey reduced costs by about 12%, from \$79,578 per 1000 completes for the telephone survey to \$70,969 per 1000 completes for the mail survey. Again, the main driver of the difference was the reduced interviewer time, which offset the increased costs for printing and other mail materials.

Other examples illustrating the benefits of using multiple modes in a survey are found in the literature. For example, using mail recruitment for a telephone interview can reduce costs.

Using time spent on calling as a cost metric, Allison, Stevenson, and Kniss (2014) used an address-based sample to select the sample for the telephone-administered 2012 Wisconsin Family Health Survey, traditionally conducted via a telephone survey with a random digit dial sample. Short mail surveys were sent to selected households without matched telephone numbers and to nonresponding households that did have matched telephone numbers. This redesign using mail surveys for the initial contact attempts reduced the number of call attempts to yield a completed case of the 2011 all-telephone data collection from 26.8 to 11.3, and reduced the hours of calling from 3.0 hours per interview to 1.5 hours per interview. The response rate was similar – 46.7% for the 2011 telephone-only iteration and 44.8% for the mail and telephone iteration in 2012. In another example, Redman, Thompson, Yost and Everts (2017) reported a decline in the number of hours per interview from 0.912 to 0.496 hours per interview when they moved the Franklin and Marshall College Poll of registered voters to web. In this study, respondents could select to participate either by web or by phone, rather than interviewing all by telephone.

9.2.2 Costs of Different Design Features in New Mode(s)

When surveys transition from one mode to another, there are many new design decisions that have to be made. This opens the opportunity for experimentation on aspects of design features in the new mode.

As discussed in the nonresponse chapter (Chapter 6), one of the most commonly tested design features in surveys that transition to self-administered or mixed modes is incentives. Generally, some level of incentive reduces the cost per complete relative to no incentive because of reducing the number of follow-up attempts that are needed. However, not all levels of incentives are equally effective at controlling costs. For instance, in a mail survey of anglers,

Andrews, Brick, and Mathiowetz (2014) examined the effects of offering no incentive, a \$1 cash incentive, a \$2 cash incentive, and a \$5 cash incentive on the direct costs of printing, postage, and incentives. They found that, relative to no incentive, the cost per complete decreased by about 20% for the conditions that were offered \$1 and \$2, but increased about 15% for the condition that was offered \$5. They concluded that the \$2 incentive was the most cost-effective option for this study.

Incentives are particularly common in surveys with two-stages of selection – the responding households are first asked to complete a screener questionnaire, and then selected persons are asked to complete a topical questionnaire. Incentive experiments were conducted for the 2016 National Study of Children’s Health (NSCH), a 2011 field test for the National Household Education Survey (NHES), and a national pilot study for the Residential Energy Consumption Survey (RECS). Each of these studies concluded that modest incentives were cost-effective.

In particular, the 2016 NSCH compared a \$2 or \$5 incentive for completing the screener survey to no incentive at all (US Census Bureau 2018). The overall mailing cost per household for the no incentive group was \$7.85, compared to \$9.82 for the \$2 incentive and \$12.57 for the \$5 incentive group. The mailing cost per completed screener also varied across incentive groups - \$22.17 for the no incentive group compared to \$25.47 for the \$2 group and \$30.97 for the \$5 group. Despite these increased costs, the \$2 initial screener incentive reduced the number of follow-up attempts that were needed, and thus reduced the cumulative cost to yield a completed topical survey from \$134.88 to \$133.02.

In a 2011 field test for the NHES, Han, Montaquila and Brick (2013) report that using a \$5 incentive for a topical questionnaire was cost-effective. In particular, the increased data

collection cost for a \$5 incentive was offset by the increased topical response rate, such that the cost per completed topical survey was the same for a topical incentive level of \$5 as it was for that of no incentive. Data collection costs per complete increased for other levels of incentives (\$10, \$15, and \$20) for the topical questionnaire; costs per complete were also increased about 20% when using a \$5 screener incentive compared to a \$2 screener incentive. Jackson, et al. (2018) updated this experiment in the 2016 NHES, using an incentive tailored to different groups as defined by a predicted response propensity. The incentive-only cost per completed case for the tailored response propensity (where the tailoring ranged from \$0 for the most likely to participate to \$10 for the least likely to participate) was \$7.80, compared to \$3.38 for a \$2 incentive and \$8.07 for a \$5 incentive. Although the incentive (only) cost per complete was lower for the tailored incentive, more contact attempts were needed to obtain a similar response rate.

In a national pilot study for the RECS, Biemer, et al. (2018) report total survey costs, including “all relevant fixed and variable costs of materials, labor, printing, incentives, postage for the mailing the questionnaires, and follow-up contacts, as well as the costs of keying the paper electronic questionnaires” (pp. 14-15), across four different mode conditions and two promised incentive conditions. Relative to the average across all conditions, a web-only survey had the lowest cost per completed interview, as did the sequential mixed-mode option that had a web survey with paper questionnaires sent in the second mailing for follow-up. Offering respondents a choice to participate in either paper or web increased costs relative to the average, but costs decreased when the choice was also tied to an additional incentive to complete the survey by web. The \$20 promised incentive increased costs per complete. They note that mixed-mode conditions that involve more paper questionnaires have more data entry costs, but those with more web surveys have more mailing and follow-up costs.

Other design options have also been examined, including the ordering of the mixed set of modes. In an establishment survey experiment for the Deaths in Custody Reporting Program, Ellis, et al. (2013) compared the standard mode offerings of web and paper survey forms, fax, email, bulk data files, and telephone to a web-mail option, where the paper forms were not sent to respondents in the first mailing, but were sent in the second mailing. The effects on costs were substantial – total survey costs for the group including the paper questionnaire was \$32,099, compared to \$22,883 for the group where the paper questionnaire was withheld until the second mailing. Unsurprisingly, most of the cost savings were incurred by reducing initial mailing costs and other costs related to processing the paper survey, yielding a savings of about \$9200, or about \$3.47 per case. Patrick, et al. (2018) compared costs for a mail survey where web was offered as an option to nonrespondents to a web-mail survey where the paper questionnaire was sent to nonrespondents, and a web-mail survey where emails were also sent to nonrespondents. Although total costs were lower in the two sequential designs that started with web, costs per complete were 8% higher in the web-mail survey than in the mail to web survey, but were 5% lower when emails were also sent to nonrespondents.

Some cost evaluations include simulating what costs would be if a different protocol had been used during data collection. For instance, in a 2014 feasibility study for the redesigned NHES, McPhee and Masterson (2015) examine the postage costs alone of alternative mailing strategies to households that are returned as “undeliverable as addressed” by the US Postal Service. The full cost for the third questionnaire mailing, conducted using FedEx, was \$582,492. They then simulated what the cost impact would be by excluding any sample unit that had multiple undeliverable returns, finding an estimated total costs of \$510,223, or a 12.4% cost savings, and that were undeliverable return plus had a vacancy indicator on the frame, with a cost

savings of 3.5%. Of course, this cost savings would come with a decrease in response rates and potential decreased representativeness of the survey data.

9.3 Costs per Complete Versus Sample Size

Studies that are transitioning from telephone to self-administered or mixed modes may need to evaluate whether a mixed-mode survey is cost-effective over a single mode. Although this is going to depend highly on the target population, sample frame, sample size, complexity of the survey questionnaire, language needs, infrastructure availability, and myriad other factors, survey organizations may be able to build a model for predicting the sample size for cost effectiveness of a mixed-mode survey over a single mode survey.

Previous studies have done this, with mixed conclusions about the sample size “tipping point” depending on the mix of modes and assumed cost structure. In a simulation of total costs, Lien (2015) estimates total costs for a phone-only survey compared to a web+phone survey by the number of completed interviews using the formula:

$$\text{Cost} = (\# \text{ completes} * \text{cost_per_complete}) + \left(\frac{\# \text{ completes}}{\text{response_rate}} * \text{cost_sampled_unit} \right) + \text{Fixed_costs}$$

Lien’s calculations reveal that total costs are less for phone-only surveys for smaller sample sizes, are about the same for surveys with about n=1000 completes, and that web+phone surveys are more cost effective when the sample size is more than 1,100 completes. Similarly, Asch (in Fricker and Schonlau 2002) reports that the extra programming and management effort of adding a web survey can be realized for surveys with at least 620 Web completes. Griffis, Goldsby, and Cooper (2003) conclude that a web survey is cost effective for surveys with over 1538 sample units compared to a mail survey.

Lesser, et al. (2017) simulated costs for studies of different sample sizes for a web-only, web with mail follow-up and four total mailings, and web with mail follow-up and five total mailings. The costs include both fixed and variable costs, where variable costs include printing/postage per unit, and the data entry costs per unit, but the questionnaire is fixed at a 12 page questionnaire, and constant response rates and proportions of responses by web are assumed for different sample sizes. In this simulation, the costs per complete for a mail survey are lower for sample sizes under about 2000 sampled units; over 2000 units, the two web-mail designs are between \$2 and \$7 less per completed case for samples ranging from 2000 to 10000 units than the mail-only designs.

These examples indicate that individual surveys that are transitioning from telephone to self-administered or mixed-mode surveys need to consider the different fixed and variable costs of modes. For instance, for some studies, it will be more cost effective to data enter mail questionnaires than to program a web survey. There may be costs that have already been sunk into developing a complicated telephone instrument that are more easily ported into a web survey than the cost for developing an easily-administered mail survey. Similarly, some survey organizations may not have the capacity for data entry, and thus need to outsource that cost to a different organization. In addition, for repeated surveys, there will be additional cost savings in web surveys given the instrument was previously programmed and thus programming time is reduced. Each of these issues must be considered when examining survey costs related to transitioning from one mode to another.

9.4 Costs for Bridge Surveys

Parallel surveys conducted in the original mode, sometimes called “bridge surveys,” allow organizations to evaluate the impact of the change in mode on survey estimates by conducting a survey in both modes simultaneously (See Chapter 7). **Bridge or parallel surveys necessarily increase the costs for the overall study because two surveys are being conducted, rather than one.** The bridge survey may be conducted in addition to any other pilot work done to determine the need for a bridge survey at all. For example, according to the Transition Plan for the Fishing Effort Survey (Marine Recreational Information Program 2015), the cost for the telephone-based Coastal Household Telephone Survey is \$1.8 million per year and the mail-based Fishing Effort Survey is estimated to cost roughly \$500,000 less than this per year, or \$1.3 million per year, at least during the bridge survey time. The National Household Travel Survey did not conduct a bridge survey for its 2017 administration, following recommendations from an expert panel that asserted the survey landscape had changed so dramatically from the previous administration in 2009 that it would not be useful (Transportation Research Board, 2016; p. 26).

Studies that do not have a formal budget to conduct a bridge or parallel study must be more creative in evaluating impact of the design on costs, especially if understanding the costs for individual components of data collection. Here, organizations should identify the individual components that are part of the study in each mode and articulate – to the best of their abilities – the costs for each component. That way, exactly where costs increased or costs decreased can be more directly evaluated.

9.5 Timeline as Costs

Transitioning from telephone to self-administered or mixed-mode surveys requires time. Some surveys may have budget and time for pretests, field tests, and evaluation of different protocols. For instance, the Panel Study of Income Dynamics (PSID) transitioned the specifications for telephone-based Core PSID instrument to a web instrument over six months, then spent six months programming and testing the web instrument, including small pilot tests with convenience samples from online panels. Additional programming to account for the longitudinal nature of the PSID took six additional months, plus two field tests with cooperative PSID respondents over one month each (McGonagle, Freedman, Griffin, and Dascola 2017). The NHES transitioned from a telephone survey to a mail survey, and included extensive pilot, cognitive, and field testing, resulting in a five year gap without national NHES survey estimates (NHES 2015, p. 3). The National Household Travel Survey (NHTS) started planning a transition from random digit dial telephone administration to an address-based sample self-administered and phone design in 2011, convening expert panels to advise on the redesign,

Other studies may conduct a pilot study and realize that the estimates from the new mode are so disparate from the previous mode that more work is needed. In particular, some surveys may find that research is needed to collect data to develop a calibration model in order to link the data across time between the different data collection modes. Here, surveys in different modes are fielded simultaneously for multiple years, and estimates are generated each year to adjust for the impact of mode differences. For instance, NOAA Fisheries decided to transition their Coastal Household Telephone Survey to a mail-based Fishing Effort Study, and encountered survey estimates substantially larger than the previous telephone-based study. This

motivated NOAA Fisheries to develop a study transition plan over three to four years, permitting additional evaluation and calibration of survey estimates (NOAA Fisheries MRIP 2015).

Additionally, **the transition from telephone to self-administered or mixed-mode surveys may add to the timeline of the study in the field or reduce it.** For instance, Allison, Stevenson, and Kniss (2014) shortened data collection from 11 months to 5 months by switching the frame to an address-based frame and a mixed-mode mail and phone design compared to the previous telephone-only design. The PSID reported that scaling up the web survey to all of the prior PSID cooperative respondents would save about 26,000 calls across the field period, and that these cooperative respondents on average cooperated within 1 to 3 three days of the initial request (McGonagle, Freedman, Griffin, and Dascola 2017).

9.6 Summary and Takeaways

- 9.6.1 Surveys that transition from telephone to self-administered or mixed modes will need to consider a variety of factors, include redesign-specific costs for sample design, sample management, contact materials, data entry, and other costs, that may affect costs when redesigning each step of the data collection process.
- 9.6.2 Little data on survey costs for surveys that transition is available from existing published studies, but these data suggest that the goal of reduced costs in the new self-administered or mixed-mode design is achieved. Inferences about costs, however, depend on how costs are measured.
- 9.6.3 Incentives can reduce the cost per complete relative to no incentive by reducing the number of follow-up attempts that are needed. However, not all levels of incentives are equally effective at controlling costs.

- 9.6.4 Previous studies have simulated costs for mixed-mode surveys with different sample sizes, coming to mixed conclusions about the “tipping point” depending on the mix of modes and assumed cost structure.
- 9.6.5 Bridge or parallel surveys necessarily increase the costs for the overall study because two surveys are being conducted, rather than one.
- 9.6.6 Transitioning from telephone to self-administered or mixed-mode surveys requires time. Some surveys may have budget and time for pretests, field tests, and evaluation of different protocols. Others may not, or may have observed such a time gap between the last iteration and the current iteration that a bridge study is not deemed necessary.

10 Human Subjects Issues

Regardless of the mode of data collection, a human subjects protections program should be guided by the ethical principles regarding research participants as presented in the report by the National Commission for the Protection of Human Subjects in Biomedical and Behavioral Research (1979), universally recognized as *The Belmont Report*. These principles are the following:

1. Respect for persons;
2. Beneficence; and
3. Justice.

The Department of Health and Human Services (HHS), in the forefront of funding studies on sensitive topics, revised and expanded its regulations for the protection of human subjects (45 C.F.R. § 46) based on the Belmont Report and other work of the National Commission. Their regulations are designed to offer basic protections to human subjects involved in both biomedical and behavioral research conducted or supported by HHS. The Office for Human Research Protections (OHRP) is the agency within HHS responsible for enforcement of these regulations (45 C.F.R. § 46.103(b)(4)).

While the sensitivity, invasiveness, and difficulty of studies varies across Federal departments and agencies, they follow regulations similar to HHS when conducting research with human subjects. At a minimum of every three years, every federal information collection of the public undergoes a formal review by the Office of Management and Budget's Office of Information and Regulatory Affairs (OMB/OIRA). The review includes publishing federal register notices for public comment in which agencies provide lengthy, detailed "supporting

statements” about their information needs, planned methods, and any mandates to justify the public reporting burden. Agencies also declare under which statutes and provisions the privacy and confidentiality of participants’ answers and identity will be assured.

A little over half of the studies in our convenience sample of survey organizations reported on involved informed consent (13 of 23 answering). Among these, about half said that the process by which they obtained informed consent changed with the transition. Descriptions of how consent changed tended to note that consent was now indicated through responses to self-administered questions, rather than through an interviewer. One respondent noted that this process “may be challenging for youth assent and literacy issues.” When considering a transition of modes, there are three main considerations to take into account regarding protection of human subjects: obtaining informed consent, protection of personally identifiable information (PII), and handling respondent distress.

10.1 Obtaining Informed Consent

Regardless of the mode of data collection, human subjects need to understand what they are being asked to do. It is important to consider the implications of a mode transition on the informed consent process. Informed consent is guided by the principle that researchers should not conduct research on human subjects unless they have obtained informed consent from the subject or the subject’s proxy. Informed consent is the process by which the prospective research participant is provided with sufficient opportunity and information concerning the conduct, purpose, use, risks, and benefits of the study to consider whether or not to participate in the study. Informed consent requires that the consent to participate in research is relevant to the research being conducted (that is, informative), understood by the potential participant, and

voluntary. Informed consent must be conducted in a manner that minimizes coercion or undue influence. These ethical principles are in the Belmont Report and have been codified in the Federal regulations to protect human research subjects, found at 45 C.F.R. § 46.116.

In interviewer-administered modes such as telephone or field surveys, the informed consent process is administered verbally. Respondents can ask questions, negotiate meaning, and confirm their understanding of the benefits and risks of the survey. When transitioning to a self-administered mode such as a web or paper survey, a few considerations arise. First, **the respondent must be sufficiently literate to read and understand the consent form**, although in a web survey, there is the possibility to have an audio feature read the consent language aloud. Second, **whereas in interviewer-administered surveys, interviewers are trained and instructed to read every word and sentence of the informed consent language aloud, there is no single mechanism or guarantee to ensure that respondents in self-administered environment do the same.** Respondent attentiveness to informed consent language may be akin to ‘speeding’ in web surveys or the blind faith often assigned to user agreements in phone apps. That is, a respondent may click or speed through the informed consent segment to get to the substance of the survey. There is also no clear way to check for understanding of the informed consent document in self-administered surveys, although in interviewer-administered surveys, interviewers can evaluate both verbal and nonverbal cues from the respondent (Walther 2002).

Efforts to ensure a respondent’s consent is informed, understood, and voluntary may involve different approaches by mode. For instance, surveys that transition to self-administered modes may find it helpful to embed some comprehension questions to confirm the respondent understands what is being asked of them, only allowing progress when they have answered the questions correctly, although we could find no examples of surveys that have done this. It may

also be important to repeat the most critical information from the informed consent in other communication materials, such as the cover letter or an informational brochure. Good user-centered design for the presentation of the information -- for example in paragraph format vs. in bullet points, using mouse hover-overs to present FAQs or definitions, or presenting information on a single web screen vs. spread out across several screens -- may improve the level of attention paid to the information.

While there is some literature on the comprehension of informed consent language within various modes, less is known about the extent to which mode and design may influence comprehension across modes. **Surveys that transition from telephone to self-administered modes would benefit from conducting experiments that vary the display, attention, or comprehension attributes of informed consent information between interviewer-administered to self-administered surveys.** Potential web-based experiments might study respondent comprehension and attention to consent language by measuring time spent on the screen and via eye-tracking, behavior associated with key focal points. Experiments may study the following:

1. Use of paragraph format vs. bullet point format (speed)
2. Extensive language vs. shorter language (reading burden)
3. Use of more legal-ese versus “plain language” (tone)
4. Use of design features like bolding and underlining versus not using those features (use of focal points)
5. Information on one screen versus on multiple screens (segment complexity)

6. Use of “hover” words for concepts that may not be readily understood, such as “privacy” or “confidential” (and measuring the number of respondents who engage the hover word)
7. Asking respondent to summarize what the key points were on the screen before continuing or conducting a quiz to ensure comprehension. (engagement and comprehension)

10.2 Protection of Personally Identifiable Information

Institutional Review Boards are bound by the Common Rule regulations to determine that adequate provisions have been made to protect the privacy of human subjects and to maintain confidentiality of their data. Privacy refers to a person’s ability to control what information that others know about him- or herself. Confidentiality pertains to the handling, storage, collection and use of an individual’s personal information.

Personally identifiable information (PII) is information that can be used on its own or with other information to identify, contact, or locate an individual. The protection of PII begins at the sampling phase of a project: respondents’ names and other identifying information (e.g., social security numbers) may be removed and replaced with unique project identifiers. During data collection, respondents may be told that they should not state any names or the names of any geographically identifiable locations in their responses. The informed consent language also typically describes the level of confidentiality that respondents can expect, with most surveys indicating that "your responses will be combined with others’ and all data will be reported in aggregate." In federal surveys, it is customary to quote the legal basis for these protections as well as penalties to the researcher should a willful breach occur. At the analysis stage, if data are

reported in tables (e.g., by respondent characteristics), there are often data suppression rules for small cell sizes.

