Oxford Handbooks Online

Longitudinal Surveys: Issues and Opportunities

D. Sunshine Hillygus and Steven Snell

The Oxford Handbook of Polling and Polling Methods
Edited by Lonna Rae Atkeson and R. Michael Alvarez

Subject: Political Science, Political Methodology, Political Behavior
Online Publication Date: Dec 2015 DOI: 10.1093/oxfordhb/9780190213299.013.7

Abstract and Keywords

Longitudinal or panel surveys, in which the same individuals are interviewed repeatedly over time, are increasingly common in the social sciences. The benefit of such surveys is that they track the same respondents so that researchers can measure individual-level change over time, offering greater causal leverage than cross-sectional surveys. Panel surveys share the challenges of other surveys while also facing several unique issues in design, implementation, and analysis. This chapter considers three such challenges: (1) the tension between continuity and innovation in the questionnaire design; (2) panel attrition, whereby some individuals who complete the first wave of the survey fail to participate in subsequent waves; and (3) specific types of measurement error—panel conditioning and seam bias. It includes an overview of these issues and their implications for data quality and outlines approaches for diagnosing and correcting for these issues in the design and analysis of panel surveys.

Keywords: panel surveys, longitudinal surveys, questionnaire design, measurement error, panel conditioning, seam bias

Introduction

Longitudinal or panel surveys, in which the same respondents are interviewed repeatedly at different points in time, are increasingly common across the academic, private, and public sectors. The major infrastructure surveys in political science, sociology, and economics—the American National Election Study (ANES), the General Social Survey (GSS), and the Panel Study on Income Dynamics (PSID)—now all contain panel components. The unique benefits of panel surveys are widely recognized: by interviewing the same subjects over time, panel surveys offer greater causal leverage than a cross-sectional survey and enable the analysis of individual-level changes in attitudes, behavior, or knowledge.

Advances in survey research technology, especially the proliferation of Internet-based surveying, have lowered the barriers to entry for longitudinal research. The emergence of online panels like GfK Knowledge Networks, YouGov, and the RAND American Life Panel makes it easier than ever for researchers to conduct repeated interviews (Hillygus, Jackson, and Young 2014; Baker et al. 2010; Yeager et al. 2011). Furthermore, in the past several years researchers have pooled their efforts and budgets in various collaborative panel studies, such as the Cooperative Congressional Election Study (CCES) and The American Panel Study (TAPS). The 2008 Associated Press-Yahoo!News Election Panel and the CBS/New York Times/YouGov 2014 Election Panel are two such projects that have involved collaborations between public opinion scholars and media organizations.

Despite their analytic strengths and increasing availability for research, panel surveys are not without their drawbacks. They share all the problems of other surveys—quality threats from sampling and nonsampling errors—while also facing several unique challenges in their design, implementation, and analysis. In this chapter we consider three such challenges: (1) a tension between continuity and innovation in the questionnaire design; (2) panel attrition, whereby some individuals who complete the first wave of the survey fail to participate in subsequent waves; and (3) types of measurement error—panel conditioning and seam bias—specific to panel surveys. We provide an overview of these various issues and their implications for data quality and also outline current approaches to diagnose and correct for these issues in the survey design and analysis.

First we define the longitudinal survey and distinguish it from related designs. We then discuss the advantages and disadvantages of longitudinal surveys, drawing attention to their unique challenges. Finally, we review best practices for avoiding the most common pitfalls and highlight avenues of future research that can improve the design and analysis of longitudinal polling.

Background

Although longitudinal surveys have a seemingly straightforward definition—they are survey projects in which respondents are interviewed at two or more points in time—it is useful to distinguish them from related designs, especially because of overlaps in usage of the relevant terminology.

The longitudinal survey belongs to the larger class of longitudinal methods because it is designed to elicit data from the same respondents at multiple time points (Menard 2002); nevertheless, there are alternative approaches for measuring individual-level change over time that do not qualify as panel surveys. Nonsurvey examples of longitudinal research abound in the social sciences, including a wide range of time series data, such as those using country-level factors (e.g., Beck, Katz, and Tucker 1998; Blais and Dobrzynska 1998), state- or county-level measures (e.g., Bishop 2013), or repeated observations of individuals from voter registration files or other nonsurvey sources (e.g., Davenport et al. 2010).

While not all longitudinal research is survey research, it is also the case that not all surveys designed to measure temporal change can be considered longitudinal surveys. A cross-sectional design, in which subjects are each interviewed only once, can be re-asked at different points in time using samples drawn independently (Visser et al. 2014; Menard 2002). An example of this repeated cross-sectional design is the typical tracking poll during a political campaign, designed to measure the ebbs and flows of candidate support. If sampling procedures and question wording are sufficiently similar, repeated cross-sectional surveys are an effective tool for detecting societal shifts in opinion. Repeated cross-sectional surveys can even be superior to panel surveys for some research questions. For example, the former might do a better job of capturing new entrants to a population, potentially providing a more accurate reflection of the population’s attitudes or behaviors in cases in which new entrants are especially different (Tourangeau 2003). Nevertheless, causal inference is generally weaker in a repeated cross-section than in a panel survey because the researcher can only compare groups of respondents rather than individuals (Visser et al. 2014; Tourangeau 2003; Bartels 1999).

Another method for measuring change is a retrospective survey design, in which respondents are asked during a single interview to recall attitudes or behaviors at several previous time periods (Menard 2002). This measurement strategy is distinct from the longitudinal survey because it relies on respondents’ retrospection rather than repeated interviews. While this approach allows researchers to measure within-subject change over time, an obvious deficiency is that it relies on memory recall, which introduces potential bias given the difficulty that some survey respondents have remembering even basic facts or behaviors (Bradburn, Rips, and Shevell 1987; Groves 2004; Tourangeau, Rips, and Rasinski 2000).

A final point of distinction exists between panel surveys and so-called online survey panels, like GfK Knowledge Networks and YouGov. Because of the difficulty of constructing a general population sample frame of email addresses, online survey panels have emerged as the primary way in which Internet polling is conducted (Groves et al. 2009). An online survey panel is simply a group of prescreened respondents who have expressed a willingness to participate in surveys, usually in exchange for money or other compensation (Baker et al. 2010).[1] The surveys in which these panelists take part might be cross-sectional or longitudinal. Despite the common use of the term "panel" to refer to this particular survey mode and sample source, this chapter focuses on the longitudinal survey design—in which the same respondents are interviewed multiple times for a given study. Such survey designs can be conducted online, by telephone, by mail, or in person.

That said, the Internet age has certainly expanded opportunities for longitudinal survey designs. The emergence of online survey panels facilitates the growing interest in longitudinal survey research by reducing the costs of subject recruitment and providing a pool of willing subjects who can be easier to locate for follow-up interviews. The willingness of online panelists to engage in additional surveys helps reduce a key cost of longitudinal research. On the other hand, the repeated interviewing of the same subjects might exacerbate the shortcomings of nonprobability online panels in particular. Researchers are increasingly concerned, for example, about the conditioning effects of repeated interviewing in both panel survey designs and online survey panels more generally (see Hillygus, Jackson, and Young 2014; Adams, Atkeson, and Karp 2012; Callegaro et al. 2014).

In addition to distinguishing what is and is not a panel survey, it is also worth highlighting the wide variability in the possible design features of panel surveys. Panel surveys can include dozens of waves or just an initial interview and a single follow-up. The ANES, for instance, typically includes one pre-election interview and one post-election interview—a two-wave panel. Panel surveys can also vary in the duration of the study and the length of time between survey interviews. The four-wave Youth-Parent Socialization Panel study spanned more than three decades, from 1965 to 1997, but most election panels span only a matter of months. Panel surveys also vary in their sampling strategy. A fixed panel design asks all respondents to participate at the same time, while a rotating panel divides the sample into different cohorts, with initial interviews staggered across survey waves. As discussed in the next section, the latter design offers useful leverage for assessing panel attrition and conditioning effects. Finally, designs differ in how they define their inferential population—some define only at the first wave, while others update at each wave. In other words, an individual who died between waves 1 and 2 would be counted as ineligible in the former and as a nonrespondent in the latter.

Eligibility for follow-up interviews can also vary—with some panels attempting to follow up with all respondents who complete the initial interview, while others select a narrower subset of respondents for subsequent interviews.[2] As with all research methodologies, the goals of the study—balanced against time and cost considerations—should guide these specific design decisions. For a more detailed overview of these and other design issues in panel surveys, see Menard (2007), Duncan and Kalton (1987), Kalton and Citro (1993), and Kasprzyk et al. (1989).