In an interviewer-administered environment, a respondent can verbalize PII at any moment during the survey by simply stating something aloud to the interviewer. The interviewer can remind the respondent not to share this information, and if the interview is being recorded, the PII can be redacted from any transcript that is created. In the data cleaning phase, researchers can review open-ended data and redact any PII.

In transitioning a survey to a self-administered environment, the nature and opportunities to reveal PII during the survey will vary. On a paper questionnaire, because the respondent may write PII anywhere on the survey, data processors should be prepared to look for and de-identify or redact this information. In a web-based environment, the respondent may only have the opportunity share PII in open-ended text boxes and may share that information regardless of whether it pertains to the question being asked. Again, the data processing team should be prepared to review all open-ended responses and de-identify or redact PII before finalizing the data set. Additional potential PII may be collected in web surveys, including IP addresses, increasing the types of PII to protect. Informed consent procedures may also benefit from including a statement that respondents should not reveal any PII when asked to type in responses on the survey.

10.3 Mandatory Reporting of Respondent Abuse or Harm to Self or Others

Another critical consideration with PII is mandatory reporting. Mandatory reporting may be required on a study if a respondent reveals that they are being abused or are at risk of harming themselves or others (AAPOR n.d.). In an interviewer-administered context, if the respondent

expresses at any point during the interview that they are being abused or are a danger to themselves or to others, in order to help protect their safety and the safety of others, the respondent is told during the informed consent process that the interviewer would need to report this to the appropriate healthcare or law enforcement agencies.

When transitioning to a self-administered environment, factors that influence mandatory reporting change. One important factor is the timeliness within which mandatory reporting can occur. It may be that a respondent reveals in an open-ended response that they are being abused or are a danger to themselves or to others. In a paper questionnaire environment, it may be days or weeks before the questionnaire is received, let alone processed and reviewed. And while a web-based questionnaire is received instantly upon completion, it also may not be processed or reviewed for many days or even weeks, depending on the length of the field period. Researchers working on sensitive topic surveys should carefully plan procedures for adhering to mandatory-reporting requirements. Plans should also consider the feasibility of reporting in the self-administered environment. For example, a web-based survey could engage an alert system in which particular words such as “kill,” “abuse,” “hurt,” “danger,” “scared,” etc. trigger an immediate review of the case to determine if mandatory reporting is needed. As with interviewer-administered modes, respondents would need to be informed in the informed consent process of any procedures that would require mandatory reporting.

10.4 Handling Respondent Distress

During research with especially sensitive content, including unwanted sexual contact, victimization, or domestic violence, respondents may become upset by the interview. Common signs of distress including crying, changes in mood, changes in facial expressions, shaking,

trembling, getting off task, or losing the ability to focus on the interview. In an interviewer-administered environment, interviewers are usually trained to recognize symptoms of low, moderate, or extreme distress. They adhere to protocols to respond to varying levels of distress, check in with the respondent, and determine if the interview should be stopped and whether a counseling resource should be engaged. These surveys may conclude with either debriefing questions about the interview or a distress check-in in which the interviewer confirms that the respondent is in an emotionally stable state of mind to conclude the interview. (If not, the interviewer will engage a prescribed distress protocol that has been part of the study training.) Typically, for studies involving sensitive topics, the interviewer will also provide a list of resources to all respondents, whether they experienced distress or not, that can be used after the interview in case the respondent would like to talk to someone about their feelings. These may be local or national hotlines or websites.

When transitioning to a self-administered survey, there is obviously no interviewer present who can detect whether the respondent is becoming upset by the interview. As such, it is impossible to engage a distress protocol. However, self-administered surveys can utilize a debriefing check-in at the end of the interview to gather feedback on how the respondent felt about the experience of completing the interviews (McClinton et al. 2015; Newman et al. 2006). They can also use the check-in to assess the respondent's emotional state and provide a list of resources (as either an insert in a paper mailing, or as a hyperlink in a web survey). In a web survey, the resources would ideally be placed in a static location on every screen so that if respondents become upset at any point during the interview, they could click on the resources link as a reference tool. For some projects, a help desk for tech support can sometimes take on

the duties of help line for emotional support as well. (Note that help desk staff would need to be trained to handle distress and provide support.)

While there is little research to look at the effects of providing distress-related information in a self-administered survey, future web-based research might look at whether people click on the resources, whether they actually utilize any of the resources, and whether they were helpful. For example, if a study involved a reliability component in which respondents were re-interviewed a few weeks after the initial interview, the second interview could ask about whether the resources had been noticed and used when provided in the first interview.

10.5 Known Adult Respondent

Self-administered surveys take the control over the selected informant or respondent out of the hands of the researcher and into the hands of the respondent. This has both statistical implications for a probability-based sample, and human subjects issues related to informed consent.

In an interviewer-administered survey, interviewers speak with and select adults from the household to be the household informant or a sampled adult respondent. The adult is then provided with the information in the informed consent document and script. **In self-administered surveys, who exactly is completing the questionnaire is out of the researcher's control.** When the survey is trying to reach a named respondent, either in a list sample or a longitudinal survey, additional steps may be required to verify the identity of the respondent. When the survey uses a within-household selection procedure to identify the respondent, although survey organizations may instruct an adult in the household to complete a survey,

whether the actual survey respondent is an adult or a child is unknown. Children cannot provide consent according to federal guidelines; rather, a parent provides consent for them, and the child assents to participate. Some have suggested asking the respondent to provide information that only an adult will know (e.g., credit card information; Kraut, et al. 2004), although this seems to increase the provision of PII and otherwise increase the risk of identification. Researchers should anticipate the risk of children or teenagers answering self-administered or mixed-mode surveys. This risk is low, given the level of burden of many surveys, but it is likely nonzero.

10.6 Summary and Takeaways

- 10.6.1 Regardless of the mode of data collection, human subjects protections should be guided by the ethical principles from *The Belmont Report*, including respect for persons; beneficence; and justice.
- 10.6.2 In interviewer-administered modes such as telephone or field surveys, the informed consent process is administered verbally by trained interviewers who should read every word of the informed consent statement, helping respondents understand meaning. In a self-administered mode, the respondent must be sufficiently literate to read and understand the consent form, and there is no guarantee that respondents in self-administered mode will pay careful attention to each word.
- 10.6.3 Experiments that vary the display, attention, or comprehension attributes of informed consent information between interviewer-administered and self-administered surveys would be beneficial future research for surveys that transition.
- 10.6.4 In transitioning a survey to a self-administered environment, the nature and opportunities to reveal personally identifiable information (PII) during the survey will vary. Survey

organizations may need to develop new protocols for how to enter and clean data with this potential for PII.

- 10.6.5 In an interviewer-administered context, if the respondent expresses at any point during the interview that they are being abused or are a danger to themselves or to others, the interviewer needs to report this to the appropriate healthcare or law enforcement agencies. When transitioning to a self-administered environment, factors that influence mandatory reporting change, especially the timeliness within which mandatory reporting can occur.
- 10.6.6 In a self-administered environment, there is no interviewer present who can detect whether the respondent is becoming upset by the interview, although counseling resources can be provided to respondents as needed, and follow-up check-ins can be conducted to evaluate respondent distress.
- 10.6.7 In interviewer-administered surveys, the interviewer can assure that an adult is a selected respondent. In self-administered surveys, who exactly is completing the questionnaire is out of the researcher's control.

11 Communicating the Impact of the Change of Modes

When changing from an interviewer-administered mode to self-administration, many things may be changing in addition to the mode of administration, as discussed throughout this report. Changing modes or adding a new mode often causes estimates to change due to mode effects, differential response bias, a change in the sampling frame and procedures, or a change in the survey content, among other reasons. For repeated cross-sectional surveys, or longitudinal or panel surveys in particular, these changes result in a break in series where estimates before the change in methodology cannot be compared to estimates after the change. It is therefore important to communicate this break in series to people who use the data and people within the organization.

Nearly all organizations (20 of 24 answering this question) reported that they communicated the transition to stakeholders and/or the public. Given the diversity of the studies and the clientele and audience for the studies, each organization's approach was somewhat unique. Commonly-mentioned approaches included posting details and analyses on websites, conducting one-on-one meetings with key clients, briefing member committees. Respondents reported communicating the costs and benefits of the transition to stakeholders, including the likelihood that key estimates would change as a result, survey content would have to be reduced, data collection time might be reduced and data quality improved. We end this report with some final thoughts about communicating breaks in a series of estimates based on mode differences.

11.1 How Do You Talk to the Public and Data Users about a Break in the Time Series?

Talking with data users about changes in the mode of data collection and other related design features is critical throughout the process. It is especially important to provide documentation to users after the transition is completed in order to summarize the decisions made and lessons learned. One common way to communicate changes in the data collection methodology and its impacts is to provide the information with the data file. When including changes in the data collection methodology along with other technical information for the data set, anyone using the data will see information about the change. Some organizations have written publicly available papers and journal articles that are available on their websites while others have developed methodology reports (see Polivka and Miller 1995; Brick, et al. 2013; Keeter 2019). It is important to identify what the original methodology was, what changed, why the change was made, and what the impact of the change is (if known). This includes any experimentation that was done or bridge studies. Reports should also contain any known differences in sample composition across the modes and any potential changes in the length of the field period, especially where events that occur during the field period may affect results (e.g., news polls about current events). Methodology reports should also note whether there were changes in question wording or format as a result of the transition; question crosswalks may be particularly helpful here. Providing this information allows data users to interpret the data correctly and draw sound conclusions.

This type of communication with data users is supported by AAPOR's Transparency Initiative. The Transparency Initiative recognizes organizations that disclose information about publicly released study methods in easy-to-understand language (AAPOR n.d.).

11.2 How Do You Communicate about a Break in the Time Series to the People in your Agency or Organization?

Equally important as communicating breaks in the time series to the public is communicating potential changes in data quality and changes in time trends within the agency or organization and with key stakeholders. This process not only helps get buy-in so everyone involved in the data release understands the methods used and the potential changes that occurred, but it also allows others to provide input to ensure the new methodology is sound. Developing new methodology and communicating it to the organization also can be done through advisory council meetings, technical review panels, or one-on-one meetings.

It is important to communicate why the change is occurring, how and why the new methodology was developed, and what the expected benefits are. Results from experimentation and related literature can be used to explain what the organization can expect to see from the survey moving forward, including costs, response rates and any changes in estimates. If changes in estimates are expected, plans for how to address the break in the time series should also be communicated, such as whether there will be a bridge study or modeling to help smooth the break, or whether the break will start the beginning of a new time series.

11.3 What Information Do We Need to Provide to Data Users on Data Files about Mode of Contact and Participation?

In addition to the response data itself, public data files typically contain information on the survey's target population, the frame, the sample design, including stratification variables, and information on the data collection, including the mode of recruitment, and response modes and the order/timing they are offered. They also indicate changes from previous data collection

cycles and their potential impact on estimates, which is particularly important when switching from a telephone survey to a self-administered survey or mixed-mode survey. Data files should contain information on the new data collection modes and specify how they are different from the previous survey cycles. Additionally, documentation should identify any changes to the sampling frame, and how the data collection procedures and question wording and formats differ, including screenshots of questionnaires in electronic modes.

To go a step further in helping data users understand the impact of change in mode, summary tables can be provided that outline the percentage of cases that responded by each mode. Additionally, flags can be placed on the response data file so users can subset by mode or compare across modes. Files can also include more detailed information, particularly for web, including the devices used to respond (smartphone, tablet, or PC) and the browsers used.

11.4 Conclusion

The transitions described in this report reflect the adaptability of the survey research profession as it confronts the profound challenges of growing nonresponse and costs, along with the opportunities provided by new technologies and databases. One clear conclusion of the report is that there is no single way that a survey is transitioned from telephone to self-administered or mixed modes of data collection. Each survey transition requires a package of decisions that affect all survey error sources. Some survey researchers prioritize comparability of survey estimates with the telephone modes of data collection, and thus make decisions to minimize any potential differences that may arise. Others prioritize maximizing the quality of the survey data collected in the new mode, and thus make decisions to optimize a design for the current set of modes. Which of these decisions is optimal is survey- and estimate-specific.

Clearly communicating these decisions, and how they may affect survey estimates, is key. If changes in estimates are expected, plans should be made and procedures for how to address the break in the time series should also be communicated. These plans and procedures may include reporting on a parallel or bridge study or statistical modeling to help smooth the changes in estimates. The plans may simply be that the new set of modes starts the beginning of a new time series. Results from experimentation and related literature can be used to explain what the organization can expect to see from the survey moving forward, including costs, response rates and any changes in estimates. We hope this report helps survey organizations consider, plan, and inform users about the important issues related to these transitions.

12 References

- 2011/12 National Survey of Children’s Health. Child and Adolescent Health Measurement Initiative (CAHMI), (2013), “2011-2012 NSCH: Child Health Indicator and Subgroups SAS Codebook, Version 1.0,” Data Resource Center for Child and Adolescent Health, sponsored by the Maternal and Child Health Bureau. www.childhealthdata.org.
- Abrajano, M., Alvarez, R. M. (2019), “Answering Questions About Race: How Racial and Ethnic Identities Influence Survey Response.” *American Politics Research*, 47(2), 250–274.
- Add Health (n.d.), “Add Health: National Longitudinal Study of Adolescent to Adult Health Wave IV,” retrieved from <https://www.cpc.unc.edu/projects/addhealth/design/wave4>.
- Al Baghal T., Lynn P. (2015), “Using Motivational Statements in Web-Instrument Design to Reduce Item-Missing Rates in a Mixed-Mode Context,” *Public Opinion Quarterly*, 79(2), 568-579.
- Allison C. M., Stevenson J., Kniss C. (2014), “Address Based Sampling as a Method of Tackling the ‘Cell Phone Problem’: Examples from the Wisconsin Family Health Survey,” Paper presented at 18th Minnesota Health Services Research Conference, Minnesota, MN.
- Allum N., Conrad F., Wenz A. (2018), “Consequences of Mid-Stream Mode-Switching in a Panel Survey,” *Survey Research Methods*, 12(1), 43-58.
- Amaya A., Biemer P., Kinyon D. (2018), “Total Error in a Big Data World: Adapting the TSE Framework to Big Data,” Paper presented at the Digital Traces Workshop, University of Bremen, Germany.

- Amaya A., Leclere F., Carris K., Liao Y. (2015), "Where to Start: An Evaluation of Primary Data-Collection Modes in an Address-Based Sampling Design," *Public Opinion Quarterly*, 79(2), 420-442.
- American Association for Public Opinion Research (n.d.), "IRB FAQs for Survey Researchers." Retrieved from <https://www.aapor.org/Standards-Ethics/Institutional-Review-Boards/IRB-FAQs-for-Survey-Researchers.aspx>, August 29, 2019.
- American Association for Public Opinion Research (n.d.), "Transparency Initiative: Latest News." Retrieved from <https://www.aapor.org/Standards-Ethics/Transparency-Initiative/Latest-News.aspx>, August 13, 2019.
- American National Election Studies (2015), "User's Guide and Codebook for the ANES 2012 Time Series Study," Ann Arbor, MI and Palo Alto, CA: the University of Michigan and Stanford University. Retrieved from https://www.electionstudies.org/wp-content/uploads/2012/02/anes_timeseries_2012_userguidecodebook.pdf.
- American National Election Studies (2018), "User's Guide and Codebook for the ANES 2016 Time Series Study," Ann Arbor, MI and Palo Alto, CA: the University of Michigan and Stanford University. Retrieved from https://electionstudies.org/anes_timeseries_2016_userguidecodebook/.
- Andreadis I. (2015), "Web Surveys Optimized for Smartphones: Are There Differences Between Computer and Smartphone Users?," *Methods, Data, Analysis*, 9(2), 213-228.
- Andrews R., Brick J. M., Mathiowetz N. (2013), "Continued Development and Testing of Dual-Frame Surveys of Fishing Effort: Testing a Dual-Frame, Mixed-Mode Survey Design-- Final Report," National Oceanic and Atmospheric Administration, Fisheries Statistics.

Retrieved from

<https://www.st.nmfs.noaa.gov/mdms/public/finalReport.jsp?ReportID=831>

Andrews R., Brick J. M., Mathiowetz N. A. (2014), "Development and Testing of Recreational Fishing Effort Surveys: Testing a Mail Survey Design," National Oceanic and Atmospheric Administration, Fisheries Statistics. Retrieved from

[https://www.cio.noaa.gov/services_programs/prplans/pdfs/ID296_MRIP_Effort_Survey_Final_Report%20\(1\).pdf](https://www.cio.noaa.gov/services_programs/prplans/pdfs/ID296_MRIP_Effort_Survey_Final_Report%20(1).pdf)

Ansolabehere S., Schaffner B. F. (2014), "Does Survey Mode Still Matter? Findings from a 2010 Multi-Mode Comparison," *Political Analysis*, 22(3), 285-303.

Ansolabehere S., Schaffner B. F. (2016), *Cooperative Congressional Election Study, 2016: Common Content*, Cambridge, MA: Harvard University.

Ansolabehere, S., Schaffner, B.F., Luks S., (2017), *Guide to the 2016 Cooperative Congressional Election Survey*. Cambridge, MA: Harvard University

<http://cces.gov.harvard.edu>

Antoun C., Couper M. P., Conrad F. G. (2017a), "Effects of Mobile Versus PC Web on Survey Response Quality: A Crossover Experiment in a Probability Web Panel," *Public Opinion Quarterly*, 81(S1), 280-306.

Antoun C., Katz J., Argueta J., Wang L. (2017b), "Design Heuristics for Effective Smartphone Questionnaires," *Social Science Computer Review*, 36(5), 557-574.

AP. (2018), "AP Votecast Methodology," AP Votecast, NORC at the University of Chicago. retrieved from <https://www.ap.org/en-us/topics/politics/votecast-methodology>.

Aquilino W. S. (1994), "Interviewer Mode Effects in Surveys of Drug and Alcohol Use: A Field Experiment," *Public Opinion Quarterly*, 58, 210-240.

- Atkeson L. R., Addams A. N., Bryant L. A., Zilberman L., Saunders K. L. (2011), "Considering Mixed Mode Surveys for Questions in Political Behavior: Using the Internet and Mail to Get Quality Data at Reasonable Costs," *Political Behavior*, 33, 161-178.
- Axinn W. G., Wagner J., Couper M. P., Crawford S. (2018), "Campus Climate Surveys of Sexual Misconduct: Limiting the Risk of Nonresponse Bias," University of Michigan: Ann Arbor, MI, PSC Research Report No. 18-887.
- Bailey J. T., Grabowski G., Link M. W. (2010), "Your Home Was Specially Selected: Using Address Based Sampling as a Recruitment Technique," Proceedings of the 65th Annual Conference for the American Association for Public Opinion Research, Chicago, IL.
- Baker R., Blumberg S. J., Brick M., Couper M. P., Courtright M., Dennis J. M., Dillman D., Frankel M. R., Garland P., Groves R. M., Kennedy C., Krosnick J., Lavrakas P. J., Lee S., Link M., Piekarski L., Rao K., Thomas R. K., Zahs D. (2010), "AAPOR Report on Online Panels," *Public Opinion Quarterly*, 74(4), 711-781.
- Baker-Prewitt M., Miller J. (2013), "Mobile Research Risk: What Happens to Data Quality When Respondents Use a Mobile Device for a Survey Designed for a PC," Paper presented at the 2013 CASRO Online Research Conference, San Francisco, CA.
- Bandilla W., Couper M., Kaczmirek L. (2014), "The Effectiveness of Mailed Invitations for Web Surveys and the Representativeness of Mixed-Mode versus Internet Only Samples," *Survey Practice*, 7(4), 1-12.
- Battaglia M. P., Link M. W., Frankel M. R., Osborn L., Mokdad A. H. (2008), "An Evaluation of Respondent Selection Methods for Household Mail Surveys," *Public Opinion Quarterly*, 72(3), 459-469.