Advantages of Longitudinal Surveys

The growth of longitudinal surveys in the last several years reflects the significant benefits of repeated interviews with the same subjects. First, longitudinal surveys are critical for understanding the dynamics of public opinion. While cross-sectional surveys are well-suited to track societal trends in opinion over time, they cannot identify within-subject change (Tourangeau 2003; Visser et al. 2014). As such, it is difficult to determine if changes in public opinion, such as Americans' dramatic shift in attitudes about same-sex marriage, are a function of sampling and cohort replacement or a reflection of real changes in individual attitudes (e.g., Baunach 2011; Brewer 2008). Without conducting repeated surveys with the same subjects, we cannot evaluate who changed their minds or why.

This ability to evaluate within-subject change is what makes panel surveys a critical tool in the study of campaigns and elections. The seminal Columbia research on voting behavior was based on panel studies, such as the seven-wave sample of twenty-four hundred voters in Erie County, Ohio, during the 1940 election (Lazarsfeld, Berelson, and Gaudet 1948; Berelson, Lazarsfeld, and McPhee 1954). A longitudinal design enabled researchers to observe which voters changed their candidate preferences during the campaign. Although voting research came to rely increasingly on national cross-sectional surveys for much of the twentieth century, the last decade or so has seen a renewed interest in panel surveys as a tool for examining the decision calculus of voters at various points in the campaign (e.g., Henderson, Hillygus, and Tompson 2010; Iyengar, Sood, and Lelkes 2012). The strength of the panel design is that by interviewing the same respondents multiple times in the course of the campaign, the researchers have a much stronger sense of the evolution of individual-level voter decision making. Consider, for instance, that cross-sectional polls typically find that roughly 5% of the electorate is undecided between the candidates at any given point in the campaign; longitudinal surveys show that it is not always the *same* 5% of the electorate in every snapshot, offering a very different portrait of the campaign (Henderson and Hillygus forthcoming).

A second, and related, advantage of the longitudinal design is that measuring within-subject change offers greater leverage in estimating causal effects. This design is especially convincing if the pre- and post-intervention surveys closely precede and follow, or bracket, an intervention.[3] Such an intervention might be naturally occurring or a manipulation of the researcher. For example, Hillygus and Jackman (2003) compare interviews before and after presidential conventions and debates to estimate the effect of these major campaign events on candidate preference. With experimental interventions, panel surveys provide the pre-treatment baseline by which the post-treatment effects are later evaluated. Examples of such analyses include surveys gauging political knowledge and attitudes before and after respondents are randomly assigned to receive a free newspaper subscription (Gerber, Karlan, and Bergan 2009) and a panel survey designed to detect the effects of a large-scale campaign against vote buying on voter turnout and vote choice (Vicente 2014).

Even without an intervention, the within-subjects nature of the panel design provides the temporal ordering of measures that is necessary (though not sufficient) to establish causality (Bartels 2006). For example, researchers have used panel data to explore the complex relationship between party identification and policy preferences (Carsey and Layman 2006) and between media messages and issue attitudes (Lenz 2009). While this approach has a somewhat weaker claim to causality, the temporal ordering of the measurement makes it far superior to traditional observational studies.

A third advantage of the longitudinal survey design is the opportunity it provides researchers to assess the reliability of the concepts being measured, a critical component of measurement error. Reliability refers to the degree to which consecutive measurements of a given concept yield the same result, provided that the meaning of the concept has not changed across time. Some phenomena can easily be measured reliably—gender and level of education, for example—while most concepts of interest to social scientists are subject to measurement error. In classical test theory, test-retest stability is a standard approach for evaluating reliability and necessitates a longitudinal design (Carmines and Zeller 1979; Bartels 2006). For example, Achen (1975) reassesses the seminal analysis of early ANES panels (Converse 1964) and finds that much of the instability across time in voter preferences is attributable to the poor reliability of survey measures. Longitudinal surveys also enable measurement error adjustments. For example, in panels with an item measured three or more times, the researcher can employ the difference in responses from one set of waves to assess the reliability of the item and to then correct appropriately for measurement bias when comparing responses to the same question across another set of waves (e.g., Bartels 1999). This calibration exercise allows researchers to control for and better distinguish measurement noise from real attitude change.

Challenges in Longitudinal Surveys

As the previous discussion makes clear, panel studies offer a number of compelling advantages for studying social, political, and economic phenomena. They do, however, come with some downsides. First, longitudinal data have a complex structure that can complicate analysis. By virtue of having multiple interviews with the same respondents, the data have a hierarchical structure that should be accounted for in the statistical modeling (Gelman and Hill 2007). There is a wide variety of modeling approaches for handling panel data: change point models, duration models, transition models, fixed effect models, hierarchical models, and so forth. Unfortunately, the substantive conclusions can differ depending on the modeling approach used, and it is not always clear which approach is best suited to the research question. Broadly, analysts can model either the level of (y) or the change in y (Δy) as a function of either the level of or change in the levels of the predictor variables, where the number of possible combinations depends on the number of survey waves used in the analysis. Given that the particular research question and data structure will determine the most appropriate modeling strategy, we refer readers to dedicated texts such as Singer and Willett (2003), Finkel (1995), and Hsiao (2003). Another complexity in analyzing longitudinal surveys is that it not always obvious which weight to use given that multiple weights are often provided.[4] Again, the decision depends on the research question and the particular combination of waves used, but generally analysts will want to use the longitudinal weight associated with the wave in which their dependent variable is measured.

Panel surveys also face a number of threats to data quality that can jeopardize the ability to make inferences about the outcomes of interest. To be sure, all survey research faces a litany of challenges that can threaten the validity and reliability of survey estimates. A rich literature across academic and professional fields has made strides in identifying potential sources of bias in survey research (e.g., Groves 2004; Groves et al. 2009; Groves and Couper 2012; Weisberg 2005). The “Total Survey Error” paradigm classifies survey error as pertaining to survey sampling, coverage, nonresponse, measurement, and postsurvey analysis and recommends best practices in survey design, implementation, and evaluation to mitigate these errors (e.g., Biemer 2011; Groves and Lyberg 2011; Weisberg 2005). In addition to these usual sources of error, however, panel surveys face additional threats to quality associated with measuring the same individuals at different points in time. We outline three such challenges here: (1) a tension between continuity and innovation in the questionnaire design; (2) panel attrition; and (3) panel conditioning and seam effects (panel-specific measurement error).

Balancing Continuity and Innovation in Panel Surveys

Given that the ability to track within-subjects change is one of the longitudinal survey design's chief benefits, it perhaps goes without saying that the basic "way to measure change is not to change the measure" (Smith 2005). Yet longitudinal studies often face a tension between the need for comparability over time and the pressure to change the question wording or other design features of the study. Especially in panels that span an extended time period, there may be compelling reasons to modify, update, or retire a question (Tourangeau 2003). For example, after nearly one hundred years of use, the U.S. Census Bureau in 2013 dropped the word "Negro" from its racial response categories. Even within a shorter panel, there can be reasons to change question wording. Within a political panel survey of an election campaign, for instance, it is common for vote choice response options to change from the choice between a generic Democrat and Republican during the nomination stage to the choice between two specific candidates after the party nominees are known. Research has consistently shown that public opinion estimates are sensitive to even small differences in question wording and response options (e.g., Green and Schickler 1993; Abramson and Ostrom 1994). Moreover, responses can also be affected by changes in other survey design features such as mode, incentives, fielding period, question order, and the like (Jackson 2011).

The point is simply that questionnaire or survey design changes should not be made lightly and require experimentation and calibration to lessen the inherent loss of continuity and comparability. Two kinds of experimentation are useful. The first is an "offline" experiment, wherein additional subjects participate in separate pilot studies, which randomize respondents to potential versions of the changes under consideration (Tourangeau 2003). Given the expense of longitudinal research, this process of independent piloting is valuable because researchers can more fully understand response properties and refine the revised survey item before interfering with the continuity of the panel. The second type of experiment is a split-ballot design within the panel survey (Tourangeau 2003). This similarly allows researchers to make between-item comparisons for the larger sample, but provides the additional benefit of sustaining the time series by presenting the original item to some subset of respondents. While experimentation should guide necessary adaptations of existing items, transparency regarding what has changed and why is the other key to successful innovation (Jackson 2011).