- Battle D., Megra M. W., Wan C. (2017), "Casting a Wide Net: Specification Error in Screening Homeschool Children," Paper presented at the 72nd Annual Conference of the American Association for Public Opinion Research, New Orleans, LA.
- Baumgardner, S. (2018), "When is the Best Time to Field your Survey? Trends in American Community Survey Response Rates." Poster presented at the 2018 American Association for Public Opinion Research Annual Conference, Denver, CO.
<https://www.census.gov/content/dam/Census/newsroom/press-kits/2018/aapor/aapor-poster-acr-response-rates.pdf>
- Beckler D. G., Ott K., Horvath P. (2005), *Indirect Monetary Incentives for the 2004 ARMS Phase III Core*, United States Department of Agriculture, National Agricultural Statistics Service, Washington, DC.
- Beebe T. J., Davern M. E., McAlpine D. D., Ziegenfuss J. K. (2007), "Comparison of Two Within-Household Selection Methods in a Telephone Survey of Substance Abuse and Dependence," *Annals of Epidemiology*, 17(6), 458-463.
- Behavioral Risk Factor Surveillance System (2006), "Summary Data Quality Report," Centers for Disease Control and Prevention, Retrieved from
https://www.cdc.gov/brfss/annual_data/2006/pdf/2006SummaryDataQualityReport.pdf
- Bentley, M. (2019), "Analyzing Self-Response in the 2018 End-to-End Test: Key Findings and Questions," Paper presented at the annual meeting of the Population Association of America, Austin, TX.
- Berkold, J. (2018), "Conducting a Mail-Push-to-Web Survey with a US General Population Audience: Lessons Learned from the Field," Paper presented at the Annual Meeting of the American Association for Public Opinion Research, Denver, CO.

- Bianchi, A., Biffignandi, S., Lynn, P. (2017), "Web-Face-to-Face Mixed-Mode Design in a Longitudinal Survey: Effects on Participation Rates, Sample Composition, and Costs," *Journal of Official Statistics*, 33(2), 385-408. doi:<https://doi.org/10.1515/jos-2017-0019>
- Biemer P. P., Murphy J., Zimmer S., Berry C., Deng G., Lewis K. (2018), "Using Bonus Monetary Incentives to Encourage Web Response in Mixed-Mode Household Surveys," *Journal of Survey Statistics and Methodology*, 6(2), 240-261.
- Biemer, P.P., Harris, K.M., Burke, B.J., Considine, K.A., Halpern, C.T., and Suchindran, C.M. 2018b. "From an In-Person to a Web-Mail Panel Survey Design: The Add Health Wave V Experience." Paper presented at the American Association for Public Opinion Research annual meeting, Denver, CO.
- Bilgen I., Dennis J. M., Ganesh N. (2019), "The Undercounted: Measuring the Impact of 'Nonresponse Follow-up' on Research Data," *AmeriSpeak White Paper*, Chicago: NORC at the University of Chicago, Retrieved from https://amerispeak.norc.org/Documents/Research/AmeriSpeak_WhitePaper2_NRFUImpactonOutcomeMeasures_APRIL2019.pdf.
- Bishop G. F., Hippler H. J., Schwarz N., Strack F. (1988), "A Comparison of Response Effects in Self-Administered and Telephone Surveys," in *Telephone Survey Methodology*, eds. Groves R. B., Biemer P. P., Lyberg L. E., Massey J. T., Nicholls II W. L., Waksberg J., pp. 321-358, New York, NY: John Wiley & Sons.
- Blake K. D., Portnoy D. B., Kaufman A. R., Jordan Lin C., Lo S. C., Backlund E., Cantor D., Hicks L., Lin A., Caporaso A., Davis T., Moser R. P., Hesse B. W. (2016), "Rationale, Procedures, and Response Rates for the 2015 Administration of NCI's Health

- Information National Trends Survey: HINTS-FDA 2015,” *Journal of Health Communication*, 21(12), 1269-1275.
- Blasius J. (2012), “Comparing Ranking Techniques in Web Surveys,” *Field Methods*, 24(4), 382-398.
- Blom A. G., Bosnjak M., Cornilleau A., Cousteaux A., Das M., Douhou S., Krieger U. (2016), “A Comparison of Four Probability-Based Online and Mixed-Mode Panels in Europe,” *Social Science Computer Review*, 34(1), 8-25.
- Blumberg S. J., Luke J. V. (2019), “Wireless Substitution: Early Release of Estimates From the National Health Interview Survey, July-December 2018,” *US Department of Health and Human Services, Centers for Disease Control and Prevention*, retrieved 8/13/2019 from <https://www.cdc.gov/nchs/data/nhis/earlyrelease/wireless201906.pdf>.
- Bonhomme S. (2018), “Active Management Framework to Monitor and Manage Data Collection,” Paper presented at the 2018 Federal Computer Assisted Survey Information Collection Workshops, Suitland, MD.
- Boonstra T. W., Larsen M. E., Townsend S., Christensen H. (2017), “Validation of a Smartphone App to Map Social Networks of Proximity,” *Plos One*, 12(12), 1-13.
- Borkan B. (2010), “The Mode Effect in Mixed-Mode Surveys: Mail and Web Surveys,” *Social Science Computer Review*, 28(3), 371-380.
- Bosa K., Gagnon F., Caron P. (2017), “Comparison of Three Methods to Select a Respondent for Household Online Surveys Using Mailed Invitations,” Paper presented at the 72nd Annual Conference of the American Association for Public Opinion Research, New Orleans, LA.

- Bosnjak M., Dannwolf T., Enderle T., Schaurer I., Struminskaya B., Tanner A., Weyandt K. W. (2018), "Establishing an Open Probability-Based Mixed-Mode Panel of the General Population in Germany: The GESIS Panel," *Social Science Computer Review*, 36(1), 103-115.
- Bosnjak M., Das M., Lynn P. (2016), "Methods for Probability-Based Online and Mixed-Mode Panels: Selected Recent Trends and Future Perspectives," *Social Science Computer Review*, 34(1), 3-7.
- Brambilla D. J., McKinlay S. M. (1987), "A Comparison of Responses of Mailed Questionnaires and Telephone Interviews in a Mixed Mode Health Survey," *American Journal of Epidemiology*, 126(5), 962-971.
- Bramlett, M. D., Blumberg, S. J., Zablotsky, B., George, J. M., Ormson, A. E., Frasier, A. M., Vsetecka, D.M., Williams, K.L., Skalland, B.J., Morrison, H.M., Santos, K.B., Pedlow, S., and Wang, F. (2017), *Design and Operation of the National Survey of Children's Health, 2011–2012* (Series 1, Number 59), Hyattsville, MD: National Center for Health Statistics.
- Braunsberger K., Wybenga H. Gates R. (2007), "A Comparison of Reliability between Telephone and Web-Based Surveys," *Journal of Business Research*, 60, 758-764.
- Breidt, J., Kreuter F., Lesser V., Moore D. L., Smyth J. D. (2018), *Comparison and Assessment of Vendor Results for the 2016 National Survey of Hunting, Fishing, and Wildlife-Associated Recreation: Evaluation Team Report*, Washington, DC: Association of Fish and Wildlife Agencies.
- Breton C., Cutler F., Lachance S., Mierke-Zatwarnicki A. (2017), "Telephone Versus Online Survey Modes for Election Studies: Comparing Canadian Public Opinion and Vote

- Choice in the 2015 Federal Election,” *Canadian Journal of Political Science*, 50(4), 1005-1036.
- Brick J. M., Andrews W. R., Brick P. D., King H., Mathiowetz N. A., Stokes L. (2012), “Methods for Improving Response Rates in Two-Phase Mail Surveys,” *Survey Practice*, 5(3), 1-7.
- Brick J. M., Andrews W. R., Mathiowetz N. A. (2016), “Single-Phase Mail Survey Design for Rare Population Subgroups,” *Field Methods*, 28(4), 381-395.
- Brick J. M., Brick P. D., Dipko S., Presser S., Tucker C., Yuan Y. (2007), “Cell Phone Survey Feasibility in the US: Sampling and Calling Cell Numbers versus Landline Numbers,” *Public Opinion Quarterly*, 71, 23-39.
- Brick J. M., Cervantes I. F., Lee S., Norman G. (2011), “Nonsampling Errors in Dual Frame Telephone Surveys,” *Survey Methodology*, 37(1), 1-12.
- Brick J. M., Lepkowski J. M. (2008), “Multiple Modes and Frame Telephone Surveys,” in *Advances in Telephone Survey Methodology*, eds. Lepkowski J. M., Tucker C., Brick J. M., de Leeuw E. D., Japac L., Lavrakas P. J., Link M. W., Sangster R. L., pp. 149-169, New York, NY: Wiley.
- Brick J. M., Lohr S., Edwards W. S., Giambo P., Broene P., Williams D., Dipko S. (2013), “National Survey of Crime Victimization Companion Study - Pilot,” Washington, DC: Bureau of Justice Statistics, retrieved from https://www.bjs.gov/content/pub/pdf/nevs-cs_prr.pdf.
- Brick J. M., Montaquila J. M., Han D., Williams D. (2012), “Improving Response Rates for Spanish Speakers in Two-Phase Mail Surveys,” *Public Opinion Quarterly*, 76(4), 721-732.

- Brick J. M., Steiger D., Bronson J., Finkelhor D., Donoghue B. (2018), "Improving the Measurement of Sexual Victimization among Children through a Redesign of the National Survey of Children's Exposure to Violence," Paper presented at the 2018 Federal Committee on Statistical Methodology, Washington, DC, Retrieved from <http://www.copafs.org/UserFiles/file/2018FCSM/G-1BrickBronsonFinkelhorSteigerandDonoghueNatSCEV.pdf>.
- Brick J. M., Waksberg J., Kulp D., Starer A. (1995), "Bias in List-Assisted Telephone Samples," *Public Opinion Quarterly*, 59(2), 218-235.
- Brick J. M., Williams D., Montaquila J. M. (2011), "Address-Based Sampling for Subpopulation Surveys," *Public Opinion Quarterly*, 75(3), 409-428.
- Brown E. M., Olson L. T., Farrelly M. C., Nonnemaker J. M., Battles H., Hampton J. (2018), "Comparing Response Rates, Costs, and Tobacco-Related Outcomes Across Phone, Mail, and Online Surveys," *Survey Practice*, 11(2), 1-14.
- Bucks, B., Fulford, S., Couper, M. (2018), "A Mixed-Mode and Incentive Experiment using Administrative Data," Paper presented at the Federal Computer Assisted Survey Information Collection Workshops, Suitland, MD.
<https://www.census.gov/fedcasic/fc2018/ppt/7ABucks.pdf>
- Buelens B., Van den Brakel J. A. (2011), *Inference in Surveys with Sequential Mixed-Mode Data Collection*, Statistics Netherlands, The Hague, Discussion paper 201121.
- Buelens B., Van den Brakel J. A. (2015), "Measurement Error Calibration in Mixed-Mode Sample Surveys," *Sociological Methods & Research*, 44(3), 391-426.

- Buelens B., Van den Brakel J. A. (2017), "Comparing Two Inferential Approaches to Handling Measurement Error in Mixed-Mode Surveys," *Journal of Official Statistics*, 33(2), 513-531.
- Bureau of Labor Statistics (2018), "Current Population Survey," US Census Bureau: Washington, DC, retrieved from <https://www.census.gov/programs-surveys/cps/technical-documentation/methodology/collecting-data.html>.
- Bureau of Labor Statistics (n.d.), "Consumer Expenditure Surveys," US Department of Labor, Washington, DC, retrieved from <https://www.bls.gov/cex/>.
- Buskirk T., Andrus C. H. (2014), "Making Mobile Browser Surveys Smarter: Results from a Randomized Experiment Comparing Online Surveys Completed via Computer or Smartphone," *Field Methods*, 26(4), 322-342.
- California Health Interview Survey (2016), *CHIS 2013-2014 Methodology Series: Report 1 - Sample Design*, Los Angeles, CA: UCLA Center for Health Policy Research.
- California Health Interview Survey (2016), *CHIS 2013-2014 Methodology Series: Report 4 - Response Rates*, Los Angeles, CA: UCLA Center for Health Policy Research.
- Callegaro M. (2010), "Do You Know Which Device Your Respondent Has Used to Take Your Online Survey?," *Survey Practice*, 3(6), 1-13.
- Callegaro M., Shand-Lubbers J., Dennis J. M. (2009), "Presentation of a Single Item versus a Grid: Effects on the Vitality and Mental Health Scales of the SF-36v2 Health Survey," Paper presented at the 64th Annual Conference of the American Association for Public Opinion Research, Hollywood, FL.

- Camillo F., D'Attoma I. (2011), "Tackling the Problem of Self Selection in the Integration of Different Data Collection Techniques," Preprint retrieved from <http://amsacta.unibo.it/3063/>.
- Cantor D., Coa K., Crystal-Mansour S., Davis T., Dipko S., Sigman R. (2009), *Health Information National Trends Survey (HINTS) 2007: Final Report*, Bethesda, MD: National Cancer Institute.
- Cantor D., Covell J., Davis T., Park I., Rizzo L. (2005), *Health Information National Trends Survey 2005 (HINTS 2005): Final Report*, Bethesda, MD: National Cancer Institute.
- Cantrell J., Hair E. C., Smith A., Bennett M., Rath J. M., Thomas R. K., Fahimi M., Dennis J. M., Vallone D. (2018), "Recruiting and Retaining Youth and Young Adults: Challenges and Opportunities in Survey Research for Tobacco Control," *Tobacco Control*, 27: 147-154.
- Carpenter H. (2018), *UK Household Longitudinal Study: Wave 8 Technical Report*. Kantar Public and ISER University of Essex, Colchester, Essex, UK. Retrieved from <https://www.understandingsociety.ac.uk/sites/default/files/downloads/documentation/mainstage/technical-reports/wave-8-technical-report.pdf>.
- Casady R. J., Lepkowski J. M. (1993), "Stratified Telephone Survey Designs," *Survey Methodology*, 19(1), 103-113.
- Catania J. A., Binson D., Canchola J., Pollack L. M., Hauck W. (1996), "Effects of Interviewer Gender, Interviewer Choice, and Item Wording on Responses to Questions Concerning Sexual Behavior," *Public Opinion Quarterly*, 60(3), 345-375.
- Centers for Disease Control and Prevention (CDC) (2017), *National Immunization Survey-Child: A User's Guide for the 2017 Public-Use Data File*, Center for Disease Control and

- Prevention National Center for Immunization and Respiratory Diseases and NORC at the University of Chicago.
- Centers for Disease Control and Prevention (CDC) (n.d.), “Behavioral Risk Factor Surveillance System,” retrieved from <https://www.cdc.gov/brfss/>.
- Cernat A. (2015), “The Impact of Mixing Modes on Reliability in Longitudinal Studies,” *Sociological Methods & Research*, 44(3), 427-457.
- Cernat A., Couper M. P., Ofstedal M. B. (2016), “Estimation of Mode Effects in the Health and Retirement Study Using Measurement Models,” *Journal of Survey Statistics and Methodology*, 4(4), 501-524.
- Chang L., Krosnick J. A. (2009), “National Surveys via RDD Telephone Interviewing Versus the Internet Comparing Sample Representativeness and Response Quality,” *Public Opinion Quarterly*, 73(4), 641-678.
- Chang L., Krosnick J. A. (2010), “Comparing Oral Interviewing with Self-Administered Computerized Questionnaires: An Experiment,” *Public Opinion Quarterly*, 74(1), 154-167.
- Chapman C., Hagedorn M. (2009), “Searching for Alternatives to a Random Digit Dial Telephone Interview — Redesigning the National Household Education Surveys,” Paper presented at the Federal Committee on Statistical Methodology Conference, Washington, DC.
- Charoenruk N. (2015), “Interviewer Voice Characteristics and Data Quality,” unpublished PhD dissertation, University of Nebraska-Lincoln.
- Charoenruk N., Olson K. M. (2018), “Do Listeners Perceive Interviewers’ Attributes from their Voices and do Perceptions Differ by Question Type?,” *Field Methods*, 30(4), 312-328.

- Cheung G., Maher P. (2015), "Lessons Learned on Preparing and Managing Mixed Mode Surveys," Paper presented at the 16th International Blaise Users Conference, Beijing, China.
- Christian L. M., Dillman D. A. (2004), "The Influence of Symbolic and Graphical Language Manipulations on Answers to Paper Self-Administered Questionnaires," *Public Opinion Quarterly*, 68, 57-80.
- Christian L. M., Dillman D. A., Smyth J. D. (2007a), "Helping Respondents Get it Right the First Time: The Relative Influence of Words, Symbols, and Graphics in Web and Telephone Surveys," *Public Opinion Quarterly*, 71(1), 113-125.
- Christian L. M., Dillman D. A., Smyth J. D. (2008), "The Effects of Mode and Format on Answers to Scalar Questions in Telephone and Web Surveys," in *Advances in Telephone Survey Methodology*, pp. 250-275. eds. Lepkowski J., Tucker C., Brick M., DeLeeuw E., Japec L., Lavrakas P., Link M., Sangster R., Hoboken, NJ: Wiley.
- Clark, S. L. (2017), *Analysis of the Household Roster Questions on the American Community Survey*. (ACS17-RER-02), Washington, DC: US Department of Commerce, US Census Bureau Retrieved from https://www.census.gov/content/dam/Census/library/working-papers/2017/acs/2017_Clark_01.pdf.
- Clements A. D., Parker C. R. (1998), "The Relationship Between Salivary Cortisol Concentrations in Frozen Versus Mailed Samples," *Psychoneuroendocrinology*, 23(6), 613-616.
- Clifford S., Jerit J. (2016), "Cheating on Political Knowledge Questions in Online Surveys: An Assessment of the Problem and Solutions," *Public Opinion Quarterly*, 80(4), 858-887.

- Coffey S. M. (2016), "Using System Paradata to Target and Evaluate Data Collection Operations," Seminar presented at the 2016 Meeting of the Washington Statistical Society, Washington, DC.
- Coffey S. M., Reist B., White M. (2013), "Monitoring Methods for Adaptive Design in the National Survey of College Graduates (NSCG)," Proceedings of the Survey Research Methods Section, American Statistical Association, Montreal, Canada, pp. 3085-3099.
- Conrad F. G., Couper M. P., Tourangeau R., Galesic M. (2005), "Interactive Feedback Can Improve Quality of Responses in Web Surveys," Proceedings of the Survey Research Methods Section, American Statistical Association, pp. 3835-3840.
- Conrad F. G., Couper M. P., Tourangeau R., Peytchev A. (2006), "Use and Non-Use of Clarification Features in Web Surveys," *Journal of Official Statistics*, 22, 245-269.
- Cornesse C., Bošnjak M. (2018), "Is There an Association Between Survey Characteristics and Representativeness? A Meta-Analysis," *Survey Research Methods*, 12(1), 1-13.
- Couper M. P. (2005), "Technology Trends in Survey Data Collection," *Social Science Computer Review*, 23(4), 486-501.
- Couper M. P., Kennedy C., Conrad F. G., Tourangeau R. (2011), "Designing Input Fields for Non-Narrative Open-Ended Responses in Web Surveys," *Journal of Official Statistics*, 27(1), 65-85.
- Couper M. P., Peterson G. J. (2017), "Why Do Web Surveys Take Longer on Smartphones?," *Social Science Computer Review*, 35(3), 357-377.
- Couper M. P., Tourangeau R., Conrad F. G., Singer E. (2006), "Evaluating the Effectiveness of Visual Analog Scales: A Web Experiment," *Social Science Computer Review*, 24(2), 227-245.