Panel Attrition

Perhaps the most well-recognized challenge to longitudinal studies is panel attrition, wherein some respondents in the sample fail to complete subsequent waves. Attrition affects longitudinal studies of all types, modes, and sponsors. For instance, the multiple-decade PSID, first fielded in 1968, lost nearly 50% of the initial sample members by 1989. The ANES 2008–2009 Panel Study lost 36% of respondents in less than a year of monthly interviews. At best, attrition reduces effective sample size, thereby decreasing analysts' abilities to discover longitudinal trends in behavior. At worst, attrition results in an available sample that is not representative of the target population, thereby introducing biases into estimates of the outcomes of interest. Recent expansions in the number and use of panel surveys, coupled with worsening response rates, make the issue of panel attrition particularly salient. It is well-documented that response rates for all surveys, including government surveys, have been in decline in recent decades (Hillygus et al. 2006). The implications may be particularly severe for panel studies since they depend on respondents participating at multiple points in time (Schoeni et al. 2013). Even high-quality government surveys have found that nonresponse and attrition have grown worse in recent years. For example, before 1992 the Survey of Income and Program Participation (SIPP) typically lost about 20% of the original sample by the final wave. That loss rate increased to 35.5% in 1996 and 36.6% in 2004 (Westat 2009, 22).

Reinterviewing the same panelists can be a labor-intensive process: researchers must locate, recontact, and persuade the panelist to participate in later waves. If any of these three steps breaks down, the case is lost (Watson and Wooden 2009). The need to track panelists to new locations can substantially increase both survey costs and the difficulty of gaining cooperation, leading some long-duration panels to alter their sampling design. For instance, the Early Childhood Longitudinal Study of the National Center for Education Statistics sampled only 50% of students who moved schools between waves to help reduce the cost of follow-up interviews. In sum, panel attrition is a problem for all panel surveys, the problem has worsened over time, and there are now more data analysts who have to contend with the problem.

The threats of panel attrition are widely recognized by public opinion researchers (e.g., Ahern and Le Brocque 2005; Traugott and Rosenstone 1994; Zabel 1998), but there is little consensus about how to handle it. Analyses of panel attrition tend to be reported and published separately from those of substantive research (e.g., Zabel 1998; Fitzgerald, Gottschalk, and Moffitt 1998; Bartels 1999; Clinton 2001; Kruse et al. 2009). Yet panel attrition is not just a technical issue of interest only to methodologists; it can have direct implications for the substantive knowledge claims that can be made from panel surveys. For example, Bartels (1999) showed that differential attrition of respondents in the 1992–1996 ANES panel resulted in an overestimation of political interest in the population. Frankel and Hillygus (2013) show that attrition in the 2008 ANES panel study biased estimates of the relationship between gender and campaign interest.

Too often, researchers simply ignore panel attrition, conducting the analysis on the subset of respondents who completed all panel waves data (e.g., Wawro 2002). In a review of the literature, Ahern and Le Brocque (2005) find that fewer than one-quarter of studies employing panel data discuss attrition or offer any analyses to detect or correct for panel attrition. In doing so, scholars make an assumption that panel attrition occurs randomly. In the language of the missing data literature (Little and Rubin 2002), any complete-case descriptive analysis assumes the missing data—subsequent survey waves, in this case—are missing completely at random (MCAR). That is, no observed or unobserved data can systematically predict or account for this missingness. Unfortunately, this assumption is almost always unfounded. Countless analyses have found that panel attrition is related to a variety of respondent characteristics (e.g., Behr 2005).

Broadly speaking, the literature on the correlates of panel attrition emphasizes that repeated participation in a panel survey depends on both the ability and motivation to cooperate. As such, characteristics like income, education, gender, race, and being foreign born correlate with attrition (Gray et al. 1996; Fitzgerald, Gottschalk, and Moffitt 1998; Loosveldt, Pickery, and Billiet 2002; Behr 2005; Lynn et al. 2005; Watson and Wooden 2009). Individuals who are more socially engaged and residentially stable—homeowners and those with children (especially young children) at home—are more likely to remain in a panel study, while younger respondents and those who live alone are more likely to drop out (Lipps 2007; Uhrig 2008; Watson and Wooden 2009; Groves and Couper 2012). Research also shows that civic engagement and interest in the survey topic are correlated with attrition; those who care more about the topic are less likely to attrit (Traugott and Morchio 1990; Traugott and Rosenstone 1994; Loosveldt and Carton 1997; Lepkowski and Couper 2001; Loosveldt, Pickery, and Billiet 2002; Voogt 2005; Smith and Son 2010). Measures of political engagement and political interest, in particular, can be predictive of attrition in surveys on all topics, but are especially predictive of attrition in surveys with political content (Brehm 1993; Traugott and Rosenstone 1994; Bartels 1999; Burden 2000; Voogt and Saris 2003; Olson and Witt 2011). For example, Olson and Witt (2011) find that political interest has been consistently predictive of retention in the ANES time series from 1964 to 2004. More recent research has also emphasized that the respondents' survey experience in the first wave will influence their likelihood of participating in future waves (e.g., Frankel and Hillygus 2013). Given the wide range of attrition correlates, Chen et al. (2015) recommend a step-by-step process of identifying the predictors of attrition based on wave 1 responses and sampling frame data.[5]

In case of expected attrition bias, there is a variety of approaches for correcting estimates to improve inference. The use of post-stratification weights is the most common correction method used, and attrition-adjusted survey weights are routinely provided by survey firms. Weighting is not without controversy, however. As Deng et al. (2013) highlight, there is wide variability in the way weights are constructed and in the variables used to account for panel attrition. While researchers typically weight to demographic benchmarks like the Current Population Survey (CPS) or American Community Survey (ACS), Vandecasteele and Debels (2006) argue that weights based on demographic variables alone are likely inadequate to correct for attrition. Weights can also result in increased standard errors and introduce instabilities in the estimates (Gelman 2007).[6]

An alternative approach is imputation, in which the attrited cases are replaced with plausible values. While there are many different imputation methods, the preferred approach is multiple imputation, in which multiple values are estimated to replace the missing data (Pasek et al. 2009; Honaker and King 2010). As with weighting, standard approaches to multiple imputation assume that missing cases are missing at random (MAR)—dependent on observed data, but not unobserved data.

Another approach for dealing with panel attrition is through specialized statistical models. In cases in which MCAR or MAR assumptions are implausible, selection models (Hausman and Wise 1979; Brehm 1993; Kenward 1998; Scharfstein, Rotnitzky, and Robins 1999) or pattern mixture models (Little 1993; Kenward, Molenberghs, and Thijs 2003) can be used to model attrition that is not missing at random (NMAR)—dependent on the values of unobserved data. These approaches, however, also require strong and untestable assumptions about the attrition process, because there is insufficient information in the original panel data to understand why some cases are missing (e.g., Schluchter 1992; Brown 1990; Diggle and Kenward 1994; Little and Wang 1996; Daniels and Hogan 2008). Recent research shows that refreshment samples can be used as leverage for modeling the attrition process (Bhattacharya 2008; Deng et al. 2013; Hirano et al. 1998, 2001; Si, Reiter, and Hillygus 2014). A refreshment sample is a new sample, independently drawn and given the same questionnaire at the same time as the original panelists. Newly introduced cohorts in a rotating panel offer similar leverage. The comparison of these new data to the original panel allows researchers to properly correct estimates from the panel data.

Because substantive results can be sensitive to the particular corrective approach employed (Zabel 1998; Kristman, Manno, and Côté 2005; Ayala, Navarro, and Sastre 2006; Basic and Rendtel 2007), the best approach for handling panel attrition is to prevent it in the first place. At the end of the chapter, we review recommendations for design decisions that can help to mitigate attrition and other panel effects.

Panel-Specific Measurement Error

It is perhaps ironic that one of the advantages of panel surveys is that they enable assessment of the reliability of survey measures because they can also introduce additional measurement error—panel conditioning and seam effects—that can threaten the validity of survey estimates. We consider each of these issues in turn.

Panel Conditioning

Panel conditioning, also known as time-in-sample bias, refers to the phenomenon in which participation in earlier waves of the panel affects responses in subsequent waves. For example, respondents might pay more attention to a political contest because they are participating in a panel about voting and know they will be asked their opinions about the candidates. Warren Miller, a pioneer of the ANES, used to joke that the study's panel design was an expensive voter mobilization effort because participation in the pre-election survey motivated respondents to show up at the polls. Conditioning effects can jeopardize the validity of survey estimates, biasing estimates of the magnitude and/or correlates of change (Kasprzyk et al. 1989; Sturgis, Allum, and Brunton-Smith 2009; Warren and Halpern-Manners 2012).

Researchers have long been concerned about panel conditioning effects.[7] In one of the earliest political panel surveys, researchers identified the potential for panel conditioning, noting that "the big problem yet unsolved is whether repeated interviews are likely, in themselves, to influence a respondent's opinions" (Lazarsfeld 1940, 128). Clausen (1968) found that those who participated in a pre-election survey in 1964 were more likely to report voting in the post-election survey—he attributed seven percentage points to the stimulating effect of participating in the pre-election interview. Traugott and Katosh (1979) replicated the study and found an even larger mobilization effect. Many others have reached similar conclusions (Kraut and McConahay 1973; Yalch 1976; Greenwald et al. 1987; Anderson, Silver, and Abramson 1988; Granberg and Holmberg 1992; Simmons, Bickart, and Lynch Jr 1993; Bartels 1999; Voogt and Van Kempen 2002). Although political interest and political knowledge are commonly found to be susceptible to panel conditioning effects, the issue is not restricted to political surveys. For example, Battaglia, Zell, and Ching (1996) found that asking mothers about the immunization status of their children led to higher vaccination rates after the interview. Unfortunately, it is not always clear when panel conditioning will be an issue. While there is considerable documentation that panel conditioning *can* exist, it is not *always* present. Some research finds limited or no panel conditioning bias (Bartels 1999; Smith, Gerber, and Orlich 2003; Kruse et al. 2009). More generally, there is a lack of clarity in the research about the conditions under which panel conditioning is expected to change attitudes, behaviors, or knowledge. In addition, panel conditioning effects might depend on the characteristics of respondents, the topic of the survey, or a variety of other survey design factors. Moreover, Mann (2005) has disputed the methodological basis of much of the previous research identifying panel conditioning effects. The common approach to diagnosing conditioning effects is to simply compare panelist responses in follow-up waves with cross-sectional measures of the same items. Even when using refreshment samples or rotating samples, it can be difficult to distinguish panel conditioning effects from attrition bias (Warren and Halpern-Manners 2012).[8] For instance, inflated turnout levels in the ANES post-election survey may be due to panel conditioning, attrition among those not interested in politics, or other sources of survey error, such as bias in initial nonresponse (Burden 2000).

The specific mechanisms by which panel conditioning effects occur also vary. Changes in behavior might occur if survey participation increases respondent motivation or interest in the topic—as is the case for political knowledge in an election panel (Bartels 1999; Kruse et al. 2009). Alternatively, survey respondents could change their responses as they become more familiar with the interview process and survey experience. The first type of panel conditioning has been referred to as "conditioning change in true status," and the second is called "conditioned reporting." Conditioned reporting is a strategic response to the interview, such as learning to give answers that reduce the number of follow-up questions. This second type of panel conditioning is closely linked with the issue of "professional" respondents in online survey panels. These are respondents who have a lot of experience with taking surveys, so they might understand how to answer in such a way as to reduce burden and maximize their paid incentives. Indeed, there may well be concerns that panel survey research that relies on samples derived from online respondent panels will have panelists who are already conditioned at the time of the first wave because they have already participated in previous surveys on related topics. It is quite common, for instance, to find that YouGov and GfK panelists are more politically knowledgeable than the general population.[9] In principle, it should be possible to distinguish conditioned reporting from conditioned responses through studies designed to specifically test these different mechanisms. Unfortunately, such research is rare.[10]

There is also little guidance about what to do if panel conditioning bias is found in a longitudinal study. Some researchers contend that "once they occur the resulting data are irredeemably biased" (Warren and Halpern-Manners 2012). This means that it is all the more important for researchers to prevent panel conditioning in the design of their surveys as we discuss in more detail at the end of the chapter. For example, research has suggested that panel conditioning effects are more common when the baseline and follow-up surveys are separated by a month or less (e.g., Bailar 1989; De Amici et al. 2000; Fitzsimons, Nunes, and Williams 2007; Levav and Fitzsimons 2006).

Seam Effects

Another source of measurement error unique to longitudinal surveys has been termed “seam bias”; it refers to the tendency of estimates of change that are measured across the “seam” of two successive survey waves to far exceed estimates of change that are measured within a single wave (Conrad, Rips, and Fricker 2009). That is, when respondents are asked to recall behaviors or conditions at multiple reference times in a single interview—for example, employment status in the current month and in the previous month—they report few changes between the referenced time periods; in contrast, estimates of change are much higher if they are measured in two separate waves of data collection. As a result, estimates of month-to-month changes in employment status are far higher when looking across survey waves than when reported within a single interview (Lynn and Sala 2006).

Seam effects have been most often studied in economics, but they have been found across a wide range of measures, recall periods, and design features (Lemaitre 1992). Seam effects were first documented in estimates of government program participation in the Census Bureau’s SIPP panel survey (Czajka 1983), but have also been found in the CPS (Cantor and Levin 1991; Polivka and Rothgeb 1993), the PSID (Hill 1987), the Canadian Survey of Labour and Income Dynamics (Brown, Hale, and Michaud 1998), and the European Community Household Panel Survey (Jackle and Lynn 2004).

Research examining the source of seam bias suggests that it stems both from respondents underestimating change within the reference period of a single interview and overestimating change across waves. Collins (1975), for example, speculates that between two-thirds and three-quarters of the observed change in various employment statistics (as measured in a monthly labor force survey) were an artifact of this type of measurement error. Lynn and Sala (2006) label the amount of change they observe from one survey wave to the next in various employment characteristics as “implausibly high.” At the same time, researchers have documented underestimates of change within a single wave, a phenomenon labeled “constant wave responding” (Martini 1989; Rips, Conrad, and Fricker 2003). Using record validation, Marquis and Moore (1989) confirm that both factors produce the seam effect.

Seam bias has largely been attributed to respondent memory issues and task difficulty. For example, there is larger seam bias found with wider time intervals between waves and the to-be-recalled change (Kalton and Miller 1991). There are also larger seam effects when the recall items are more cognitively difficult (Lynn and Sala 2006). Some have suggested that seam bias can be further exacerbated by panel conditioning because individuals learn that it is less burdensome to give the same response for each referenced time than to report change (Rips, Conrad, and Fricker 2003).

A related phenomenon identified in political surveys is a sharp discrepancy in the stability of vote choice or time of vote decision when measured via recall in a post-election survey compared to estimation based on measures of candidate support from multiple waves of panel data (Plumb 1986; Chaffee and Rimal 1996; Fournier et al. 2004). Researchers have found discrepancies at both the aggregate and individual levels (Plumb 1986; Chaffee and Rimal 1996). For example, in an analysis of vote intention stability in the four-wave ANES 1980 panel study, Plumb (1986) finds that just 40% of respondents had the same time of decision with both methods. Critically, some find that the recall measure produces higher levels of stability (Plumb 1986), while others find it produces lower levels of stability (Katz 1971; Kogen and Gottfried 2012). Several explanations have been offered. First, it may be difficult for respondents to remember when the decision was made, especially if asked several months after the fact. Second, there might be issues of social desirability, whereby respondents might prefer to indicate that they delayed their decisions in order to appear neutral or independent. Alternatively, some—especially partisans—might claim they knew all along, not wanting to admit that they were ever undecided.

In terms of mitigating seam bias, the preponderance of research has focused on efforts to improve respondent recall (Callegaro 2008). For example, Rips, Conrad, and Fricker (2003) demonstrate that researchers can reduce seam effects by altering question order. They reason that seam bias is a predictable pattern of satisficing given the usual grouping of questions by topic instead of time period (Rips, Conrad, and Fricker 2003; Conrad, Rips, and Fricker 2009). Furthermore, respondents did best when time was ordered backwards, or in reverse chronological order—asking first about the most recent week and then about earlier and earlier weeks (Rips, Conrad, and Fricker 2003).