- Couper M. P., Tourangeau R., Conrad F. G., Zhang C. (2013), "The Design of Grids in Web Surveys," *Social Science Computer Review*, 31(3), 322-345.
- Couper M. P., Traugott M. W., Lamias M. J. (2001), "Web Survey Design and Administration," *Public Opinion Quarterly*, 65, 230-253.
- Couper, M. P. (2017), "New Developments in Survey Data Collection," *Annual Review of Sociology*, 43(1), 121-145.
- Couper, M. P., Tourangeau, R., Conrad, F. G., Singer, E. (2006), "Evaluating the Effectiveness of Visual Analog Scales: A Web Experiment," *Social Science Computer Review*, 24(2), 227-245.
- Curtin, R. (2019), *Survey of Consumers*. Retrieved from: <https://data.sca.isr.umich.edu/>
- Data Resource Center for Child & Adolescent Health (n.d.), "Full-Length NSCH Survey Instruments," Health Resources and Services Administration (HRSA) of the U.S. Department of Health and Human Services (HHS), retrieved from <https://www.childhealthdata.org/learn-about-the-nsch/survey-instruments>.
- David M., Little R. J. A., Samuhel M. E., Triest R. K. (1986), "Alternative Methods for CPS Income Imputation," *Journal of the American Statistical Association*, 81(393), 29-41.
- de Bruijne M., Das M., van Soest A., Wijnant A. (2015), "Adapting Grid Questions for Mobile Devices," Paper presented at the 2015 Biannual Meeting of the European Survey Research Association (ESRA), Reykjavik, Iceland.
- de Bruijne M., Wijnant A. (2013), "Comparing Survey Results Obtained via Mobile Devices and Computers: An Experiment With a Mobile Web Survey on a Heterogeneous Group of Mobile Devices versus a Computer-Assisted Web Survey," *Social Science Computer Review*, 31(4), 482-502.

- de Leeuw E. D. (1992), *Data quality in mail, telephone, and face-to-face surveys*, Amsterdam, The Netherlands: TT Publications.
- 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. (2018), "Mixed-Mode: Past, Present, and Future," *Survey Research Methods*, 12(2), 75-89.
- de Leeuw E. D., Hox, J.J., Dillman D. A., (2008), "Mixed-Mode Surveys: When and Why," in *International Handbook of Survey Methodology*, eds. de Leeuw E. D., Hox J. J., Dillman D. A., pp. 299-316, New York, NY: Taylor & Francis Group.
- de Leeuw E. D., Suzer-Gurtekin Z. T., Hox J. J. (2018), "The Design and Implementation of Mixed-Mode Surveys," in *Advances in Comparative Survey Methods: Multinational, Multiregional, and Multicultural Contexts (3MC)*, eds. Johnson T. P., Pennell B., Stoop I., Dorer B., pp. 387-409, Hoboken, NJ: John Wiley & Sons.
- de Leeuw E. D., Toepoel V. (2017), "Mixed-Mode and Mixed-Device Surveys," in *The Palgrave Handbook of Survey Research*, eds. Vannette D. L., Krosnick J. A., pp. 51-61, London: Palgrave Macmillan.
- de Leeuw E. D., Van der Zouwen J. (1988), "A Methodological Comparison of the Data Quality in Telephone and Face to Face Surveys: A Comparative Meta-Analysis," in *Telephone Survey Methodology*, eds. Groves R. M., Biemer P. P., Lyberg L. E., Massey J. T., Nicholls W. L., Waksberg J., pp. 283-299, New York, NY: Wiley.
- DeBell M., Amsbary M., Meldener V., Brock S., Maisal N. (2018), "Methodology Report for the ANES 2016 Time Series Study," Palo Alto, CA and Ann Arbor, MI: Stanford University and the University of Michigan.

- DeBell M., Jackman S., Maisel N., Amsbary M., Meldener V., Brick J. M., Krupenkin M., Peterson E. (2017), "Methodology and Findings of the ANES 2016 Recruitment Pretest Study," ANES Technical Report No. nes006978, American National Election Studies.
- Dennis J. M. (2019), "Technical Overview of the AmeriSpeak Panel: NORC's Probability-Based Household Panel," *AmeriSpeak White Paper*, Chicago: NORC at the University of Chicago, Retrieved from <http://amerispeak.norc.org/Documents/Research/AmeriSpeak%20Technical%20Overview%202019%2002%2018.pdf>.
- Derouvray C., Couper M. P. (2002), "Designing a Strategy for Reducing 'No Opinion' Responses in Web-Based Surveys," *Social Science Computer Review*, 20(1), 3-9.
- Dillman D. A. (2009), "Some Consequences of Survey Mode Changes in Longitudinal Surveys," in *Methodology of Longitudinal Surveys*, ed. P. Lynn, pp. 127-140, West Sussex, UK: John Wiley & Sons.
- Dillman D. A. (2012), "Introduction to Special Issue of Survey Practice on Item Nonresponse," *Survey Practice*, 5(2), 1-4.
- Dillman D. A., Allen T. B. (1995), *Census Booklet Questionnaire Evaluation Test: Phase 1 -- Summary of 20 Taped Interviews (Technical Report No. 95-41)*, Pullman, WA: Washington State University, Social and Economic Sciences Research Center.
- Dillman D. A., Brown T. L., Carlson J., Carpenter E. H., Lorenz F. O., Mason R., Saltiel J., Songster R. L. (1995), "Effects of a Category Order on Answers to Mail and Telephone Surveys," *Rural Sociology*, 60, 674-687.
- Dillman D. A., Christian L. M. (2005), "Survey Mode as a Source of Instability in Responses Across Surveys," *Field Methods*, 17(1), 30-52.

- Dillman D. A., Edwards M. L. (2016), "Designing a Mixed-Mode Survey," in *The Sage Handbook of Survey Methodology*, eds. Wolf C., Joye D., Smith T. W., Fu Y., pp. 255-268, London: SAGE Publications.
- Dillman D. A., Parsons N. L., Mahon-Haft T. (2004), *Connections Between Optical Features and Respondent Friendly Design: Cognitive Interview Comparisons of the Census 2000 Form and New Possibilities (Technical Report No. 04-030)*, Pullman, WA: Washington State University, Social and Economic Sciences Research Center.
- Dillman D. A., Phelps G., Tortora R., Swift K., Kohrell J., Berck J., Messer B. J. (2009), "Response Rate and Measurement Differences in Mixed Mode Surveys Using Mail, Telephone, Interactive Voice Response, and the Internet," *Social Science Research*, 38(1), 1-18.
- Dillman D. A., Redline C. D. (2004), "Testing Paper Self-Administered Questionnaires: Cognitive Interviewing and Field Test Comparisons," in *Methods for Testing and Evaluating Survey Questionnaires*, eds. Presser S., Couper M. P., Lessler J. T., Martin E., Martin J., Rothgeb J. M., Singer E., pp. 239-317, New York: John Wiley.
- Dillman D. A., Sangster R. L., Tarnai J., Rockwood T. H. (1996), "Understanding Differences in People's Answers to Telephone and Mail Surveys," in *New Directions For Evaluation. Current Issues for Survey Research*, eds. Braverman M. T., Slater J. K., pp. 46-52, San Francisco, CA: Jossey-Bass.
- Dillman D. A., Smyth J. D., Christian L. M. (2009), *Internet, Mail and Mixed-Mode Surveys: The Tailored Design Method*, Hoboken, NJ: John Wiley & Sons Inc.
- Dillman D. A., Smyth J. D., Christian L. M. (2014), *Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Methods (4th ed.)*, Hoboken, NJ: John Wiley & Sons Inc.

- Dillman D. A., Tarnai J. (1991), "Mode effects of Cognitively-Designed Recall Questions: A Comparison of Answers to Telephone and Mail Surveys," in *Measurement Errors in Surveys*, eds. Biemer P. P., Groves R. M., Lyberg L. E., Mathiowetz N. A., Sudman S., 73-93, New York, NY: John Wiley & Sons.
- DiSogra C., Callegaro M. (2016), "Metrics and Design Tool for Building and Evaluating Probability-Based Online Panels," *Social Science Computer Review*, 34(1), 26-40.
- DiSogra C., Dennis J. M. Fahimi M. (2010), "On the Quality of Ancillary Data Available for Address-Based Sampling," Proceedings of the Survey Research Methods Section, American Statistical Association, Vancouver, British Columbia.
- Domnich A., Panatto D., Signori A., Bragazzi N. L., Cristina M. L., Amicizia D., Gasparini R. (2015), "Uncontrolled Web-Based Administration of Surveys on Factual Health-Related Knowledge: A Randomized Study of Untimed Versus Timed Quizzing," *Journal of Medical Internet Research*, 17(4), e94.
- Durdiakova J., Fabryova H., Koborova I., Ostatnikova D., Celec P. (2013), "The Effects of Saliva Collection, Handling and Storage on Salivary Testosterone Measurement," *Steroids*, 78, 1325-1331.
- Dutwin D. (2019), "Non-Probability Samples: Emerging Methods and Models for High Quality Research," SSRS. retrieved from <https://ssrs.com/non-probability-samples-research-methods/>.
- Dutwin D., Buskirk T. D. (2017), "Apples to Oranges or Gala versus Golden Delicious? Comparing Data Quality of Nonprobability Internet Samples to Low Response Rate Probability Samples," *Public Opinion Quarterly*, 81(S1), 213-239.

- Dutwin D., Lavrakas P. (2016), "Trends in Telephone Outcomes, 2008-2015," *Survey Practice*, 9(2), 1-9.
- Dutwin D., Malarek D. (2014), "The Use of Recent Activity Flags to Improve Cellular Telephone Efficiency," *Survey Practice*, 7(1), 1-10.
- Dykema J., DiLoreto K., Croes K. D., Garbarski D., Beach J. (2017), "Factors Associated with Participation in the Collection of Saliva Samples by Mail in a Survey of Older Adults," *Public Opinion Quarterly*, 81(1), 57-85.
- Dykema J., Jones N. R., Piché T., Stevenson J. (2013), "Surveying Clinicians by Web: Current Issues in Design and Administration," *Evaluation & the Health Professions*, 36, 352-381.
- Edgar J., Scanlon P. (2017), "Apples or Oranges: What is the Right Question When Comparing Web Probing and Cognitive Interviewing?," Paper presented at the 72nd Annual Conference of the American Association for Public Opinion Research, New Orleans, LA.
- Edwards B., Maitland A., Connor S. (2017), "Measurement Error in Survey Operations Management: Detection, Quantification, Visualization, and Reduction," in *Total Survey Error in Practice*, eds. P. Biemer, E. deLeeuw, S. Eckman, B. Edwards, F. Kreuter, L. E. Lyberg, N. C. Tucker, B. West, pp. 255-278, Hoboken, NJ: John Wiley & Sons.
- Elkasabi M., Suzer-Gurtekin Z. T., Lepkowski J. M., Kim U., Curtin R., McBee R. (2014), "A Comparison of ABS Mail and RDD Surveys for Measuring Consumer Attitudes," *International Journal of Market Research*, 56(6), 737-756.
- Elliott M. R., Raghunathan T. E., Schenker N. (2018), *Combining Estimates from Multiple Surveys*, Hoboken, NJ: John Wiley & Sons.
- Elliott M. R., Valliant R. (2017), "Inference for Nonprobability Samples," *Statistical Science*, 32(2), 249-264.

- Elliott M. R., West B. T. (2015), "Clustering by Interviewer: A Source of Variance That is Unaccounted for in Single-Stage Health Surveys," *American Journal of Epidemiology*, 182(2), 118-126.
- Elliott, M. N., Zaslavsky, A. M., Goldstein, E., Lehrman, W., Hambarsoomians, K., Beckett, M. K., Giordano, L. (2009), "Effects of Survey Mode, Patient Mix, and Nonresponse on CAHPS® Hospital Survey Scores," *Health Services Research*, 44(2p1), 501-518.
- Ellis C., Aspinwall K., Heinrich T., Ginder S., McDonald H. S., Noonan M. (2013), "The Effects of Pushing Web in a Mixed-Mode Establishment Data Collection," Paper presented at the 68th Annual Conference of the American Association for Public Opinion Research, Boston, MA.
- Epi Info (n.d.), "Epi Info for Windows," retrieved from <https://www.cdc.gov/epiinfo/pc.html>.
- Fahimi M., Kulp D., Brick J. M (2009), "A Reassessment of List-Assisted RDD Methodology," *Public Opinion Quarterly*, 73(4), 751-760.
- Federal Communications Commission (Updated May 18, 2016), "Wireless Local Number Portability (WLNP)," Retrieved from <https://www.fcc.gov/general/wireless-local-number-portability-wlnp>
- Federal Highway Administration and Westat (2018), *2017 NHTS Data User Guide*, Federal Highway Administration, Office of Policy Information, Washington, DC. Retrieved from https://nhts.ornl.gov/assets/NHTS2017_UsersGuide_04232019_1.pdf
- Fessler P., Kasy M., Lindner P. (2018), "Survey Mode Effects on Measured Income Inequality," *The Journal of Economic Inequality*, 16(4), 487-505.

- Finamore J., Coffey S., Reist B. (2015), "Enhancing the Use of Adaptive Design in a National Survey of College Graduates," Paper presented at the 70th Annual Conference of the American Association for Public Opinion Research, Hollywood, FL.
- Finamore, John (2019), "The Role of Survey Paradata in a Federal Statistical Agency's Commitment to Quality." Paper presented at the 2019 FedCASIC Workshop, Washington DC.
- Finkel S. E., Guterbock T. M., Borg M. J. (1991), "Race of Interviewer Effects in a Preelection Poll Virginia 1989," *Public Opinion Quarterly*, 55(3), 313-330.
- Fisher G. G., Ryan L. H. (2018), "Overview of the Health and Retirement Study and Introduction to the Special Issue," *Work, Aging and Retirement*, 4(1), 1-9.
- Fishing Effort Survey (N.A.), National Oceanic and Atmospheric Administration, Fisheries Service, Department of Commerce. Retrieved from <https://www.fisheries.noaa.gov/webdam/download/76441998>.
- Fowler F. J., Gallagher P. M., Nederend S. (1999), "Comparing Telephone and Mail Responses to the CAHPS (TM) Survey Instrument," *Medical Care*, 37(3), MS41-MS49.
- Fowler F. J., Mangione T. W. (1990), *Standardized Survey Interviewing*, Newbury Park, CA: Sage Publications.
- Francis J., Laflamme G. (2015), "Evaluating Web Data Collection in the Canadian Labour Force Survey," Paper presented at the Federal Committee on Statistical Methodology Research Conference. Retrieved from https://nces.ed.gov/fcsm/pdf/H2_Francis_2015FCSM.pdf.
- Frauke K., Presser S., Tourangeau R. (2008), "Social Desirability Bias in CATI, IVR, and Web Surveys: The Effects of Mode and Question Sensitivity," *Public Opinion Quarterly*, 72(5), 847-865.

- Freedman, V. A. (2017), "The Panel Study of Income Dynamics' Wellbeing and Daily Life Supplement (PSID-WB) User Guide: Final Release 1," Institute for Social Research, University of Michigan.
- Fricker R. D., Schonlau M. (2002), "Advantages and Disadvantages of Internet Research Surveys: Evidence from the Literature," *Field Methods*, 14(4), 346-367.
- Fricker S., Galesic M., Tourangeau R., Yan T. (2005), "An Experimental Comparison of Web and Telephone Surveys," *Public Opinion Quarterly*, 69(3), 370-392.
- Fuchs M. (2009a), "Asking for Numbers and Quantities: Visual Design Effects in Paper & Pencil Surveys," *International Journal of Public Opinion Research*, 21(1), 65-84.
- Fuchs M. (2009b), "Differences in the Visual Design Language of Paper-and-Pencil Surveys Versus Web Surveys: A Field Experimental Study on the Length of Response Fields in Open-Ended Frequency Questions," *Social Science Computer Review*, 27, 213-227.
- Fulton J. A. (2012), "Respondent Consent to Use Administrative Data," unpublished dissertation, University of Maryland, College Park.
- Funke F., Reips U. D., Thomas R. K. (2011), "Sliders for the Smart: Type of Rating Scale on the Web Interacts With Educational Level," *Social Science Computer Review*, 29(2), 221-231.
- Galesic M., Tourangeau R., Couper M. P., Conrad F. G. (2008), "Eye-Tracking Data: New Insights on Response Order Effects and Other Cognitive Shortcuts in Survey Responding," *Public Opinion Quarterly*, 72(5), 892-913.
- Gallagher P. M., Fowler F. J., Stringfellow V. L. (1999), "Respondent Selection by Mail: Obtaining Probability Samples of Health Plan Enrollees," *Medical Care*, 37(3), MS50-MS58.

- Gatny H. H., Couper M. P., Axinn W. G. (2013), "New Strategies for Biosample Collection in Population-Based Social Research," *Social Science Research*, 42, 1402-1409.
- Gaziano C. (2005), "Comparative Analysis of Within-Household Respondent Selection Techniques," *Public Opinion Quarterly*, 69(1), 124-157.
- Ghandour R. M., Jones J. R., Lebrun-Harris L. A., Minnaert J., Blumberg S. J., Fields J., Bethell C., Kogan M. D. (2018), "The Design and Implementation of the 2016 National Survey of Children's Health," *Maternal and Child Health Journal*, 22(8), 1093-1102.
- Gooch A., Vavreck L. (2019), "How Face-to-Face Interviews and Cognitive Skill Affect Item Non-Response: A Randomized Experiment Assigning Mode of Interview," *Political Science Research and Methods*, 7(1), 143-162.
- Gotschi E., Delve R., Freyer B. (2009), "Participatory Photography as a Qualitative Approach to Obtaining Insights into Farmer Groups," *Field Methods*, 21, 290-308.
- Grady, R. H., Greenspan, R. L., Liu, M. (2019), "What Is the Best Size for Matrix-Style Questions in Online Surveys?" *Social Science Computer Review*, 37(3), 435-445.
- Graesser A. C., Cai Z., Louwse M. M., Daniel F. (2006), "Question Understanding Aid (QUAID): A Web Facility that Tests Question Comprehensibility," *Public Opinion Quarterly*, 70, 3-22.
- Greene J., Speizer H., Wiitala W. (2008), "Telephone and Web: Mixed-Mode Challenge," *Health Services Research*, 43(1), 230-248.
- Gregoski M. J., Mueller M., Vertegel A., Shaporev A., Jackson B. B., Frenzel R. M., Sprehn S. M., Treiber F. A. (2012), "Development and Validation of a Smartphone Heart Rate Acquisition Application for Health Promotion and Wellness Telehealth Applications," *International Journal of Telemedicine and Applications*, Article ID 696324, 1-7.

- Griffin D. "What the American Community Survey Can Tell Us about Mixed-Mode Surveys." Brown Bag Seminar, University of Michigan, Institute for Social Research, Ann Arbor, MI, February 3, 2009. Retrieved from <https://www.psc.isr.umich.edu/dis/workshop/references/ISRMixedMode.pdf>, August 13, 2019.
- Griffis S. E., Goldsby T. J., Cooper M. (2003), "Web-Based and Mail Surveys: A Comparison of Response, Data, and Cost," *Journal of Business Logistics*, 24(2), 237-258.
- Groves R. M. (1989; 2004), *Survey Errors and Survey Costs*, Hoboken, NJ: John Wiley & Sons, Inc.
- Groves R. M., Couper M. P., Presser S., Singer E., Tourangeau R., Acosta G. P., Nelson L. (2006), "Experiments in Producing Nonresponse Bias," *International Journal of Public Opinion Quarterly*, 70(5), 720-736.
- Groves R. M., Fultz N. H. (1985), "Gender Effects Among Telephone Interviewers in a Survey of Economic Attitudes," *Sociological Methods and Research*, 14(1), 31-52.
- Groves R. M., Heeringa S. G. (2006), "Responsive Design for Household Surveys: Tools for Actively Controlling Survey Errors and Costs," *Journal of the Royal Statistical Society*, 169(3), 439-457.
- Groves R. M., Kahn R. L. (1979), *Surveys by Telephone: A National Comparison with Personal Interviews*, New York, NY: Academic Press.
- Groves, R. M., Lepkowski, J. M. (1985), "Dual frame, mixed mode survey designs." *Journal of Official Statistics*, 1(3), 263-286
- Guidry K. R. (2012), "Response Quality and Demographic Characteristics of Respondents Using a Mobile Device on a Web-Based Survey," unpublished manuscript, Indiana University.