The other innovation that targets seam effects at the design stage is dependent interviewing (DI), which addresses the issue of seam bias straight on by automatically populating a panelist's previous response and asking if the response still holds (Conrad, Rips, and Fricker 2009; Moore et al. 2009; Lynn et al. 2005). The previous response serves as a reminder or anchor by which the respondent can compare the present, perhaps causing reflection on any change and when it may have occurred (Moore et al. 2009). Dependent interviewing is increasingly common, having been employed in the Census Bureau's SIPP and CPS projects (Conrad, Rips, and Fricker 2009), and is thought to improve interview times and general data quality; nevertheless, Lynn et al. (2005) cautions that the method may underestimate change across waves if it induces acquiescence bias among respondents who want to tell the interviewer that the previous response is still accurate.

As with panel conditioning, the best solution for seam effects is to prevent them. Though there are some post-survey methods for dealing with seam bias, many of them effectively throw away data. For an overview of such methods, see Lynn et al. (2005).

Recommendations for Researchers

The trend favoring longitudinal surveys will almost certainly continue given the method's ability to track within-subject change. Nevertheless, as with all survey methods, longitudinal surveys face several challenges to their validity and reliability. Responsible researchers must acknowledge the potential impact of these challenges on substantive knowledge claims. In addition to threats from declining response rates, concerns about the representativeness of survey respondents, and difficulties measuring various attitudes and behaviors—issues that arise in all survey designs—longitudinal surveys can face the unique challenges of comparability issues, panel attrition, panel conditioning, and seam effects. Researchers should grapple with potential biases from attrition and measurement error as a matter of course. Analyses should routinely include assessments of the quality of panel composition and resulting data, using whatever information about attrition can be gleaned by comparing later waves to earlier waves on observable factors like respondent demographics, survey satisfaction, or other measures related to respondent experience. Despite some potential limitations of weighting as a correction for attrition, we recommend that—at minimum—analysts calculate estimates using the longitudinal survey weights. Better still, researchers should leverage refreshment samples or rotating panels, if available, to better understand the impact of attrition bias and panel conditioning on the survey estimates.

It is the producers of new longitudinal surveys, however, who bear the greatest responsibility for preventing panel effects. Those designing panel surveys can take several measures to reduce panel survey error and improve the reliability and validity of the resulting data. Given findings about the relationship between the survey experience and attrition, the researcher should first ensure that the questionnaire, especially the questionnaire for the first survey wave, adheres to best practices in questionnaire design.[11] Furthermore, the researcher should enact protocols to make certain that interviewers are well trained, as poor interviewer performance decreases panelists' propensity to respond in later waves. Even in Internet polling, in which there is no traditional interviewer, the survey design must take into account potential technological issues and general user-friendliness, as difficulties with the online interface similarly cause panelists to attrit (Frankel and Hillygus 2013).

This also points to the need to explicitly measure respondents' survey experience, such as including a survey satisfaction item at the end of the first wave questionnaire. Where respondents report satisfaction with the interviewer, the researcher can reduce nonresponse in later waves by assigning the same interviewer to all follow-up interviews. When a respondent is found to be at risk of attriting, design adaptations can be made to increase the likelihood of response—for example, increasing the incentive payments for those with a high propensity to attrit (Laurie and Lynn 2009; Schoeni et al. 2013). The researcher executing a panel survey design must also take great care to keep track of panelists. Lepkowski and Couper (2001) identify the researcher's inability to locate and contact panelists as a major source of panel attrition. When respondents cannot be identified at the time of a later survey, cases are lost, resulting in a reduction of effective sample size and potentially biasing estimates for the remaining cases. The researcher can prevent lost cases by undertaking several activities to track respondents, such as instigating communication with the panelist between waves that are spaced far apart, collecting additional contact information (e.g., a mailing address and phone number, even if the primary means of communication is email), and using public records and administrative data sources for tracing respondents. For example, the PSID regularly updates panelist addresses using the United States Postal Service national change of address service and offers respondents a $10 payment to simply return a prepaid postcard verifying their full contact information (Schoeni et al. 2013). This sort of mailing belongs to the broad class of "keeping in touch exercises" (KITEs) (Laurie 2007). Another activity to improve tracking of panelists is the use of a dedicated website for respondents with information about the study, past results, and a change-of-address form.

The researcher can also address measurement error through careful survey design. A researcher concerned about panel conditioning might interview respondents less frequently, since panel conditioning can be exacerbated by frequent and closely spaced interviews. On the other hand, infrequent waves that are spaced far apart might rely more heavily on recall regarding the period between waves, which can induce seam effects. The researcher is left to balance these different considerations, with the optimal design depending on the research question and variables of interest. For instance, panel conditioning has been shown to have relatively limited effects on attitudinal questions, but strong effects on political knowledge. If the researcher wants to engage questions about the relationship between political knowledge and various outcomes, the best design would minimize conditioning effects by asking political knowledge questions infrequently and perhaps by developing new political knowledge items. On the other hand, if the primary goal is to observe change in some attitude or behavior, the researcher might do best to field many waves close together—thereby minimizing seam effects at the possible risk of inducing some conditioning.

As we hope this chapter makes clear, there are many opportunities for future research that could inform the design, conduct, and analysis of panel surveys. Researchers could build into the panel design observational or experimental features to distinguish and measure the various sources of longitudinal survey error. For example, a new panel of respondents for a longitudinal survey might gain traction on the distinction between panel attrition and conditioning by drawing on a very rich sampling frame, such as a voter registration database enhanced with commercial data. This kind of list would provide relatively straightforward criteria for measuring nonrandom attrition, by comparing the pre-study covariates of returning panelists and those who drop out and would also provide some leverage on conditioning, by allowing the researcher to compare the respondents' predicted and actual responses and behaviors.

Experimental designs might manipulate the panel survey experience for some respondents in order to gain a clearer understanding of how to minimize survey error. For instance, building on the earlier discussion of panel conditioning versus seam effects, the researcher could randomize respondents to complete several or few surveys that are spaced near or far apart. Similarly, the researcher can evaluate other design tradeoffs by randomizing design differences across panelists. For example, previous research suggests that the researcher can stem panel attrition by increasing communication with panelists, directing them to a study website, and sharing details of study findings with them. These measures are meant to increase panelists' interest in and commitment to the panel survey (Schoeni et al. 2013), but the researcher should consider whether these efforts—especially the provision of study results—contribute to panel conditioning. An experimental design could randomize the use of these particular retention efforts to estimate their effect on attrition and panel conditioning.

In addition, given the extent to which longitudinal survey research is being conducted with online panels, more research should consider how the online setting reduces or exacerbates the various types of error unique to the longitudinal survey design. Building on Adams, Atkeson, and Karp (2012) and Hillygus, Jackson, and Young (2014), such research might compare the panel conditioning effects of new recruits who enter either online panels or other types of panel surveys. Research on survey error in online surveys would be greatly enhanced if collaborations with the proprietors of online panels provided not just the number of surveys completed and panelists' time in the panel (Clinton 2001; Adams, Atkeson, and Karp 2012), but also information about the kinds of surveys to which the panelist has been invited and the kinds of surveys that the panelist has actually completed.

It is our hope that future research on panel survey error will not only provide a more comprehensive list of best practices to prevent and to measure survey error, but also will mitigate these biases when they are found in existing longitudinal survey data.

Acknowledgments

This work was supported by NSF Grant SES-10-61241. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

Abramson, P. R., and C. W. Ostrom. 1994. “Question Wording and Partisanship: Change and Continuity in Party Loyalties During the 1992 Election Campaign.” *Public Opinion Quarterly* 58 (1): 21.

Achen, C. H. 1975. “Mass Political Attitudes and the Survey Response.” *American Political Science Review* 69 (4): 1218–1231.

Adams, A. N., L. R. Atkeson, and J. A. Karp. 2012. “Panel Conditioning in Online Survey Panels: Problems of Increased Sophistication and Decreased Engagement.” Prepared for delivery at the American Political Science Association Annual Meeting. New Orleans.

Ahern, K., and R. Le Brocque. 2005. “Methodological Issues in the Effects of Attrition: Simple Solutions for Social Scientists.” *Field Methods* 17 (February): 53–69.

Anderson, B A., B. D. Silver, and P. R. Abramson. 1988. “The Effects of the Race of the Interviewer on Race-related Attitudes of Black Respondents in SRC/CPS National Election Studies.” *Public Opinion Quarterly* 52 (3): 289–324.

Ayala, L., C. Navarro, and M. Sastre. 2006. *Cross-country Income Mobility Comparisons under Panel Attrition: The Relevance of Weighting Schemes*. Technical report, Instituto de Estudios Fiscales.

Bailar, B. A. 1989. “Information Needs, Surveys, and Measurement Errors.” In *Panel Surveys*, edited by D. Kasprzyk, G Duncan, G. Kalton, and M. P. Singh, 1–24. New York: Wiley.

Baker, R., S. J. Blumberg, J. M. Brick, M. P. Couper, M. Courtright, J. M. Dennis, … D. Zahs. 2010. “Research Synthesis: AAPOR Report on Online Panels.” *Public Opinion Quarterly* 74 (October): 711–781.

Bartels, L. M. 1999. “Panel Effects in the American National Election Studies.” *Political Analysis* 8 (January): 1–20.

Bartels, L. M. 2006. “Three Virtues of Panel Data for the Analysis of Campaign Effects.” In *Capturing Campaign Effects*, edited by H. E. Brady and R. Johnston, 134–163. Ann Arbor: University of Michigan Pres.

Basic, E., and U. Rendtel. 2007. “Assessing the Bias due to Non-coverage of Residential Movers in the German Microcensus Panel: An Evaluation Using Data from the Socio-Economic Panel.” *AStA: Advances in Statistical Analysis* 91 (3): 311–334.

Battaglia, M. P., E. R. Zell, and P. L. Y. H. Ching. 1996. “Can Participating in a Panel Sample Introduce Bias into Trend Estimates?” In *Proceedings of the Survey Research Methods Section, 1010–1013.* Alexandria, VA: American Statistical Association. Retrieved from .

Baunach, D. M. 2011. “Decomposing Trends in Attitudes Toward Gay Marriage, 1988–2006.” *Social Science Quarterly* 92 (June): 346–363.

Beck, N., J. N. Katz, and R. Tucker. 1998. “Taking Time Seriously: Time-Series-Cross-Section Analysis with a Binary Dependent Variable.” *American Journal of Political Science* 42 (4): 1260–1288.

Behr, A. 2005. “Extent and Determinants of Panel Attrition in the European Community Household Panel.” *European Sociological Review* 21 (July): 489–512.

Berelson, B. R., P. F. Lazarsfeld, and W. N. McPhee. 1954. *Voting: A Study of Opinion Formation in a Presidential Campaign*. Chicago: University of Chicago Press.

Bhattacharya, D. 2008. “Inference in Panel Data Models under Attrition Caused by Unobservables.” *Journal of Econometrics* 144 (2): 430–446.

Biemer, P. P. 2011. “Total Survey Error: Design, Implementation, and Evaluation.” *Public Opinion Quarterly* 74 (February): 817–848.

Bishop, B. H. 2013. “Drought and Environmental Opinion A Study of Attitudes toward Water Policy.” *Public Opinion Quarterly* 77 (3): 798–810.

Blais, A., and A. Dobrzynska. 1998. “Turnout in Electoral Democracies.” *European Journal of Political Research* 33: 239–261.

Bradburn, N. M., L. J. Rips, and S. K. Shevell. 1987. “Answering Autobiographical Questions: The Impact of Memory and Inference on Surveys.” *Science* 236 (April): 157–161.

Brehm, J. 1993. *The Phantom Respondents: Opinion Surveys and Political Representation*. Ann Arbor: University of Michigan Press.

Brewer, P. R. 2008. "The Shifting Foundations of Public Opinion about Gay Rights." *Journal of Politics* 65 (July): 1208–1220.

Brown, A., A. Hale, and S. Michaud. 1998. "Use of Computer Assisted Interviewing in Longitudinal Surveys." In *Computer Assisted Survey Information Collection*, edited by M. P. Couper, R. P. Baker, J. Bethlehem, C. Z. F. Clark, J. Martin, W. L. Nicholls, II, J. M. O'Reilly, 185–200. New York: John Wiley & Sons.

Brown, C. H. 1990. "Protecting against Nonrandomly Missing Data in Longitudinal Studies." *Biometrics* 46 (1): 143–155.

Burden, B. C. 2000. "Voter Turnout and the National Election Studies." *Political Analysis* 8 (4): 389–398.

Callegaro, M. 2008. "Seam Effects in Longitudinal Surveys." *Journal of Official Statistics* 24 (3): 387–409.

Callegaro, M., R. Baker, J. Bethlehem, A. S. Goritz, J. A. Krosnick, and P. J. Lavrakas. 2014. "Online Panel Research." In *Online Panel Research: A Data Quality Perspective*, edited by Callegaro, M., R. Baker, J. Bethlehem, A. S. Goritz, J. A. Krosnick, and P. J. Lavrakas, 1–22. New York: John Wiley & Sons.

Cantor, D., and K. Levin. 1991. *Summary of Activities to Evaluate the Dependent Interviewing Procedure of the Current Population Survey*. Report submitted to the Bureau of Labor Statistics by Westat, Inc. (Contract No. J-9-J-8-0083).

Carmines, E. G., and R. A. Zeller. 1979. *Reliability and Validity Assessment*. Thousand Oaks, CA: Sage.

Carsey, T. M., and G. C. Layman. 2006. "Changing Sides or Changing Minds? Party Identification and Policy Preferences in the American Electorate." *American Journal of Political Science* 50 (April): 464–477.

Chaffee, S. H., and R. N. Rimal. 1996. "Time of Vote Decision and Openness to Persuasion." In *Political Persuasion and Attitude Change*, edited by D. Mutz, P. Sniderman, and R. Brody, 267–291. Ann Arbor: University of Michigan Press.

Chen, Q., A Gelman, M. Tracy, F. H. Norris, and S. Galea. 2015. "Incorporating the Sampling Design in Weighting Adjustments for Panel Attrition." *Statistics in Medicine*.

Clausen, A. R. 1968. “Response Validity: Vote Report.” *Public Opinion Quarterly* 32 (4): 588–606.

Clinton, J. D. 2001. “Panel Bias from Attrition and Conditioning: A Case Study of the Knowledge Networks Panel.” Unpublished manuscript, Stanford University. Retreived from .

Collins, C. 1975. “Comparison of Month-to-month Changes in Industry and Occupation Codes with Respondent’s Report of Change: CPS Job Mobility Study.” US Census Bureau, Response Research Staff Report (75-5).

Conrad, F. G., L. J. Rips, and S. S. Fricker. 2009. “Seam Effects in Quantitative Responses.” *Journal of Official Official Statistics* 25 (3): 339–361.

Converse, P. E. 1964. “The Nature of Belief Systems in Mass Publics.” In *Ideology and Discontent*, edited by David E. Apter, 206–261. Ann Arbor: University of Michigan Press.

Czajka, J. 1983. “Subannual Income Estimation.” In *Technical, Conceptual and Administrative Lessons of the Income Survey Development Program (ISDP)*, 87–97. New York: Social Science Research Council.

Daniels, M. J., and J. W. Hogan. 2008. *Missing Data in Longitudinal Studies: Strategies for Bayesian Modeling and Sensitivity Analysis*. New York: CRC Press.

Das, M., V. Toepoel, and A. van Soest. 2011. “Nonparametric Tests of Panel Conditioning and Attrition Bias in Panel Surveys.” *Sociological Methods & Research* 40 (January): 32–56.

Davenport, T. C., A. S. Gerber, D. P. Green, C. W. Larimer, C. B. Mann, and C. Panagopoulos. 2010. “The Enduring Effects of Social Pressure: Tracking Campaign Experiments Over a Series of Elections.” *Political Behavior* 32 (May): 423–430.

De Amici, D, C. Klersy, F. Ramajoli, L. Brustia, and P. Politi. 2000. “Impact of the Hawthorne Effect in a Longitudinal Clinical Study.” *Controlled Clinical Trials* 21 (April): 103–114.

Deng, Y., D. S. Hillygus, J. P. Reiter, Y. Si, and S. Zheng. 2013. “Handling Attrition in Longitudinal Studies: The Case for Refreshment Samples.” *Statistical Science* 28 (May): 238–256.

Diggle, P., and M. G. Kenward. 1994. “Informative Drop-out in Longitudinal Data Analysis.” *Applied Statistics* 43 (1): 49–93.

Duncan, G. J., and G. Kalton. 1987. "Issues of Design and Analysis of Surveys across Time." *International Statistical Review/Revue Internationale de Statistique* 55 (1): 97–117.

Finkel, S. E. 1995. *Causal Analysis with Panel Data*. Thousand Oaks, CA: Sage Publications.