- Guidry K. R. (2012), "Response Quality and Demographic Characteristics of Respondents Using a Mobile Device on a Web-Based Survey," Paper presented at the 67th Annual Conference of the American Association for Public Opinion Research, Orlando, FL.
- Hagan D. E., Collier C. M. (1983), "Must Respondent Selection Procedures for Telephone Surveys be Invasive?" *Public Opinion Quarterly*, 47(4), 547-556.
- Han D., Cantor D., Brick P. D., Sigman R., Aponte M. (2010), "Findings from a Two-Phase Mail Survey for a Study of Veterans," Proceedings of the Survey Research Methods Section, American Statistical Association, Vancouver, British Columbia, pp. 2776-2782.
- Han D., Montaquila J. M., Brick J. M. (2013), "An Evaluation of Incentive Experiments in a Two-Phase Address-Based Sample Mail Survey," *Survey Research Methods*, 7(3), 207-218.
- Hansen T. O., Simonsen M. K., Nielsen F. C., Andersen Hundrup Y. (2007), "Collection of Blood, Saliva, and Buccal Cell Samples in a Pilot Study on the Danish Nurse Cohort: Comparison of the Response Rate and Quality of Genomic DNA," *Cancer Epidemiology, Biomarkers & Prevention*, 16(10), 2072-2076.
- Harris K.M. (2018), "Overview of Add Health for New Data Users," Paper presented at the 2018 Add Health Users Conference, Bethesda, MD.
- Harris L. E., Weinberger M., Tierney W. M. (1997), "Assessing Inner City Patients' Hospital Experiences. A Controlled Trial of Telephone Interviews Versus Mailed Surveys," *Medical Care*, 35(1), 70-76.
- Harter R., Battaglia M. P., Buskirk T. D., Dillman D. A., English N., Fahimi M., Frankel M. R., Kennel T., McMichael J. P., McPhee C. B., Montaquila J., Yancey T., Zukerberg A. L. (2016), "AAPOR Report: Address-Based Sampling," retrieved 5/16/2018 from

<https://www.aapor.org/Education-Resources/Reports/Address-based-Sampling.aspx#SECTION%203>.

Hatchett S., Schuman H. (1975), "White Respondents and Race-of-Interviewer Effects," *Public Opinion Quarterly*, 39(4), 523-528.

HCAHPS. (2008), *Mode and Patient-mix Adjustment of the CAHPS® Hospital Survey (HCAHPS)* Retrieved from <https://www.hcahpsonline.org/globalassets/hcahps/mode-patient-mix-adjustment/final-draft-description-of-hcahps-mode-and-pma-with-bottom-box-modedoc-april-30-2008.pdf>

Health and Retirement Study (2017), "2017 Consumption And Activities Mail Survey (CAMS 2017) Final, Version 1.0," Institute for Social Research, University of Michigan, Ann Arbor, MI, retrieved from

http://hrsonline.isr.umich.edu/modules/meta/2017/cams/desc/2017CAMS_DD.pdf

Health and Retirement Study (2019), "Data Collection Path Diagram," Institute for Social Research, University of Michigan, Ann Arbor, MI, retrieved from

<https://hrs.isr.umich.edu/data-products/collection-path>.

Health and Retirement Study (n.d.), "Available Products: Administrative Linkages," retrieved from <https://hrs.isr.umich.edu/data-products/restricted-data/available-products>.

Health and Retirement Study (n.d.), "Questionnaires." Retrieved from

<https://hrs.isr.umich.edu/documentation/questionnaires>.

Heerwegh D. (2009), "Mode Differences Between Face-to-Face and Web Surveys: An Experimental Investigation of Data Quality and Social Desirability Effects," *International Journal of Public Opinion Research*, 21, 111-121.

- Heerwegh D., Loosveldt G. (2008), "Face-to-Face versus Web Surveying in a High-Internet-Coverage Population," *Public Opinion Quarterly*, 72(5), 836-846.
- Hepner K. A., Brown J. A., Hays R. D. (2005), "Comparison of Mail and Telephone in Assessing Patient Experiences in Receiving Care from Medical Group Practices," *Evaluation the Health Professions*, 28(4), 377-389.
- Hicks W., Cantor D. (2012), "Evaluating Methods to Select a Respondent for a General Population Mail Survey," Paper presented at the 67th Annual Conference of the American Association for Public Opinion Research, Orlando, FL.
- Hochstim J. R. (1967), "A Critical Comparison of Three Strategies of Collecting Data From Households," *Journal of the American Statistical Association*, 62, 976-989.
- Hoebel J., von der Lippe E., Lange C., Ziese T. (2014), "Mode Differences in a Mixed-Mode Health Interview Survey Among Adults," *Archives of Public Health*, 72(46), 1-12.
- Holland J., Chrisian L. M. (2007), "The Influence of Interactive Probing on Response to Open-Ended Questions in a Web Survey," Paper presented at the Southern Association for Public Opinion Research Annual Conference, Raleigh, NC.
- Horwitz R., Vasquez J., personal email correspondence, Assistant Survey Director, National Survey of College Graduates. Subject line: NSCG Debit Cards. Date: 5/8/2019.
- Hox J. J., de Leeuw E. D., Klausch T. (2015), "Mixed Mode Research: Issues in Design and Analysis," Paper presented at the International Conference on Total Survey Error: Improving Quality in the Era of Big Data, Baltimore, MD.
- Hox J. J., de Leeuw E. D., Klausch T. (2017), "Mixed-Mode Research: Issues in Design and Analysis," in *Total Survey Error in Practice*, eds. Biemer P. P., de Leeuw E. D., Eckman S., pp. 511-531, Hoboken, NJ: John Wiley & Sons.

- Hox J. J., de Leeuw E. D., Zijlmans E. A. O. (2015), "Measurement Equivalence in Mixed Mode Surveys," *Frontiers in Psychology*, 5(February), 1-11.
- Huddy L., Billig J., Bracciodieta J., Hoeffler L., Moynihan P. J., Pugliani P. (1997), "The Effect of Gender on the Survey Response," *Political Behavior*, 19(3), 197-220.
- Hunsecker J. (2018), "Analog or Digital: Methods for Pre-testing Surveys and Websites." Paper presented at the Annual Conference of the American Association for Public Opinion Research, Denver, CO.
- Hyman H. H., Cobb W. J., Feldman J. J., Hart C. W., Stember C. H. (1954), *Interviewing in Social Research*, Chicago, IL: The University of Chicago Press.
- Hyman H. H., Sheatsley P. B. (1950), "The Current Status of American Public Opinion," in *The Teaching of Contemporary Affairs: Twenty-First Yearbook of the National Council for the Social Studies*, eds. Payne J. C., pp. 11-34, New York, NY: National Education Association.
- Iannacchione V. G. (2011), "The Changing Role of Address-Based Sampling in Survey Research," *Public Opinion Quarterly*, 75(3), 556-575.
- Ingels S. J., Pratt D. J., Herget D. R., Burns L. J., Dever J. A., Ottem R., Rogers J. E., Jin Y., Leinwand S., LoGerfo L. (2011), "High School Longitudinal Study of 2009 (HSL:09): Base-Year Data File Documentation," NCEES 2011-328, National Center for Education Statistics, US Department of Education.
- IPUMS-USA (n.d.), "IPUMS USA RESPMODE," Minnesota Population Center, University of Minnesota, retrieved from https://usa.ipums.org/usa-action/variables/RESPMODE#codes_section.

- Israel G., Lamm A. (2012), "Item Nonresponse in a Client Survey of the General Public," *Survey Practice*, 5(2), 1-6.
- Jackle A., Callegaro M. (2008), "Dependent Interviewing," in *Encyclopedia of Survey Research Methods*, eds. P. J. Lavrakas, pp. 187-188, Thousand Oaks, CA: SAGE Publications, Inc.
- Jackle A., Lynn P., Burton J. (2015), "Going Online with a Face-to-Face Household Panel: Effects of a Mixed Mode Design on Item and Unit Non-Response," *Survey Research Methods*, 9(1), 57-70.
- Jackle A., Roberts C., Lynn P. (2010), "Assessing the Effect of Data Collection Mode on Measurement," *International Statistical Review*, 78(1), 3-20.
- Jackman S. (2015), "Election Studies in the 21st Century," presented at the 70th Annual Conference of the American Association for Public Opinion Research, Hollywood, FL.
- Jackson M. T., McPhee C. B., Lavrakas P. J. (2019), "Using Response Propensity Modeling to Allocate Noncontingent Incentives to an Address-Based Sample: Evidence from a National Experiment," *Journal of Survey Statistics and Methodology*, smz007.
- Jans M., Grant D., Lee A., Park R., Edwards S., Rauch J., Flores-Cervantes I. (2013), "Address-Based Sampling (ABS) as an "Alternative" to RDD: A Test in California," Paper presented at the 68th Annual Conference of the American Association for Public Opinion Research, Boston, MA.
- Jans M., Park R., Rauch J., Grant D., Edwards S. (2015), "Logos and Inserts can Reduce Survey Return Rates: An Experiment in California," *Survey Practice*, 8(4), 1-12.
- Jans M., Sirkis R., Morgan D. (2013), "Managing Data Quality Indicators with Paradata Based Statistical Quality Control Tools: The Keys to Survey Performance," in *Improving*

- Surveys with Paradata: Analytic Uses of Process Information*, ed. F. Kreuter, pp. 191-229, Hoboken, NJ: John Wiley & Sons.
- Japac L. (2008), "Interviewer Error and Interviewer Burden," in *Advances in Telephone Survey Methodology*, eds. J. M. Lepkowski, C. Tucker, J. M. Brick, E. D. de Leeuw, L. Japac, P. J. Lavrakas, M. W. Link, R. L. Sangster, pp. 187-211, Hoboken, NJ: John Wiley & Sons, Inc.
- Javeline D. (1999), "Response Effects in Polite Cultures: A Test of Acquiescence in Kazakhstan," *Public Opinion Quarterly*, 63(1), 1-28.
- Jenkins K., Koning A. (2019), "Joint Rutgers-Eagleton/FDU Poll: Political Leaders Not Really Doing It for Garden Staters," Rutgers University Center for Public Interest Polling, retrieved from <http://eagletonpoll.rutgers.edu/rutgers-eagleton-fdu-nj-favorable-ratings/>.
- Jensen C., Thomsen J. P. F. (2014), "Self-Reported Cheating in Web Surveys on Political Knowledge," *Quality & Quantity*, 48(6), 3343-3354.
- Jones J., Saad L., Newport F., Marken S. (2015), "A Mail Survey Experiment Using Gallup's Annual Crime Survey," Paper presented at the 70th Annual Conference of the American Association for Public Opinion Research, Hollywood, FL.
- Kali J., Flores Cervantes I. (2016), "Conducting a Telephone Survey Using an ABS Sample: A Case Study of the California Health Interview Survey," Paper presented at the 71st Annual Conference for the American Association for Public Opinion Research, Austin, TX.
- Kappelhof J. (2015), "The Impact of Face-to-Face vs Sequential Mixed-Mode Designs on the Possibility of Nonresponse Bias in Surveys Among Non-Western Minorities in the Netherlands," *Journal of Official Statistics*, 31(1), 1-31.

Keeter S. (2019), "Growing and Improving Pew Research Center's American Trends Panel,"
Pew Research Center, Washington, DC, retrieved from

<https://www.pewresearch.org/methods/2019/02/27/growing-and-improving-pew-research-centers-american-trends-panel/>.

Keeter S., Hatley, N., Kennedy, C., Lau, A. (2017), *What Low Response Rates Mean for Telephone Surveys*. Pew Research Center, Washington, DC:

<https://assets.pewresearch.org/wp-content/uploads/sites/12/2017/05/12154630/RDD-Non-response-Full-Report.pdf>

Keeter S., Kennedy C. (2006), *National Polls Not Undermined by Growing Cell-Only Population: The Cell Phone Challenge to Survey Research*, Pew Research Center,

Washington, DC, retrieved from <https://www.pewresearch.org/wp-content/uploads/sites/4/legacy-pdf/276.pdf>.

Keeter S., McGeeney, K., Igielnik, R., Mercer, A., Mathiowetz, N. A. (2015), *From Telephone to the Web: The Challenge of Mode of Interview Effects in Public Opinion Polls*. Pew Research Center: Washington, DC. Retrieved from

<https://www.pewresearch.org/methods/2015/05/13/from-telephone-to-the-web-the-challenge-of-mode-of-interview-effects-in-public-opinion-polls/>

Kennedy C. (2010), "Nonresponse and Measurement Error in Mobile Phone Surveys," unpublished PhD dissertation, University of Michigan-Ann Arbor.

Kennedy C., Blumenthal M., Clement S., Clinton J. D., Durand C., Franklin C., McGeeney K., Miringoff L., Olson K., Rivers D., Saad L., Witt G. E., Wlezin C. (2018), "An Evaluation of the 2016 Election Polls in the United States," *Public Opinion Quarterly*, 82(1), 1-33.

- Kennedy C., Everett S. E. (2011), "Use of Cognitive Shortcuts in Landline and Cell Phone Surveys," *Public Opinion Quarterly*, 75, 336-348.
- Kennedy C., Mercer A., Keeter S., Hatley N., McGeeney K., Gimenez A. (2016), *Evaluating Online Nonprobability Surveys*, Pew Research Center, Washington, DC, retrieved from <https://www.pewresearch.org/methods/2016/05/02/evaluating-online-nonprobability-surveys/>.
- Keusch F., Yan T. (2016), "Web versus Mobile Web: An Experimental Study of Device Effects and Self-Selection Effects," *Social Science Computer Review*, 35(6), 751-769.
- Kim Y., Dykema J., Stevenson J., Black P., Moberg D. P. (2018), "Straightlining: Overview of Measurement, Comparison of Indicators, and Effects in Mail-Web Mixed-Mode Surveys," *Social Science Computer Review*, 37(2), 214-233.
- Kirgis N. G., Lepkowski J. M. (2013), "Design and Management Strategies for Paradata-Driven Responsive Design: Illustrations from the 2006-2010 National Survey of Family Growth," in *Improving Surveys with Paradata: Analytic Uses of Process Information*, ed. F. Kreuter, pp. 123-144, Hoboken, NJ: John Wiley & Sons, Inc.
- Kish L. (1949), "A Procedure for Objective Respondent Selection Within the Household," *Journal of the American Statistical Association*, 57, 92-115.
- Kitada H. (2016), "Examining Trends in the Presence of Survey Mode Effects," Paper presented at the 71st Annual Conference of the American Association for Public Opinion Research, Austin, TX.
- Klausch T., Hox J. J., Schouten B. (2013), "Measurement Effects of Survey Mode on Equivalence of Attitudinal Rating Scale Questions," *Sociological Methods & Research*, 52(3), 227-263.

- Klausch T., Hox J. J., Schouten B. (2015), "Selection Error in Single- and Mixed Mode Surveys of the Dutch General Population," *Journal of the Royal Statistical Society*, 178(4), 945-961.
- Klausch T., Schouten B. (2018), "Evaluating and Reducing Biases in Mixed Mode Survey Data," Webinar presented as a part of the American Association for Public Opinion Research, retrieved from <https://www.aapor.org/Education-Resources/Online-Education/Webinar-Details.aspx?webinar=WEB1118>.
- Klausch T., Schouten B., Buelens B., van den Brakel J. (2017), "Adjusting Measurement Bias in Sequential Mixed-Mode Surveys Using Re-Interview Data," *Journal of Survey Statistics and Methodology*, 5(4), 409-432.
- Klausch T., Schouten B., Hox J. J. (2017), "Evaluating Bias of Sequential Mixed-Mode Designs Against Benchmark Surveys," *Sociological Methods & Research*, 46(3), 456-489.
- Kolenikov S., Kennedy C. (2014), "Evaluating Three Approaches to Statistically Adjust for Mode Effects," *Journal of Survey Statistics and Methodology*, 2(2), 126-158.
- Kraut, R. Olson, J., Banaji, M., Bruckman, A., Cohen, J., Couper, M. (2004), "Psychological Research Online: Report of Board of Scientific Affairs' Advisory Group on the Conduct of Research on the Internet," *American Psychologist*, 59(2), 105-117. doi:10.1037/0003-066X.59.2.105
- Krenn P. J., Titze S., Oja P., Jones A., Ogilvie D. (2011), "Use of Global Positioning Systems to Study Physical Activity and the Environment: A Systematic Review," *American Journal of Preventive Medicine*, 41(5), 508-515.
- Kreuter F. (2013), *Improving Surveys with Paradata*, Hoboken, NJ: John Wiley & Sons, Inc.

- Kreuter F., McCulloch S., Presser S., Tourangeau R. (2011), "The Effects of Asking Filter Questions in Interleaved versus Grouped Format," *Sociological Methods & Research*, 40(1), 88-104.
- Kreuter F., Olson K. (2013), "Paradata for Nonresponse Error Investigation," in *Improving Surveys with Paradata: Analytic Uses of Process Information*, ed. F. Kreuter, pp. 13-42, Hoboken, NJ: John Wiley & Sons, Inc.
- Kreuter F., Presser S., Tourangeau R. (2008), "Social Desirability Bias in CATI, IVR, and Web Surveys: The Effects of Mode and Question Sensitivity," *Public Opinion Quarterly*, 72(5), 847-865.
- Krosnick J. A. (1991), Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, 5(3), 213-236.
- Krosnick J. A., Alwin D. F. (1987), "An Evaluation of a Cognitive Theory of Response-Order Effects in Survey Measurement," *Public Opinion Quarterly*, 51, 201-219.
- Krug S. (2014), *Don't Make Me Think, Revisited. A Common Sense Approach to Web Usability* (3rd ed.), Berkeley, CA: New Riders.
- Krysan M., Couper M. P., (2003), "Race in the Live and the Virtual Interview: Racial Deference, Social Desirability, and Activation Effects in Attitude Surveys," *Social Psychology Quarterly*, 66, 364-383.
- Krysan M., Schuman H., Scott L. J., Beatty P. (1994), "Response Rates and Response Content in Mail versus Face-to-Face Surveys," *Public Opinion Quarterly*, 58, 381-399.
- Krzyzanowski M., Qin Y., Robinson J., Sikes N. (2018), "Complex Use of Voxco, Commercial-Off-the-Shelf Software, for Data Collection," Paper presented at the Federal Computer Assisted Survey Information Collection (FedCASIC) Workshops, Suitland, MD.

- Laflamme F., Maydan M., Miller A. (2008), "Using Paradata to Actively Manage Data Collection Survey Process," Paper presented at the 2008 JSM Proceedings, Section on Survey Research Methods, American Statistical Association, Alexandria, VA.
- Lai J. W., Vanno L., Link M., Pearson J., Makowska H., Benezra K., Green M. (2010), "Life360: Usability of Mobile Devices for Time Use Surveys," *Survey Practice*, 3(1), 1-6.
- Lambert A. D., Miller A. L. (2015), "Living with Smartphones: Does Completion Device Affect Survey Responses?," *Research in Higher Education*, 56, 166-177.
- Lau A., Kennedy C. (2019), *When Online Survey Respondents Only 'Select Some That Apply'*, Pew Research Center, Washington, DC, retrieved from <https://www.pewresearch.org/methods/2019/05/09/when-online-survey-respondents-only-select-some-that-apply/>.
- Lavrakas P. J., Benson G., Blumberg S., Buskirk T., Cervantes I. F., Christian L., Dutwin D., Fahimi M., Fienberg H., Guterbock T., Keeter S., Kelly J., Kennedy C., Peytchev A., Piekarski L., Shuttles C. (2017), "Report From the AAPOR Task Force on: The Future of U.S. General Population Telephone Survey Research," retrieved 5/16/2019 from <https://www.aapor.org/Education-Resources/Reports/The-Future-Of-U-S-General-Population-Telephone-Sur.aspx>.
- LeClere F. B., Vanicek J. S., Xia K., Amaya A. E., Murphy W. E., Fiorio L., Carris K. L. (2012), "Changing Survey Modes: Does it Matter How You Get There," Proceedings of the Survey Research Methods Section, American Statistical Association, pp. 5344-5354.
- Lepkowski J. M., Mosher W. D., Groves R. M., West B. T., Wagner J., Gu H. (2013), "Responsive Design, Weighting, and Variance Estimation in the 2006-2010 National

- Survey of Family Growth, National Center for Health Statistics,” *Vital and Health Statistics*, 2(158), 1-52.
- Lepkowski J. M., Sandosky S. A., Couper M. P., Chardoul S., Carn L., Scott L. J. (1995), “A Comparison of Recording Errors Between CATI and Paper-and-Pencil Data Collection Methods,” *Proceedings of the American Statistical Association, Survey Research Methods Section*, 521-526.
- Lepkowski J.M., Groves R.M.. 1986. "A Mean Squared Error Model for Dual Frame, Mixed Mode Survey Design." *Journal of the American Statistical Association*, 81(396): 930-937.
- Groves R. M., Magilavy L. J. (1986), “Measuring and Explaining Interviewer Effects in Centralized Telephone Surveys,” *Public Opinion Quarterly*, 50, 251-266.
- Lesser V. M., Newton L. D., Yang D. K., Sifneos J. C. (2016), “Mixed-Mode Surveys Compared with Single Mode Surveys: Trends in Responses and Methods to Improve Completion,” *Journal of Rural Social Sciences*, 31(3), 7-34.
- Lesser V., Nawrocki K., Newton L. (2017), “Improving Response in Multimode and Single Mode Probability Based Surveys Compared to a Non-Probability Survey,” Paper presented at the 2017 conference of the European Survey Research Association, Lisbon, Portugal.
- Lesser V., Newton L., Yang D. (2012), “Comparing Item Nonresponse Across Different Delivery Modes in General Population Surveys,” *Survey Practice*, 5(2), 1-5.
- Liao D., Biemer P. P., Mullan Harris K., Burke B. J., Halpern C. T. (2019), “Transitioning from In-Person Mode to Web-Mail Mixed Mode in a Panel Survey,” Paper presented at the 74th Annual Conference of the American Association for Public Opinion Research, Toronto, Ontario, Canada.

- Lien R. (2015), "Adding a Web Mode to Phone Surveys: Effectiveness Cost Implications," Paper presented at the 70th Annual Conference of the American Association for Public Opinion Research, Hollywood, FL.
- Link M. (2013), "Measuring Compliance in Mobile Longitudinal Repeated-Measures Design Study," *Survey Practice*, 6(4), 1-11.
- Link M. W., Battaglia M. P., Frankel M. R., Osborn L., Mokdad A. H. (2008), "A Comparison of Address-Based Sampling (ABS) Versus Random-Digit Dialing (RDD) for General Population Surveys," *Public Opinion Quarterly*, 72(1), 6-27.
- Link M. W., Burks A. T. (2013), "Leveraging Auxiliary Data, Differential Incentives, and Survey Mode to Target Hard-to-Reach Groups in an Address-Based Sample Design," *Public Opinion Quarterly*, 77(3), 696-713.
- Link M., Battaglia M., Frankel M., Osborn L., Mokdad A. (2006), "Address-Based Versus Random-Digit Dialed Surveys: Comparison of Key Health and Risk Indicators," *American Journal of Epidemiology*, 164, 1019-1025.
- Link M., Murphy J., Schober M. F., Buskirk T. D., Hunter Childs J., Langer Tesfaye C. (2014), "Mobile Technologies for Conducting, Augmenting, and Potentially Replacing Surveys: Report of the AAPOR Task Force on Emerging Technologies in Public Opinion Research," retrieved 12/18/2018 from <https://www.aapor.org/Education-Resources/Reports.aspx>.
- Liu M. (2018). "Data Collection Mode Effect on Abortion Questions: A Comparison of Face-To-Face and Web Surveys." *Gender and Women's Studies*. 1(1):2.
- Liu M., Cernat A. (2018), "Item-by-item Versus Matrix Questions: A Web Survey Experiment," *Social Science Computer Review*, 36(6), 690-706.

- Liu M., Wang Y. (2014), "Data Collection Mode Effects on Political Knowledge," *Survey Methods: Insights from the Field*. Retrieved from <http://surveyinsights.org/?p=5317>
- Liu M., Wang Y. (2015), "Data Collection Mode Effect on Feeling Thermometer Questions: A Comparison of Face-to-Face and Web Surveys," *Computers in Human Behavior*, 48, 212-218.
- Lohr S., Brick M. (2014), "Allocation for Dual Frame Telephone Surveys with Nonresponse," *Journal of Survey Statistics and Methodology*, 2(4), 388-409.
- Lozar Manfreda K., Bosnjak M., Berzelak J., Haas I., Vehovar V. (2008), "Web Surveys versus Other Survey Modes: A Meta-Analysis Comparing Response Rates," *International Journal of Market Research*, 50(1), 79-104.
- Lugtig P., Jackle A. (2014), "Can I Just Check...? Effects of Edit Check Questions on Measurement Error and Survey Estimates," *Journal of Official Statistics*, 30(1), 45-62.
- Lugtig P., Lensvelt-Mulders G. J. L. M., Frerichs R., Greven A. (2011), "Estimating Nonresponse Bias and Mode Effects in a Mixed-Mode Survey," *International Journal of Market Research*, 53(5), 669-686.
- Lykes V., Meyers J. (2017), "Got Mail? Drivers of Mail-to-Online Response Rates," Paper presented at the 72nd Annual Conference of the American Association for Public Opinion Research, New Orleans, LA.
- Lykke L. C., and Garcia Trejo Y.A. (2018), Results from the Usability Testing of the Spanish Language Version of the 2020 Census Barriers, Attitudes and Motivators Survey (CBAMS), Research and Methodology Directorate, Center for Behavioral Science Methods Research Report Series (Survey Methodology #2018-19), U.S. Census Bureau.

Available online at <http://www.census.gov/content/dam/Census/library/working-papers/2018/adrm/rsm2018-19.pdf>

Lynn P., Hope S., Jackle A., Campanelli P., Nicolaas G. (2011), “The Role of Visual and Aural Communication in Producing Mode Effects on Answers in Survey Questions,” Institute for Social and Economic Research Working Paper Series #20141-20, University of Essex.

Lynn P., Kaminska O. (2012), “The Impact of Mobile Phones on Survey Measurement Error,” *Public Opinion Quarterly*, 77, 586-605.

MacInnis B., Krosnick J. A., Ho A. S., Cho M. (2018), “The Accuracy of Measurements with Probability and Nonprobability Survey Samples: Replication and Extension,” *Public Opinion Quarterly*, 82(4), 707-744.

Marine Recreational Information Program (2015), *Transition Plan for the Fishing Effort Survey*, Silver Spring, MD: U.S. Department of Commerce, National Oceanic and Atmospheric Administration.

Marken S. (2018), “Still Listening: The State of Telephone Surveys,” Gallup Methodology Blog. Retrieved from <https://news.gallup.com/opinion/methodology/225143/listening-state-telephone-surveys.aspx>.

Marken S., Auter Z., Marlar J. (2018), “Mail or web first that is our question: A comparison of multi- and sequential mode offerings.” Paper presented at the Annual meeting of the American Association for Public Opinion Research, Denver, CO.

Marken S., Marlar J., Buckles G. (2018), “Mail or Web First That is Our Question: A Comparison of Multi- and Sequential Mode Offerings,” Paper presented at the 73rd

Annual Conference of the American Association for Public Opinion Research, Denver, CO.

Marlar J., Chattopadhyay M., Ander S., Kanitkar K., Andrews R., Foster J., Kitts-Jensen R. (2017), “Leveraging ABS to Conduct a Mixed-Mode, Multiphase Survey,” Paper presented at the 72nd Annual Conference of the American Association for Public Opinion Research, New Orleans, LA.

Marlar J., Chattopadhyay M., Jones J., Marken S., Kreuter F. (2018), “Within-Household Selection and Dual-Frame Telephone Surveys: A Comparative Experiment of Eleven Different Selection Methods,” *Survey Practice*, 11(2), 1-26.

Martinez W. L. (2018), “U.S. Federal Committee on Statistical Methodology Working Group on Transparent Quality Reporting of Editing and Imputation when Integrating Data from Multiple Data Sources,” Paper presented at the United Nations Economic Commission for Europe (UNECE) Workshop on Statistical Data Editing, Neuchatel, Switzerland.

Martinez, M., Eggleston, C.M., Katz, J.M., Morales, G.D. (2018), *Cognitive Pretesting of the 2017 American Community Survey Mail Design Test: Full Redesign Materials*. Research and Methodology Directorate, Center for Behavioral Science Methods Research Report Series (Survey Methodology #2018-18), U.S. Census Bureau. Available online at <http://www.census.gov/content/dam/Census/library/working-papers/2018/adrm/rsm2018-18.pdf>

Mathews M., Parast L., Tolpadi A., Elliott M., Flow-Delwiche E., Becker K. (2017), “Emergency Department Patient Experience of Care Survey in the Discharged to Community Setting - A Randomized Feasibility Study,” Paper presented at the 72nd

- Annual Conference of the American Association for Public Opinion Research, New Orleans, LA.
- Mathews, K., Phelan, J., Jones, N. A., Konya, S., Marks, R., Pratt, B. M., Coombs, J., Bentley, M. (2017a), *2015 National Content Test Race and Ethnicity Analysis Report*. Washington, DC: U.S. Census Bureau.
- Mathiowetz N. A., Brick J. M., Stokes L., Andrews R. A., Muzzy S. (2010), "A Pilot Test of a Dual Frame Mail Survey as an Alternative to an RDD Survey," Paper presented at the Joint Statistical Meetings, Vancouver, B. C.
- Mathiowetz N., McGonagle K. (2000), "An Assessment of the Current State of Dependent Interviewing in Household Surveys," *Journal of Official Statistics*, 16(4), 401-418.
- Mauz E., von der Lippe E., Allen J., Schilling R., Muters S., Hoebel J., Schmich P., Wetzstein M., Kamtsiuris P., Lange C. (2018), "Mixing Modes in a Population-Based Interview Survey: Comparison of a Sequential and a Concurrent Mixed-Mode Design for Public Health Research," *Archives of Public Health*, 76(8), 1-17.
- Mavletova A. (2013), "Data Quality in PC and Mobile Web Surveys," *Social Science Computer Review*, 31(6), 725-743.
- Mavletova A., Couper M. P. (2013), "Sensitive Topics in PC Web and Mobile Web Surveys: Is There a Difference," *Survey Research Methods*, 7(3), 191-205.
- Mavletova A., Couper M. P. (2015), "A Meta-Analysis of Breakoff Rates in Mobile Web Surveys," in *Mobile Research Methods: Opportunities and Challenges of Mobile Research Methodologies*, eds. Toninelli R., Pinter R., de Padraza P., London: Ubiquity Press.

- Mavletova A., Couper M. P. (2016), "Grouping of Items in Mobile Web Questionnaires," *Field Methods*, 28(2), 170-193.
- Mavletova A., Couper M. P., Lebedev D. (2018), Grid and Item-by-Item Formats in PC and Mobile Web Surveys. *Social Science Computer Review*, 36(6), 647-668.
doi:10.1177/0894439317735307
- Mavoa S., Oliver M., Witten K., Badland H. M. (2011), "Linking GPS and Travel Diary Data Using Sequence Alignment in a Study of Children's Independent Mobility," *International Journal of Health Geographics*, 10(1), 1-10.
- Mayfield A., Frasier A., Vanicek J., Li Y., English N., Greene J., Leidy M. (2015), "Knowing When to Stop: Evaluating First 5 LA Family Survey Data Based on Data Collection Mode and Difficulty to Complete an Interview," Paper presented at the 70th Annual Conference of the American Association for Public Opinion Research, Hollywood, FL.
- McClain C., Crawford S. D. (2013), "Grid Formats, Data Quality, and Mobile Device Use: Toward a Questionnaire Design Approach," Paper presented at the 68th Annual Conference of the American Association for Public Opinion Research, Boston, MA.
- McClinton Apollis T., Lund C., de Vries P. J., Mathews C. (2015), "Adolescents' and Adults' Experiences of Being Surveyed About Violence and Abuse: A Systematic Review of Harms, Benefits and Regrets," *American Journal of Public Health*, 105(2), e31-e45.
- McGeeney K., Yan H. Y. (2016), *Text Message Notification for Web Surveys: Sending Texts to Survey Panel Members Shortens Response Time*, Pew Research Center, Washington, DC.
Retrieved from <https://www.pewresearch.org/methods/2016/09/07/text-message-notification-for-web-surveys/>

- McGeeney, K., Kennedy, C. (2016), "Cellphone Activity Flags: A Trade-off between Efficiency and Coverage." Pew Research Center. Retrieved from <https://www.pewresearch.org/methods/2016/10/24/cellphone-activity-flags/>.
- McGonagle K. A., Freedman V., Griffin J., Dascola M. (2017), "Web Development in the PSID: Transition and Testing of a Web Version of the 2015 PSID Telephone Instrument," Technical Series Paper #17-02, Institute for Social Research, University of Michigan.
- McGonagle K. A., Schoeni R. F., Sastry N., Freedman V. A. (2012), "The Panel Study of Income Dynamics: Overview, Recent Innovations, and Potential for Life Course Research," *Longitudinal and Life Course Studies*, 3(2), 268-284.
- McMaster H. S., Stander V. A., Williams C. S., Woodall K. A., O'Malley C. A., Bauer L. M., Davila E. P. (2018), "Engaging Military Couples in Marital Research: Does Requesting Referrals from Service Members to Recruit Their Spouses Introduce Sample Bias?," *BMC Medical Research Methodology*, 18, 1-13.
- McPhee C. (2012), "Making the Money Count: Maximizing the Utility of Incentives in a Two-Stage Mail Survey," Paper presented at the 67th Annual Conference of the American Association for Public Opinion Research, Orlando, FL.
- McPhee C., Bielick S., Masterton M., Flores L., Parmer R., Amchin S., Stern S., McGowan H. (2015), *National Household Education Surveys Program of 2012: Data File User's Manual* (NCES 2015-030), National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education, Washington, DC.
- McPhee C., Jackson M., Bielick S., Masterson M., Battle D., McQuiggan M., Payri M., Cox C., Medway R. (2018), *National Household Education Surveys Program of 2016: Data File*

- User's Manual* (NCES 2018-100), National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education, Washington, DC.
- McPhee C., Masterton M. (2015), "To Re-Mail or Not to Re-Mail: Evaluating Occupancy Status in an Address-Based Household Mail Survey," Paper presented at the 70th Annual Conference of the American Association for Public Opinion Research, Hollywood, FL.
- McPhee C., Zukerberg A. (2018), "To Re-Contact or Not to Re-Contact: Using Auxiliary Data to Model Address Eligibility in a Household Survey," Paper presented at the 73rd Annual Conference of the American Association for Public Opinion Research, Denver, CO.
- McQuiggan M., Medway R., Zhang M., Megra M. (2015), "Prepaid Incentives in ABS Surveys: Effect on Nonresponse and Measurement Errors," Paper presented at the 2015 International Total Survey Error Conference, Baltimore, MD.
- Medway R. L., Fulton J. (2012), "When More Gets You Less: A Meta-Analysis of the Effect of Concurrent Web Options on Mail Survey Response Rates," *Public Opinion Quarterly*, 76(4), 733-746.
- Medway R., Battle L. (2018), *Administering a Single-Phase, All-Adults Mail Survey: A Methodological Evaluation of the 2013 NATES Pilot Study* (NCES 2018-121), National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education, Washington, DC.
- Mercer A., Caporaso A., Cantor D., Townsend R. (2015), How Much Gets You How Much? Monetary Incentives and Response Rates in Household Surveys. *Public Opinion Quarterly*, 79(1), 105-129.
- Mercer A., Lau A., Kennedy C. (2018), *For Weighting Online Opt-In Samples, What Matters Most?* Pew Research Center, Washington, DC. Retrieved from

<https://www.pewresearch.org/methods/2018/01/26/for-weighting-online-opt-in-samples-what-matters-most/>.

Messer B., Dillman D. A. (2011), "Surveying the General Public Over the Internet Using Addressed Based Sampling and Mail Contact Procedures," *Public Opinion Quarterly*, 75(3), 429-457.

Messer B., Edwards M., Dillman D. (2012), "Determinants of Item Nonresponse to Web and Mail Respondents in Three Address-Based Mixed-Mode Surveys of the General Public," *Survey Practice*, 5(2), 1-9.

Millar M. M., Schmuhl P., Page K., Genovesi A. L., Ely M., Hemingway C., Olson L. M. (2018), "Improving Response to an Establishment Survey Through the use of Web-Push Data Collection Methods," *Mathematical Population Studies*, 25(3), 168-179.

Millar M., Dillman D. (2012), "Do Mail and Internet Surveys Produce Different Item Nonresponse Rates? An Experiment Using Random Mode Assignment," *Survey Practice*, 5(2), 1-7.

Mitchell S., Stroble M., Fahrney K., Nguyen M., Bibb B., Thissen M. R., Stephenson W. (2008), "Using Computer Audio-Recorded Interviewing to Assess Interviewer Coding Error," Proceedings of the American Statistical Association, Survey Research Methods Section, 4414-4421.

Montaquila J. M., Brick J. M., Kim K. (2012), "Methodological Findings from a Two-Phase Address Based Sample Fielded by Mail," Paper presented at the 67th Annual Conference of the American Association for Public Opinion Research, Orlando, FL.

- Montaquila J. M., Brick J. M., Williams D., Kim K., Han D. (2013), “A Study of Two-Phase Mail Survey Data Collection Methods,” *Journal of Survey Statistics and Methodology*, 2013(1), 66-87.
- Murphy J., Biemer P., Berry C. (2018), “Transitioning a Survey to Self-Administration using Adaptive, Responsive, and Tailored (ART) Design Principles and Data Visualization,” *Journal of Official Statistics*, 34(3), 625-648.
- Murphy J., Link M. W., Hunter Childs J., Langer Tesfaye C., Dean E., Stern M., Pasek J., Cohen J., Callegaro M., Harwood P. (2014), *Social Media in Public Opinion Research: Report of the AAPOR Task Force on Emerging Technologies in Public Opinion Research*. American Association of Public Opinion Research, Retrieved from https://www.aapor.org/AAPOR_Main/media/MainSiteFiles/AAPOR_Social_Media_Report_FNL.pdf.
- Murphy J., Mayclin D., Richards A., Roe D. (2015), “A Multi-Method Approach to Survey Pretesting,” Paper presented at the Federal Committee on Statistical Methodology, retrieved from https://nces.ed.gov/fcsm/pdf/D3_Murphy_2015FCSM.pdf.
- Murphy W., Harter R., Xia K. (2010), “Design and Operational Changes for the REACH U.S. Risk Factor Survey,” Proceedings of the American Statistical Association, Survey Research Methods Section, Vancouver, British Columbia.
- Murphy, J. J., Duprey, M. A., Chew, R. F., Biemer, P. P., Harris, K. M., and Halpern, C. T. (2019), *Interactive Visualization to Facilitate Monitoring Longitudinal Survey Data and Paradata*. RTI Press Publication No. OP-0061-1905. Research Triangle Park, NC: RTI Press. <https://doi.org/10.3768/rtipress.2019.op.0061.1905>

- National Academies of Sciences, Engineering, and Medicine (2017a), *Federal Statistics, Multiple Data Sources, and Privacy Protection: Next Steps*, Washington, DC: The National Academies Press.
- National Academies of Sciences, Engineering, and Medicine (2017b), *Innovations in Federal Statistics: Combining Data Sources While Protecting Privacy*, Washington, DC: The National Academies Press.
- National Academies of Sciences, Engineering, and Medicine (2018), *Measuring the 21st Century Science and Engineering Workforce Population: Evolving Needs*, Washington, DC: The National Academies Press.
- National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research (1979), *The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research*, Bethesda, MD: The Commission.
- National Comorbidity Survey (n.d.), “National Comorbidity Survey (NCS),” Harvard Medical School. retrieved from <https://www.hcp.med.harvard.edu/ncs/>.
- National Oceanic and Atmospheric Administration (NOAA) (n.d.), “Our Free eVTR App is Available for Online Reporting,” retrieved from https://www.greateratlantic.fisheries.noaa.gov/mediacenter/2018/02/28_our_free_evtr_app_is_available_for_online_reporting.html.
- Newman E., Risch E., Kassam-Adams N. (2006), “Ethical Issues in Trauma-Related Research: A Review,” *Journal of Empirical Research on Human Research Ethics*, 1(3), 29-46.
- Nicolaas G., Tipping S. (2006), “Mode Effects in Social Capital Surveys,” *National Statistics Survey Methodology Bulletin*, 58, 59-74.