Fitzgerald, J., P. Gottschalk, and R. Moffitt. 1998. *An Analysis of Sample Attrition in Panel Data: The Michigan Panel Study of Income Dynamics*. Technical report.

Fitzsimons, G. J., J. C. Nunes, and P. Williams. 2007. "License to Sin: The Liberating Role of Reporting Expectations." *Journal of Consumer Research* 34 (1): 22–31.

Fournier, P., R. Nadeau, A. Blais, E. Gidengil, and N. Nevitte. 2004. "Time-of-voting Decision and Susceptibility to Campaign Effects." *Electoral Studies* 23 (4): 661–681.

Frankel, L. L., and D. S. Hillygus. 2013. "Looking Beyond Demographics: Panel Attrition in the ANES and GSS." *Political Analysis* 22 (October): 336–353.

Gelman, 2007. "Struggles with Survey Weighting and Regression Modeling." *Statistical Science* 22 (2): 153–164.

Gelman, A., and J. Hill. 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge: Cambridge University Press.

Gerber, A. S., D. Karlan, and D. Bergan. 2009. "Does the Media Matter? A Field Experiment Measuring the Effect of Newspapers on Voting Behavior and Political Opinions." *American Economic Journal: Applied Economics* 1 (March): 35–52.

Granberg, D., and S. Holmberg. 1992. "The Hawthorne Effect in Election Studies: The Impact of Survey Participation on Voting." *British Journal of Political Science* 22 (02): 240–247.

Gray, R., P. Campanelli, K. Deepchand, and P. Prescott-Clarke. 1996. "Exploring Survey Non-response: The Effect of Attrition on a Follow-up of the 1984–85 Health and Life Style Survey." *The Statistician* 45 (2): 163–183.

Green, D. P., and E. Schickler. 1993. "Multiple-Measure Assessment of Party Identification." *Public Opinion Quarterly* 57 (4): 503.

Greenwald, A. G., C. G. Carnot, R. Beach, and B. Young. 1987. "Increasing Voting Behavior by Asking People If They Expect to Vote." *Journal of Applied Psychology* 72 (2): 315.

Groves, R. M. 2004. *Survey Errors and Survey Costs*. New York: John Wiley & Sons.

Groves, R. M., and L. Lyberg. 2011. “Total Survey Error: Past, Present, and Future.” *Public Opinion Quarterly* 74 (February): 849–879.

Groves, R. M., and M. P. Couper. 2012. *Nonresponse in Household Interview Surveys*. New York: John Wiley & Sons.

Groves, R. M., F. J. Fowler Jr., M. P. Couper, J. M. Lepkowski, E. Singer, and R. Tourangeau. 2009. *Survey Methodology*. 2nd ed. New York: Wiley.

Hausman, J. A., and D. A. Wise. 1979. “Attrition Bias in Experimental and Panel Data: The Gary Income Maintenance Experiment.” *Econometrica* 47 (2): 455–473.

Henderson, M., and D. S. Hillygus. Forthcoming. “Changing the Clock: The Role of Campaigns in the Timing of Vote Decision.” *Public Opinion Quarterly*.

Henderson, M., D. S. Hillygus, and T. Tompson. 2010. “‘Sour Grapes’ or Rational Voting? Voter Decision Making Among Thwarted Primary Voters in 2008.” *Public Opinion Quarterly* 74 (March): 499–529.

Hill, Daniel. 1987. “Response Errors around the Seam: Analysis of Change in a Panel with Overlapping Reference Periods.” In *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 210–215. Retreived from .

Hillygus, D. S., and S. Jackman. 2003. “Voter Decision Making in Election 2000: Campaign Effects, Partisan Activation, and the Clinton Legacy.” *American Journal of Political Science* 47 (4): 583–596.

Hillygus, D. S, N. Jackson, and M. Young. 2014. “Professional Respondents in Nonprobability Online Panels.” In *Online Panel Research: A Data Quality Perspective*, edited by M. Callegro, R. Baker, J. Bethlehem, A. S. Goritz, J. A. Krosnick, and P. J. Lavrakas, 219–237. New York: Wiley.

Hillygus, D. S., N. H. Nie, K. Prewitt, and H Pals. 2006. *The Hard Count: The Political and Social Challenges of Census Mobilization*. New York: Russell Sage Foundation.

Hirano, K, G. W. Imbens, G. Ridder, and D. B. Rebin. 1998. “Combining Panel Data Sets with Attrition and Refreshment Samples.” Working Paper 230, National Bureau of Economic Research.

Hirano, K., G. W. Imbens, G. Ridder, and D. B. Rubin. 2001. “Combining Panel Data Sets with Attrition and Refreshment Samples.” *Econometrica* 69 (6): 1645–1659.

Honaker, J., and G. King. 2010. "What to Do about Missing Values in Time-Series Cross-Section Data." *American Journal of Political Science* 54 (April): 561–581.

Hsiao, C. 2003. *Analysis of Panel Data*. 2nd ed. Cambridge: Cambridge University Press.

Iyengar, S., G. Sood, and Y. Lelkes. 2012. "Affect, Not Ideology: A Social Identity Perspective on Polarization." *Public Opinion Quarterly* 76 (September): 405–431.

Jackle, A., and P. Lynn. 2004. "Dependent Interviewing and Seam Effects in Work History Data." ISER Working Paper 2004-24, Institute for Social and Economic Research, University of Essex, Colchester.

Jackson, N. 2011. *Questionnaire Design Issues in Longitudinal and Repeated Cross-Sectional Surveys*. Report of the Duke Initiative on Survey Methodology Workshop on Questionnaire Design Issues in Longitudinal and Repeated Cross-Sectional Surveys, February 18.

Kalton, G., and C. F. Citro. 1993. *The Future of the Survey of Income and Program Participation*. Washington, D.C.: National Academy Press.

Kalton, G., and M. E. Miller. 1991. "The Seam Effect with Social Security Income in the Survey of Income and Program Participation." *Journal of Official Statistics* 7 (2): 235–245.

Kasprzyk, D., G. Duncan, G. Kalton, and M. P. Singh. 1989. *Panel Surveys*. New York: Wiley.

Katz, 1971. "Platforms & Windows: Broadcasting's Role in Election Campaigns." *Journalism & Mass Communication Quarterly* 48 (2): 304–314.

Kenward, M. G. 1998. "Selection Models for Repeated Measurements with Non-random Dropout: An Illustration of Sensitivity." *Statistics in Medicine* 17 (23): 2723–2732.

Kenward, M. G., G. Molenberghs, and H. Thijs. 2003. "Pattern-mixture Models with Proper Time Dependence." *Biometrika* 90 (1): 53–71.

Kogen, L., and J. A. Gottfried. 2012. "I Knew It All Along! Evaluating Time-of-decision Measures in the 2008 US Presidential Campaign." *Political Behavior* 34 (4): 719–736.

Kraut, R. E., and J. B. McConahay. 1973. "How Being Interviewed Affects Voting: An Experiment." *Public Opinion Quarterly* 37 (3): 398–406.

Kristman, V. L., M. Manno, and P. Côté. 2005. "Methods to Account for Attrition in Longitudinal Data: Do They Work? A Simulation Study." *European Journal of Epidemiology* 20 (8): 657–662.

Kruse, Y., M. Callegaro, J. M. Dennis, C. DiSogra, S. Subias, M. Lawrence, and T. Thompson. 2009. "Panel Conditioning and Attrition in the AP-Yahoo! News Election Panel Study." Presented at the Annual Meeting of the American Association for Public Opinion Research. Hollywood, FL Retreived from .

Landsberger, H. A. 1958. *Hawthorne Revisited: Management and the Worker, Its Critics, and Developments in Human Relations in Industry*. Ithaca: Cornell University Press.

Laurie, H. 2007. "Minimizing Panel Attrition." In *Handbook of Longitudinal Research: Design, Measurement, and Analysis*, edited by Scott Menard, 167–184. Burlington, MA: Elsevier.

Laurie, H., and P. Lynn. 2009. "The Use of Respondent Incentives on Longitudinal Surveys." In *Methodology of Longitudinal Surveys*, edited by Peter Lynn, 205–234. Chichester, UK: John Wiley & Sons.

Lazarsfeld, P. F. 1940. "'Panel' Studies." *Public Opinion Quarterly* 4 (1): 122–128.

Lazarsfeld, P. F., B. Berelson, and H. Gaudet. 1948. *The People's Choice: How the Voter Makes Up His Mind in a Presidential Campaign*. New York: Columbia University Press.