- Office of Management and Budget (OMB) (2006), *Standards and Guidelines for Statistical Surveys*, Washington, DC, retrieved from https://unstats.un.org/unsd/dnss/docs-nqaf/USA_standards_stat_surveys.pdf.
- Office of Management and Budget (OMB) (2017), “2017 National Survey of College Graduates,” Washington, DC, retrieved from https://www.reginfo.gov/public/do/PRAViewIC?ref_nbr=201612-3145-001icID=204164.
- Olmsted-Hawala E., Nichols E., Myers M. (2018), “Iterative Usability Testing on Surveys that Span Years.” Paper presented at the Annual meeting of the American Association for Public Opinion Research, Denver, CO.
- Olson K. (2010), “An Examination of Questionnaire Evaluation by Expert Reviewers,” *Field Methods*, 22(4), 295-318.
- Olson K., Parkhurst B. (2013), “Collecting Paradata for Measurement Error Evaluations.” In F. Kreuter (Ed.), *Improving Surveys with Paradata: Analytic Uses of Process Information* (pp. 43-72), Hoboken, NJ: John Wiley & Sons.
- Olson K., Peytchev A. (2007), “Effect of Interviewer Experience on Interview Pace and Interviewer Attitudes,” *Public Opinion Quarterly*, 71(2), 273-286.
- Olson K., Smyth J. D. (2014), “Accuracy of Within-Household Selection in Web and Mail Surveys of the General Population,” *Field Methods*, 26, 56-59.
- Olson K., Smyth J. D. (2017), “Within-Household Selection in Mail Surveys: Explicit Questions are Better than Cover Letter Instruction,” *Public Opinion Quarterly*, 81(3), 688-713.

- Olson K., Smyth J. D., Ganshert A. (2019), "The Effects of Respondent and Question Characteristics on Respondent Answering Behaviors in Telephone Interviews," *Journal of Survey Statistics and Methodology*, 7(2), 275-308.
- Olson K., Smyth J. D., Phillips A. (2018), "Grids vs. Item-by-Item Design and Visual Design for a Mixed Mode Web Push Survey on Nondifferentiation," Paper presented at the 73rd Annual Meeting of the Midwest Association for Public Opinion Research conference, Chicago, IL.
- Olson K., Stange M., Smyth J. D. (2014), "Assessing Within-Household Selection Methods in Household Mail Surveys," *Public Opinion Quarterly*, 78(3), 656-678.
- Olson K., Wagner J., Anderson R. (2018), "Survey Costs: A Typology and Evaluative Criteria," PORTAL session presented at the American Association for Public Opinion Research annual meeting, Denver, CO.
- Olson, K., Smyth, J.D., Phillips, A., and Stenger, R. 2019. Four Questionnaire Experiments in Mixed-Mode, Mixed-Device Surveys: Answer Boxes, Response Option Order, Check-All versus Forced-Choice, and Ordinal Scale versus Number Box Items. Paper presented at the American Association for Public Opinion Research annual meeting, Toronto, Ontario, Canada.
- Oudejans M., Christian L. M. (2011), "Using Interactive Features to Motivate and Probe Responses to Open-Ended Questions," in *Social and Behavioral Research and the Internet: Advances in Applied Methods and Research Strategies*, eds. Das M., Ester P., Kaczmirek L., pp. 215-244, New York, NY: Routledge.

- Parast L., Elliott M. N., Hambarsoomian K., Teno J., Anhang Price R. (2018), "Effects of Survey Mode on Consumer Assessment of Healthcare Providers and Systems (CAHPS) Hospice Survey Scores," *Journal of American Geriatric Society*, 66, 546-552.
- Pasek J., Jang S. M., Cobb III C. L., Dennis J. M., Disogra C. (2014), "Can Marketing Data Aid Survey Research? Examining Accuracy and Completeness in Consumer-File Data," *Public Opinion Quarterly*, 78(4), 889-916.
- Patrick M. E., Couper M. P., Laetz V. B., Schulenberg J. E., O'Malley P. M., Johnston L. D., Miech, R. A. (2018), "A Sequential Mixed-Mode Experiment in the US National Monitoring the Future Study," *Journal of Survey Statistics and Methodology*, 6(1), 72-97.
- Penn State Harrisburg Center for Survey Research (2019a), *Benefits of the Lion Poll Methodology*, Penn State Harrisburg Center for Survey Research, Harrisburg, PA, retrieved from <https://csr.hbg.psu.edu/Lion-Poll/Benefits-of-the-Lion-Poll-Methodology>.
- Penn State Harrisburg Center for Survey Research. (2019b), *Lion Poll Survey: Benefits, Challenges, and Study Methods*. Penn State Harrisburg Center for Survey Research, Harrisburg, PA, Retrieved from https://csr.hbg.psu.edu/Portals/44/Lion%20Poll_Benefits_Challenges_Study%20Methods_May2019_1.pdf
- Peterson G., Mechling J., LaFrance J., Swinehart J., Ham G. (2013), "Solving the Unintentional Mobile Challenge," Paper presented at the CASRO Online Research Conference, San Francisco, CA.
- Pew Research Center (2006), *The Cell Phone Challenge to Survey Research*, Pew Research Center, Washington, DC, retrieved from <http://www.people-press.org/2006/05/15/the-cell-phone-challenge-to-survey-research/>.

Pew Research Center (2015), “Advances in Telephone Survey Sampling: Balancing Efficiency and Coverage Using Several New Approaches,” Pew Research Center, Washington, DC, retrieved from <https://www.pewresearch.org/methods/2015/11/18/advances-in-telephone-survey-sampling/>.

Pew Research Center (2015), “From Telephone to the Web: The Challenge of Mode of Interview Effects in Public Opinion Polls,” Pew Research Center, Washington, DC, retrieved from <http://www.pewresearch.org/methods/2015/05/13/from-telephone-to-the-web-the-challenge-of-mode-of-interview-effects-in-public-opinion-polls/>.

Pew Research Center (2017), “Are Telephone Polls Underestimating Support for Trump?,” Pew Research Center, Washington, DC, retrieved from <http://www.pewresearch.org/methods/2017/03/31/are-telephone-polls-understating-support-for-trump/>.

Pew Research Center (2019), “Internet/Broadband Fact Sheet,” Pew Research Center, Washington, DC, retrieved from <https://www.pewinternet.org/fact-sheet/internet-broadband/>.

Pew Research Center (2019), “The American Trends Panel Survey Methodology,” Pew Research Center, Washington, DC, retrieved from <https://www.pewresearch.org/methods/u-s-survey-research/american-trends-panel/>.

Peytchev A. (2007), “Participation Decisions and Measurement Error in Web Surveys,” unpublished doctoral dissertation, University of Michigan, Ann Arbor, MI.

Peytchev A., (2012), “Multiple Imputation for Unit Nonresponse and Measurement Error,” *Public Opinion Quarterly*, 76(2), 214–237.

- Peytchev A., Conrad F. G., Couper M. P., Tourangeau R. (2010), Increasing Respondents' Use of Definitions in Web Surveys. *Journal of Official Statistics*, 26(4), 633-650.
- Peytchev A., Hill C. A. (2010), "Experiments in Mobile Web Survey Design: Similarities to Other Modes and Unique Considerations," *Social Science Computer Review*, 28(3), 319-335.
- Peytchev A., Ridenhour J., Krotki K. (2010), "Differences Between RDD Telephone and ABS Mail Survey Design: Coverage, Unit Nonresponse, and Measurement Error," *Journal of Health Communication: International Perspectives*, 15, 117-134.
- Phelan J. (2016), *2020 Research and Testing: 2015 National Content Test Optimizing Self-Response Report*, Washington, DC: US Census Bureau.
- Pierzchala, M., Wright, D., Wilson, C., Guerino, P. (2004), *Instrument Design for a Blaise Multimode Web, CATI, and Paper Survey*. Paper presented at the Blaise Users Conference, Gatineau, Québec. <http://www.blaiseusers.org/2004/papers/24.pdf>
- Poehler E., Barth D. (2017), "Influencing Respondent Mode Choices in the American Community Survey," Paper presented at the 72nd Annual Conference of the American Association for Public Opinion Research, New Orleans, LA.
- Polivka A. E., Miller S. M. (1998), The CPS after the Redesign: Refocusing the Economic Lens. In J. Haltiwanger, M. E. Manser, R. Topel (Eds.), *Labor Statistics Measurement Issues* pp. 249-289, Chicago, IL: University of Chicago Press.
- Powers J. R., Mishra G., Young A. F. (2005) Differences in mail and telephone responses to self-rated health: use of multiple imputation in correcting for response bias. *Australian and New Zealand Journal of Public Health*, 29(2): 149-154.

- Preisendorfer P., Wolter F. (2014), "Who is Telling the Truth? A Validation Study on Determinants of Response Behavior in Surveys," *Public Opinion Quarterly*, 78(1), 126-146.
- Presser S., Rothgeb J. M., Couper M. P., Lessler J. T., Martin E., Martin J., Singer E. (2004). *Methods for Testing and Evaluating Survey Questionnaires*. Hoboken, NJ: Wiley.
- QuestionPro (n.d.), "Question Pro: Survey software that gets the job done," retrieved from <https://www.questionpro.com>.
- Rasinski K. A., Mingay D., Bradburn N. M. (1994), "Do Respondents Really 'Mark All That Apply' on Self-Administered Questions?," *Public Opinion Quarterly*, 58(3), 400-408.
- Redford J., Hastedt S. (2011), "Who Missed the Skips? Empirical Results from a Self-Administered Survey," Paper presented at the 66th Annual Conference of the American Association for Public Opinion Research, Phoenix, AZ.
- Redline C. (2011), *Clarifying Survey Questions*, unpublished dissertation, the Joint Program in Survey Methodology, University of Maryland, College Park.
- Redline C. (2013), "Clarifying Categorical Concepts in a Web Survey," *Public Opinion Quarterly*, 77(S1), 89-105.
- Redline C., Dillman D. A., Dajani A. N., Scaggs M. A. (2003), "Improving Navigational Performance in US Census 2000 by Altering the Visually Administered Languages of Branching Instructions," *Journal of Official Statistics*, 19, 403-420.
- Redman J., Thompson S., Yost B., Everts K. (2017), "Results of a Multi-Mode Design on Pre-Election Surveys," Paper presented at the 72nd Annual Conference of the American Association for Public Opinion Research, New Orleans, LA.

- Reich J., Yates W., Woolson R. (1986), "Kish Method for Mail Survey Respondent Selection," *American Journal of Public Health*, 76, 206.
- Reist B. (2014), "Early Experience of Adaptive Design Work in the NSCG," Paper presented at the Federal Economics Statistics Advisory Committee Meeting, US Census Bureau, June 2014.
- Residential Energy Consumption Survey (n.d.), US Energy Information Administration, Retrieved from https://www.eia.gov/survey/form/eia_457/2015_EIA-475A_paper.pdf.
- Residential Energy Consumption Survey (RECS) (2013), "2009 Technical Documentation-Summary," Washington, DC: US Department of Energy.
- Revilla M., Couper M. P. (2017), "Comparing Grids with Vertical and Horizontal Item-by-Item Formats for PCs and Smartphones," *Social Science Computer Review*, 36(3), 349-368.
- Revilla M., Ochoa C. (2016), "Open Narrative Questions in PC and Smartphones: Is the Device Playing a Role?," *Quality Quantity*, 50(6), 2495-2513.
- Richards A., Powell R., Murphy J., Nguyen M., Yu S. (2016), "Gridlocked: The Impact of Adapting Survey Grids for Smartphones," *Survey Practice*, 9(3), 1-15.
- Rizzo L., Brick J., Park I. (2004), "A Minimally Intrusive Method for Sampling Persons in Random-Digit Dial Surveys," *Public Opinion Quarterly*, 68(2), 267-274.
- Romuald K. S., Haggard L. M. (1994), "The Effect of Varying the Respondent Selection Script on Respondent Self-Selection in RDD Telephone Surveys," American Statistical Association Proceedings, Survey Research Methods Section, 1299-1304.
- Rothhaas C. A., Bently M., Hill J. M., Lestina F. (2011), "2010 Census: Bilingual Questionnaire Assessment Report," US Census Bureau, retrieved from

<https://www2.census.gov/programs-surveys/decennial/2010/program-management/5-review/cpex/2010-memo-156.pdf>

- Russ D. E., Ho K., Colt J. S., Armenti K. R., Baris D., Chow W., Davis F., Johnson A., Purdue M. P., Karagas M. R., Schwartz K., Schwenn M., Silverman D. T., Johnson C. A., Friesen M. C. (2016), "Computer-Based Coding of Free-Text Job Descriptions to Efficiently Identify Occupations in Epidemiological Studies," *Occupational and Environmental Medicine*, 73(6), 417-424.
- Rustemeyer A. (1997), "Measuring Interviewer Performance in Mock Interviews," Proceedings of the American Statistical Association, Social Statistics Section, 341-346.
- Rylander-Rudqvist T., Hakansson N., Tybring G., Wolk A. (2006), "Quality and Quantity of Saliva DNA Obtained from the Self-Administrated Orangene Method - A Pilot Study on the Cohort of Swedish Men," *Cancer Epidemiology, Biomarkers Prevention*, 15(9), 1742-1745.
- Sakshaug J. W., Cernat A., Raghunathan T. E. (2019), "Do Sequential Mixed-Mode Surveys Decrease Nonresponse Bias, Measurement Error Bias, and Total Bias? An Experimental Study," *Journal of Survey Statistics and Methodology*, published advanced access online. <https://doi.org/10.1093/jssam/smy024>
- Sakshaug J. W., Couper M. P., Ofstedal M. B. (2010), "Characteristics of Physical Measurement Consent in a Population-Based Survey of Older Adults," *Medical Care*, 48(1), 64-71.
- Sakshaug J. W., Hulle S., Schmucker A., Leibig S. (2017), "Exploring the Effects of Interviewer- and Self-Administered Survey Modes on Record Linkage Consent Rates and Bias," *Survey Research Methods*, 11(2), 171-188.

- Sakshaug J. W., Ofstedal M. B., Guyer H., Beebe T. J. (2015), "The Collection of Biospecimens in Health Surveys," in *Handbook of Health Survey Methods*, 1st ed., eds. Johnson T. P., pp. 383-419, Hoboken, NJ: John Wiley & Sons, Inc.
- Sala E., Lynn P. (2009), "The Potential of a Multi-Mode Data Collection Design to Reduce Non Response Bias. The Case of a Survey of Employers," *Quality & Quantity*, 43(1), 123-136.
- Saris W. E., Gallhofer I. (2007a), *Design, Evaluation, and Analysis of Questionnaires for Survey Research*, Hoboken, NJ: John Wiley & Sons, Inc.
- Saris W. E., Gallhofer I. (2007b), "Estimation of the Effects of Measurement Characteristics on the Quality of Survey Questions," *Survey Research Methods*, 1(1), 29-43.
- Sawyer S., Dillman D. A. (2002), *How Graphical, Numerical, and Verbal Languages Affect the Completion of the Gallup Q-12 on Self-Administered Questionnaires: Results from 22 Cognitive Interviews and Field Experiment (Technical Report No. 02-26)*, Pullman, WA: Washington State University, Social and Economic Sciences Research Center.
- Schaeffer N. C., Dykema J., Maynard D. W. (2010), "Interviewers and Interviewing," in *Handbook of Survey Research* (2nd ed.), eds. Wright J. D., Marsden P. V., pp. 437-470, Bingley, UK: Emerald Group Publishing Limited.
- Schlosser S., Mays A. (2017), "Mobile and Dirty: Does Using Mobile Devices Affect the Data Quality and the Response Process of Online Surveys?," *Social Science Computer Review*, 36(2), 212-230.
- Schoeni R. F., Stafford F., McGonagle K. A., Andreski P. (2013), "Response Rates in National Panel Surveys," *The ANNALS of the American Academy of Political and Social Science*, 645(1), 60-87.

- Schonlau M., Couper M. P. (2016), "Semi-Automated Categorization of Open-Ended Questions," *Survey Research Methods*, 10(2), 143-152.
- Schouten B., Peytchev A., Wagner J. (2017), *Adaptive Survey Design*, London, New York: CRC Press, Taylor & Francis Group.
- Schuman H., Converse J. (1971), "The Effects of Black and White Interviewers on Black Responses in 1968," *Public Opinion Quarterly*, 35, 44-68.
- Schuman H., Presser S. (1981), *Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording, and Context*, New York, NY: Academic Press.
- Schwarz N., Hippler H. J., Noelle-Neumann E. (1992), "A Cognitive Model of Response Order Effects in Survey Measurement," in *Context Effects in Social and Psychological Research*, eds. Schwarz N., Sudman S., pp. 187-199, New York, NY: Springer-Verlag.
- Schwarz N., Strack F., Hippler H. J., Bishop G. (1991), "The Impact of Administration Mode on Response Effects in Survey Measurement," *Applied Cognitive Psychology*, 5(3), 193-212.
- Seem E., Coombs J. (2017), "2020 Research and Testing: 2015 National Content Test Relationship Question Experiment Analysis Report," Washington, DC: U.S. Census Bureau, retrieved from <https://www2.census.gov/programs-surveys/decennial/2020/program-management/final-analysis-reports/2015nct-relationship-question-experiment.pdf>.
- Seeskin Z. H. (2016), *Evaluating the Use of Commercial Data to Improve Survey Estimates of Property Taxes*. US Census Bureau: Washington, DC. Retrieved from <https://www.census.gov/content/dam/Census/library/working-papers/2016/adrm/carra-wp-2016-06.pdf>

- SESTAT (2018), “SESTAT Data Tool,” National Center for Science and Engineering Statistics, National Science Foundation, Alexandria, VA, Retrieved from <https://ncesdata.nsf.gov/datadownload/>.
- SESTAT (2018), “SESTAT Metadata Explorer,” National Center for Science and Engineering Statistics, National Science Foundation, Alexandria, VA, retrieved from <https://ncesdata.nsf.gov/metadataexplorer/metadataexplorer.html>.
- Singer E. (2002), “The Use of Incentives to Reduce Nonresponse in Household Surveys,” in *Survey Nonresponse*, eds. Groves R. M., Dillman D. A., Eltinge J. L., Little J. A., pp. 163-177, New York, NY: John Wiley & Sons.
- Singer E., 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.
- Sinozich S., Langer G., Filer C., De Jong A. (2019), “New and Improved? Investigating Mode Effects in Two RDD-Online Transitions,” Paper presented at the 74th Annual Conference of the American Association for Public Opinion Research, Toronto, Ontario, Canada.
- Skalland B., George J., Welch V., Hill H A., Elam-Evans L. D., Knighton C., Smith C. (2017), “Testing the Impact of Mail Materials on Web Participation in the National Immunization Survey,” *Proceedings of the Survey Research Methods Section, American Statistical Association*. Alexandria, VA: American Statistical Association. Pp. 3708-3732.
- Skalland B., Khare, M. (2013), “Geographic Inaccuracy of Cell Phone Samples and the Effect on Telephone Survey Bias, Variance, and Cost,” *Journal of Survey Statistics and Methodology*, 1(1), 45-65.
- Slud, E. V. (2015), *Impact of Mode-Based Imputation on ACS Estimates*. (#ACS15-RER-07), Washington, DC: US Department of Commerce, US Census Bureau.

- Smit P. R., Rijk J. V. (2014), "History of the Dutch Crime Victimization Survey(s)," in *Encyclopedia of Criminology and Criminal Justice*, eds. Bruinsma G., Weisburd D., pp. 2286-2296, New York, NY: Springer.
- Smyth J. D., Christian L. M., Dillman D. A. (2008), "Does 'Yes or No' on the Telephone Mean the Same as 'Check-All-That-Apply' on the Web?," *Public Opinion Quarterly*, 72(1), 103-113.
- Smyth J. D., Dillman D. A., Christian L. M., McBride M. (2009), "Open-Ended Questions in Web Surveys: Can Increasing the Size of Answer Spaces and Providing Extra Verbal Instructions Improve Response Quality?," *Public Opinion Quarterly*, 73, 325-337.
- Smyth J. D., Dillman D. A., Christian L. M., Stern M. J. (2006), "Comparing Check-All and Forced-Choice Question Formats in Web Surveys," *Public Opinion Quarterly*, 70(1), 66-77.
- Smyth J. D., Olson K. M. (2018), "Mixed-Mode, Mixed-Device Self-Administered Surveys: Mail-Centric Versus Web-Centric Questionnaire Design and Layout," Paper presented at the 73rd Annual Conference of the American Association for Public Opinion Research, Denver, CO.
- Smyth J. D., Olson K. M. (2019), "The Effects of Mismatches between Survey Question Stems and Response Options on Data Quality and Responses," *Journal of Survey Statistics and Methodology*, 7(1), 34-65.
- Smyth J. D., Olson K. M. (Forthcoming), "How Well Do Interviewers Record Responses to Numeric, Interviewer Field-Code, and Open-Ended Narrative Questions in Telephone Surveys?" *Field Methods*.

- Smyth J. D., Olson K. M., Stange M. (Forthcoming), "Within-Household Selection Methods: A Critical Review and Experimental Examination," in *Experimental Methods in Survey Research: Techniques that Combine Random Sampling with Random Assignment*, eds. Lavrakas P., de Leeuw E., Traugott M., Kennedy C., Holbrook A., West B., Hoboken, NJ: John Wiley & Sons, Inc.
- Sommer J., Diedenhofen B., Musch J. (2016), "Not to Be Considered Harmful: Mobile-Device Users Do Not Spoil Data Quality in Web Surveys," *Social Science Computer Review*, 35(3), 378-387.
- Sonnega A., Faul J. D., Ofstedal M. B., Langa K. M., Phillips J. W. R., Weir D. R. (2014), "Cohort Profile: the Health and Retirement Study (HRS)," *International Journal of Epidemiology*, 43(2), 576-585.
- Stange M., Smyth J. D., Olson K. M. (2016), "Using a Calendar and Explanatory Instructions to Aid Within-Household Selection in Mail Surveys," *Field Methods*, 28(1), 64-78.
- Stange M., Smyth J. D., Olson K. M. (2019), "Drawing on LGB Identity to Encourage Participation and Disclosure of Sexual Orientation in Surveys," *The Sociological Quarterly*, 60(1), 168-188.
- Statistics Canada (2018), "Travel Survey of Residents of Canada (TSRC)," Statistics Canada, Ottawa, Ontario, retrieved from <http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=3810>.
- Steele E. A., Marlar J., Allen L., Kanitkar K. N. (2016), "Effects of an Initial Offering of Multiple Survey Response Options on Response Rates," Paper presented at the 71st Annual Conference of the American Association for Public Opinion Research, Austin, TX.

- Stern M., Sterrett D., Bilgen I. (2016), "The Effects of Grids on Web Surveys Completed with Mobile Devices," *Social Currents*, 3(3), 217-233.
- Sterrett D., Malato D., Stern M. J., Tompson T., Benz J., Reimer B. (2015), "Benefits and Challenges of Web Surveys in Mix-Mode Designs: Demographic and Data Quality Differences Across Modes in Survey of Households Recovering from Superstorm Sandy," Paper presented at the 70th Annual Conference of the American Association for Public Opinion Research, Hollywood, FL.
- Strobl M., Fahrney K., Nguyen M., Bibb B., Thissen M. R., Stephenson W., Mitchell S. (2008), "Using Computer Audio-Recorded Interviewing to Assess Interviewer Coding Error," Paper presented at the 63rd Annual Conference of the American Association for Public Opinion Research, New Orleans, LA.
- Struminskaya B., Weyandt K., Bosnjak M. (2015), "The Effects of Questionnaire Completion Using Mobile Devices on Data Quality," *Methods, Data, Analyses*, 9(2), 261-292.
- Survey of Consumers Mail Team (2012), "Methodology March 2011/April 2012 Survey of Consumers Mail Pilot Study." Report prepared by Survey of Consumers Mail Team for the Surveys of Consumers.
- Suzer-Gurtekin Z. T. (2013), "Investigating the Bias Properties of Alternative Statistical Inference Methods in Mixed-Mode Surveys," unpublished dissertation, University of Michigan.
- Suzer-Gurtekin Z. T., Valliant R., Heeringa S. G., de Leeuw E. D. (2018), "Mixed-Mode Surveys: Design, Estimation and Adjustment Methods," in *Advances in Comparative Survey Methodology*, eds. Johnson T. P., Pennell B. E., Stoop I., Dorer B., pp. 409-430, Hoboken, NJ: John Wiley & Sons.

- Tarnai J., Dillman D. A. (1992), "Questionnaire Context as a Source of Response Differences in Mail versus Telephone Surveys," in *Context Effects in Social and Psychological Research*, eds. N. Schwarz, S. Sudman, pp. 115-129, New York, NY: Springer-Verlag.
- Thieme M., Reist B. (2017), "Center for Adaptive Design Update," Paper presented to the Census Scientific Advisory Committee, March 30, 2017.
- Thomas R. K., Klein J. D. (2006), "Merely Incidental? Effects of Response Format on Self-Reported Behavior," *Journal of Official Statistics*, 22, 221-244.
- Tijdens K. (2014), "Dropout Rates and Response Times of an Occupation Search Tree in a Web Survey." *Journal of Official Statistics*, 30(1), 23-43.
- Tijdens K. (2015), "Self-identification of occupation in web surveys: requirements for search trees and look-up tables." *Survey Insights: Methods from the Field*. Retrieved from <https://surveyinsights.org/?p=6967>
- Timbrook J., Olson K., Smyth J. D. (2018), "Why do Cell Phone Interviews Last Longer? A Behavior Coding Perspective," *Public Opinion Quarterly*, 82(3), 553-582.
- Toepoel V., Das M., van Soest A. (2009), "Design of Web Questionnaires: A Test for Number of Items per Screen," *Field Methods*, 21, 200-213.
- Toepoel V., Lugtig P. (2014), "What Happens if You Offer a Mobile Option to Your Web Panel? Evidence From a Probability-Based Panel of Internet Users," *Social Science Computer Review*, 32(4), 454-560.
- Toepoel V., Lugtig P. (2015), "Online Surveys are Mixed-Device Surveys," *MDS Journal for Quantitative Methods and Survey Methodology*, 9(2), 155-162.

- Toninelli D., Revilla M. (2016), "Smartphones vs PCs: Does the Device Affect the Web Survey Experience and the Measurement Error for Sensitive Topics? A Replication of the Mavletova Couper's 2013 Experiment," *Survey Research Methods*, 10(2), 153-169.
- Tourangeau R. (2017), "Mixing Modes: Tradeoffs Among Coverage, Nonresponse, and Measurement Error," in *Total Survey Error in Practice*, eds. Biemer P., de Leeuw E. D., Eckman S., pp. 115-132, Hoboken, NJ: John Wiley & Sons.
- Tourangeau R., Conrad F. G., Arens Z., Fricker S., Lee S., Smith E. (2006), "Everyday Concepts and Classification Errors: Judgments of Disability and Residence," *Journal of Official Statistics*, 22, 385-418.
- Tourangeau R., Couper M. P., Conrad F. G. (2004), "Spacing, Position, and Order: Interpretive Heuristics for Visual Features of Survey Questions," *Public Opinion Quarterly*, 68, 368-393.
- Tourangeau R., Maitland A., Rivero G., Sun H., Williams D., Yan T. (2017), "Web Surveys by Smartphone and Tablets: Effects on Survey Responses," *Public Opinion Quarterly*, 81(4), 896-929.
- Tourangeau R., Maitland A., Steiger D., Yan T. (forthcoming), "A Framework for Making Decisions About Question Evaluation Methods." In *Advances in Questionnaire Design, Development, Evaluation and Testing*, eds. Beatty P.C. , Wilmot A., Collins D., Kaye L., Padilla J. L., Willis G. Hoboken, NJ: Wiley.
- Tourangeau R., Maitland A., Yan H. Y. (2016), "Assessing the Scientific Knowledge of the General Public: The Effects of Question Format and Encouraging or Discouraging Don't Know Responses," *Public Opinion Quarterly*, 80(3), 741-760.

- Tourangeau R., Rasinski K., Jobe J. B., Smith T. W., Pratt W. F. (1997), "Sources of Error in a Survey of Sexual Behavior," *Journal of Official Statistics*, 13, 341-365.
- Tourangeau R., Smith T. W. (1996), "Asking Sensitive Questions: The Impact of Data Collection Mode, Question Format, and Question Context," *Public Opinion Quarterly*, 60(2), 275-304.
- Tourangeau R., Sun H., Maitland A., Rivero G., Williams D. (2017), "Web Surveys by Smartphones and Tablets: Effects on Data Quality," *Social Science Computer Review*, 36(5), 542-556.
- Tourangeau R., Yan T. (2007), "Sensitive Questions in Surveys," *Psychological Bulletin*, 133(5), 859-883.
- Tourangeau, R. (2019), "How Errors Cumulate: Two Examples," *Journal of Survey Statistics and Methodology*. <https://doi.org/10.1093/jssam/smz019>
- Townsend S., Larsen M. E., Boonstra T. W., Christensen H. (2015), "Using Bluetooth Low Energy in Smartphones to Map Social Networks," Cornell University, retrieved from <https://arxiv.org/abs/1508.03938>.
- Transportation Research Board. (2016), *Exploring New Directions for the National Household Travel Survey: Phase Two Report of Activities* Transportation Research Circular E-C217. Retrieved from <http://onlinepubs.trb.org/Onlinepubs/circulars/ec217.pdf>
- Troldahl V. C., Carter Jr. R. E. (1964), "Random Selection of Respondents Within Households in Phone Surveys," *Journal of Marketing Research*, 1(2), 71-76.
- Tucker C., Lepkowski J. M., Piekarski L. (2002), "The Current Efficiency of List-Assisted Telephone Sampling Designs," *Public Opinion Quarterly*, 66(3), 321-338.

- U.S. Census Bureau (2014), *American Community Survey Design and Methodology (January 2014)*, Washington, DC: U.S. Bureau of the Census, retrieved from https://www2.census.gov/programs-surveys/acs/methodology/design_and_methodology/acs_design_methodology_report_2014.pdf.
- U.S. Census Bureau (2017), “Proposed Information Collection; Comment Request; 2018 National Sample Survey of Registered Nurses,” FR-2017-13292, Washington, DC: U.S. Bureau of the Census, retrieved from <https://www.federalregister.gov/documents/2017/06/26/2017-13293/proposed-information-collection-comment-request-2018-national-sample-survey-of-registered-nurses>.
- U.S. Census Bureau (2018a), *2017 National Survey of Children’s Health Methodology Report*, Washington, DC: U.S. Bureau of the Census, retrieved from <https://www.census.gov/content/dam/Census/programs-surveys/nsch/tech-documentation/methodology/2017-NSCH-Methodology-Report.pdf>.
- U.S. Census Bureau (2018b), *2016 National Survey of Children’s Health Methodology Report*, Washington, DC: U.S. Bureau of the Census, retrieved from <https://www.census.gov/content/dam/Census/programs-surveys/nsch/tech-documentation/methodology/2016-NSCH-Methodology-Report.pdf>.
- U.S. Department of Transportation, Federal Highway Administration, (2011), “2009 National Household Travel Survey User’s Guide”, Federal Highway Administration: Retrieved from <https://nhts.ornl.gov/2009/pub/UsersGuideV2.pdf>.

- U.S. Energy Information Administration (2017), *2015 RECS Square Footage Methodology*, Washington, DC, retrieved from https://www.eia.gov/consumption/residential/reports/2015/squarefootage/pdf/2015_recs_squarefootage.pdf.
- U.S. Energy Information Administration (2018), *Residential Energy Consumption Survey (RECS): 2015 Household Characteristics Technical Documentation Summary*, Washington, DC, retrieved from <https://www.eia.gov/consumption/residential/reports/2015/methodology/pdf/RECSmethodology2015.pdf>.
- U.S. Energy Information Administration (n.d.), *Residential Energy Consumption Survey (RECS)*, retrieved from <https://www.eia.gov/survey/#eia-457>.
- U.S. General Services Administration (n.d.), “Section 508: GSA Government-Wide IT Accessibility Program,” retrieved from <https://www.section508.gov/>.
- United Nations Economic Commission for Europe (UNECE), (2019), *Generic Statistical Business Process Model (GSBPM) (Version 5.1, January 2019)*, Retrieved from <https://statswiki.unece.org/display/GSBPM/GSBPM+v5.1>
- United Nations Economic Commission for Europe UNECE (2013), “Generic Statistical Business Process Model (GSBPM) Version 5.0, December 2013,” retrieved from <https://statswiki.unece.org/display/GSBPM/GSBPM+v5.0>.
- University of Michigan (n.d.), “Surveys of Consumers,” Institute for Social Research, University of Michigan, Ann Arbor, MI, retrieved from <https://data.sca.isr.umich.edu/>.
- Valliant R., Dever J. A. (2018), *Survey Weights: A Step-by-Step Guide to Calculation*, College Station, TX: Stata Press.

- Valliant R., Hubbard F., Lee S., Chang C. (2014), "Efficient Use of Commercial Lists in US Household Sampling," *Journal of Survey Statistics and Methodology*, 2(2), 182-209.
- Vannieuwenhuyze J., Loosveldt G., Molenberghs G. (2010), "A Method for Evaluating Mode Effects in Mixed-Mode Surveys," *Public Opinion Quarterly*, 74(5), 1027-1045.
- Vannieuwenhuyze J., Loosveldt G., Molenberghs G. (2012), "A Method to Evaluate Mode Effects on the Mean and Variance of a Continuous Variable in Mixed-Mode Surveys," *International Statistical Review*, 80(2), 306-322.
- Vannieuwenhuyze J., Loosveldt G., Molenberghs G. (2014), "Evaluating Mode Effects in Mixed-Mode Survey Data Using Covariate Adjustment Models," *Journal of Official Statistics*, 30(1), 1-21.
- Vicente P., Reis E., Santos M. (2009), "Using Mobile Phones for Survey Research," *International Journal of Market Research*, 51, 613-33.
- Wagner J., Olson K., Edgar M. (2017), "Assessing Potential Errors in Level-of-Effort Paradata Using GPS Data," *Survey Research Methods*, 11(3), 218-233.
- Walther J. B. (2002), "Research Ethics in Internet-Enabled Research: Human Subjects Issues and Methodological Myopia," *Ethics and Information Technology*, 4, 205-216.
- Wang J., Frechtel P., Sukasih A., Kinyon D. (2017), "Accounting for Data Collection Mode in Hot Deck Imputation," Paper presented at the 2017 Joint Statistical Meetings, Baltimore, MD.
- Wang W., Rothschild D., Goel S., Gelman A. (2015), "Forecasting Elections With Non-Representative Polls," *International Journal of Forecasting*, 31(3), 980-991.

- Weaver L., Beebe T. J., Rockwood T. (2019), “The Impact of Survey Mode on the Response Rate in a Survey of the Factors that Influence Minnesota Physicians’ Disclosure Practices,” *BMC Medical Research Methodology*, 19(1), 73-79.
- Wells B. M., Hughes T., Park R., CHIS Redesign Working Group, Rogers T. B., Ponce N. (2018), *Evaluating the California Health Interview Survey of the Future: Results from a Methodological Experiment to Test an Address-Based Sampling Mail Push-to-Web Data Collection*, Los Angeles, CA: UCLA Center for Health Policy Research, retrieved from <http://healthpolicy.ucla.edu/chis/design/Documents/CHIS%20Spring%202018%20ABS%20Web%20Field%20Experiment%20Report.pdf>.
- Wells T., Bailey J. T., Link M. W. (2014), “Comparison of Smartphone and Online Computer Survey Administration,” *Social Science Computer Review*, 32(2), 238-255.
- Wells, B.M., Hughes, T., Park, R., CHIS Redesign Working Group, Ponce, N. (2019), *Evaluating the California Health Interview Survey of the Future: Results from a Statewide Pilot of an Address-Based Sampling Mail Push-to-Web Data Collection*. Los Angeles, UCLA Center for Health Policy Research. Retrieved from [http://healthpolicy.ucla.edu/chis/design/Documents/CHIS%20Fall%202018%20ABS%20Web%20Pilot%20Report%20for%20DHCS%20\(July%202019\).pdf](http://healthpolicy.ucla.edu/chis/design/Documents/CHIS%20Fall%202018%20ABS%20Web%20Pilot%20Report%20for%20DHCS%20(July%202019).pdf).
- Wernimont J., Snowden R. (2015), “Integrated Management of Survey Modes,” Paper presented at the 2015 research conference of the Federal Committee on Statistical Methodology, Washington, DC.
- West B. T., Blom A. G. (2017), “Explaining Interviewer Effects: A Research Synthesis,” *Journal of Survey Statistics and Methodology*, 5, 175-211.

- West B. T., Wagner J., Hubbard F., Gu H. (2015), "The Utility of Alternative Commercial Data Sources for Survey Operations and Estimation: Evidence from the National Survey of Family Growth," *Journal of Survey Statistics and Methodology*, 3(2), 240-264.
- Westat. (2010), *National Household Education Survey Redesign: Report of Spring 2009 Cognitive Research*, Rockville, MD: Westat.
- Westat. (2013), *Health Information National Trends Survey 4 (HINTS 4): Cycle 2 Methodology Report*. Rockville, MD: Westat.
- Westat. (2018), *Health Information National Trends Survey 5 (HINTS 5): Cycle 2 Methodology Report*. Rockville, MD: Westat.
- Wettergren L., Mattsson E., von Essen L. (2011), "Mode of Administration Only Has a Small Effect on Data Quality and Self-Reported Health Status and Emotional Distress Among Swedish Adolescents and Young Adults," *Journal of Clinical Nursing*, 20(11-12), 1568-1577.
- Wilkinson-Flicker S., McPhee C., Medway R., Kaiser A., Cutts K. (2016), "Mixing Modes: Challenges (and Tradeoffs) of Adapting a Mailed Paper Survey to the Web," Paper presented at the 71st Annual Conference of the American Association for Public Opinion Research, Austin, TX.
- Williams D., Brick J. M. (2018), "Trends in U.S. Face-To-Face Household Survey Nonresponse and Level of Effort." *Journal of Survey Statistics and Methodology*, 6(2), 186-211.
- Williams D., Edwards S., Giambo P., Kena G. (2018), "Cost Effective Mail Survey Design," Paper presented at the Federal Committee on Statistical Methodology Research and Policy Conference, Washington, DC.

- Williams D., Sun H., Elkin I., To N. (2018), "Usability testing of an online consumer expenditure diary." Paper presented at the Annual Conference of the American Association for Public Opinion Research, Denver, CO.
- Willis G. B. (2005), *Cognitive Interviewing: A Tool for Improving Questionnaire Design*, Thousand Oaks, CA: Sage.
- Willis G. B. (2015), *Analysis of the Cognitive Interview in Questionnaire Design: Understanding Qualitative Research*, New York, NY: Oxford University Press.
- Willis G., Lessler J. (1999), *The BRFSS-QAS: A Guide for Systematically Evaluating Question Wording*, Rockville, MD: Research Triangle Institute.
- Winneg K., Ben-Porath E., Jamieson K. H. (2017), "Learning from the U.S. General Election Presidential Rebates: What Difference Does Mode Make?," Paper presented at the 72nd Annual Conference of the American Association for Public Opinion Research, New Orleans, LA.
- Wolter K. K., Smith P. J., Khare M., Welch B., Copeland K. R., Pineau V. J., Davis N. (2017), "Statistical Methodology of the National Immunization Survey, 2005-2014," *National Center for Health Statistics: Vital Health Statistics*, 1(61), 1-96.
- Yan T., Curtin R., Jans M. (2010), "Trends in Income Nonresponse Over Two Decades," *Journal of Official Statistics*, 26(1), 145-164.
- Ye C., Fulton J., Tourangeau R. (2011), "More Positive or More Extreme? A Meta-Analysis of Mode Differences in Response Choice," *Public Opinion Quarterly*, 75(2), 349-365.
- Yeager D. S., Krosnick J. A., Chang L., Javitz H. S., Levendusky M. S., Simpser A., Wang R. (2011), "Comparing the Accuracy of RDD Telephone Surveys and Internet Surveys

Conducted with Probability and Non-Probability Samples,” *Public Opinion Quarterly*,
75(4), 709-747.

Zuckerberg A., Mamedova S. (2012), “Speaking the Same Language: Effective Techniques for Reaching Spanish Speaking Households in a Mail Survey,” Paper presented at the 67th Annual Conference of the American Association for Public Opinion Research, Orlando, FL.