Lemaitre, G. 1992. *Dealing with the Seam Problem for the Survey of Labour and Income Dynamics*. Ottawa: Statistics Canada.

Lenz, G. S. 2009. "Learning and Opinion Change, Not Priming: Reconsidering the Priming Hypothesis." *American Journal of Political Science* 53 (4): 821–837.

Lepkowski, J. M., and M. P. Couper. 2001. "Nonresponse in the Second Wave of Longitudinal Household Surveys." In *Survey Nonresponse*, edited by R. M. Groves, D. A. Dillman, J. L. Eltinge, and R. J. Little, 259–272. New York: Wiley and Sons.

Levav, J., and G. J. Fitzsimons. 2006. "When Questions Change Behavior: The Role of Ease of Representation." *Psychological Science* 17 (March): 207–213.

Lipps, O. 2007. "Attrition in the Swiss Household Panel." *Methoden–Daten–Analysen* 1 (1): 45–68.

Little, R. J. A. 1993. "Pattern-mixture Models for Multivariate Incomplete Data." *Journal of the American Statistical Association* 88 (421): 125–134.

Little, R. J. A, and D. B. Rubin. 2002. *Statistical Analysis with Missing Data*. New York: Wiley.

Little, R. J. A, and Y. Wang. 1996. "Pattern-mixture Models for Multivariate Incomplete Data with Covariates." *Biometrics* 58 (1): 98–111.

Loosveldt, G., and A. Carton. 1997. "Evaluation of Nonresponse in the Belgian Election Panel Study '91–'95." In *Proceedings of the Survey Research Methods Section, American Statistical Association*, 1017–1022. Retreived from .

Loosveldt, G., J. Pickery, and J. Billiet. 2002. "Item Nonresponse as a Predictor of Unit Nonresponse in a Panel Survey." *Journal of Official Statistics* 18 (4): 545–558.

Lynn, P., and E. Sala. 2006. "Measuring Change in Employment Characteristics: The Effects of Dependent Interviewing." *International Journal of Public Opinion Research* 18 (4): 500–509.

Lynn, P., N. Buck, J. Burton, A. Jackle, and H. Laurie. 2005. "A Review of Methodological Research Pertinent to Longitudinal Survey Design and Data Collection." ISER Working Paper 2005-29, Institute for Social and Economic Research, University of Essex, Colchester.

Mann, C. B. 2005. "Unintentional Voter Mobilization: Does Participation in Preelection Surveys Increase Voter Turnout?" *ANNALS of the American Academy of Political and Social Science* 601 (1): 155–168.

Marquis, K. H, and J. C. Moore. 1989. "Some Response Errors in SIPP—With Thoughts About Their Effects and Remedies." In *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 381–386. Retreived from .

Martini, A. 1989. "Seam Effect, Recall Bias, and the Estimation of Labor Force Transition Rates from SIPP." In *Proceedings of the Survey Research Methods Section, American Statistical Association*,. 387–392. Retreived from .

Menard, S. 2002. *Longitudinal Research*. Vol. 76. 2nd ed. Thousand Oaks: Sage Publications.

Menard, S., ed. 2007. *Handbook of Longitudinal Research: Design, Measurement, and Analysis*. Burlington, MA: Elsevier.

Moore, J., N. Bates, J. Pascale, and A. Okon. 2009. "Tackling Seam Bias Through Questionnaire Design." In *Methodology of Longitudinal Surveys*, edited by Peter Lynn, 72–92. New York: John Wiley & Sons.

Olson, K., and L. Witt. 2011. “Are We Keeping the People Who Used to Stay? Changes in Correlates of Panel Survey Attrition Over Time.” *Social Science Research* 40 (4): 1037–1050.

Pasek, J., A. Tahk, Y. Lelkes, J. A. Krosnick, B. K. Payne, O. Akhtar, and T. Tompson. 2009. “Determinants of Turnout and Candidate Choice in the 2008 US Presidential Election Illuminating the Impact of Racial Prejudice and Other Considerations.” *Public Opinion Quarterly* 73 (5): 943–994.

Plumb, E. 1986. “Validation of Voter Recall: Time of Electoral Decision Making.” *Political Behavior* 8 (4): 302–312.

Polivka, A. E., and J. M. Rothgeb. 1993. “Redesigning the CPS Questionnaire.” *Monthly Labor Review* September: 10–28.

Rips, L. J., F. G. Conrad, and S. S. Fricker. 2003. “Straightening the Seam Effect in Panel Surveys.” *Public Opinion Quarterly* 67 (4): 522–554.

Scharfstein, D. O., A. Rotnitzky, and J. M. Robins. 1999. “Adjusting for Nonignorable Drop-out Using Semiparametric Nonresponse Models.” *Journal of the American Statistical Association* 94 (448): 1096–1120.

Schluchter, M. D. 1992. “Methods for the Analysis of Informatively Censored Longitudinal Data.” *Statistics in Medicine* 11 (14–15): 1861–1870.

Schoeni, R. F., F. Stafford, K. A. McGonagle, and P. Andreski. 2013. “Response Rates in National Panel Surveys.” *Annals of the American Academy of Political and Social Science* 645 (January): 60–87.

Si, Y., J. P. Reiter, and D. S. Hillygus. 2014. “Semi-parametric Selection Models for Potentially Non-ignorable Attrition in Panel Studies with Refreshment Samples.” *Political Analysis* (June): 1–21.

Simmons, C. J., B. A. Bickart, and J. G. Lynch Jr. 1993. “Capturing and Creating Public Opinion in Survey Research.” *Journal of Consumer Research* 20 (2): 316–329.

Singer, J. D., and J. B. Willett. 2003. *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*. New York: Oxford University Press.

Smith, J. K., A. S. Gerber, and A. Orlich. 2003. “Self-Prophecy Effects and Voter Turnout: An Experimental Replication.” *Political Psychology* 24 (3): 593–604.

Smith, T. W. 2005. “The Laws of Studying Societal Change.” General Social Survey Social Change Report, No. 50.

Smith, T. W., and J. Son. 2010. “An Analysis of Panel Attrition and Panel Change on the 2006-2008 General Social Survey Panel.” General Social Survey Methodological Report, No. 118.

Sturgis, P., N. Allum, and I. Brunton-Smith. 2009. “Attitudes Over Time: The Psychology of Panel Conditioning.” In *Methodology of Longitudinal Surveys,* edited by P. Lynn, 113–126. Chichester, UK: John Wiley & Sons.

Tourangeau, R. 2003. *Recurring Surveys: Issues and Opportunities*. Report to the National Science Foundation on a workshop held on March 28–29. Arlington, VA: National Science Foundation.

Tourangeau, R., L. J. Rips, and K. Rasinski. 2000. *The Psychology of Survey Response*. New York: Cambridge University Press.

Traugott, M. W., and J. P. Katosh. 1979. “Response Validity in Surveys of Voting Behavior.” *Public Opinion Quarterly* 43 (3): 359.

Traugott, S., and G. Morchio. 1990. “Assessment of Bias Due to Attrition and Sample Selection in the NES 1989 Pilot Study.” ANES Technical Report, Center for Political Studies, University of Michigan, Ann Arbor.

Traugott, S., and S. J. Rosenstone. 1994. *Panel Attrition Among the 1990–1992 Panel Respondents*. Technical Report, Center for Political Studies.

Uhrig, S. C. N. 2008. “The Nature and Causes of Attrition in the British Household Panel Study.” ISER Working Paper 2008-05, Institute for Social and Economic Research, University of Essex, Colchester.

Vandecasteele, L., and A. Debels. 2006. “Attrition in Panel Data: The Effectiveness of Weighting.” *European Sociological Review* 23 (December): 81–97.

Vicente, P. C. 2014. “Is Vote Buying Effective? Evidence from a Field Experiment in West Africa.” *Economic Journal* 124 (574): F356–F387.

Visser, P. S., J. A. Krosnick, P. J. Lavrakas, and N. Kim. 2014. “Survey Research.” In *Handbook of Research Methods in Social Psychology*, 2nd ed., edited by H. T. Reis and C. M. Judd, 223–252. Cambridge: Cambridge University Press.

Voogt, R. J. J. 2005. “An Alternative Approach to Correcting Response and Nonresponse Bias in Election Research.” *Acta Politica* 40 (1): 94–116.

Voogt, R. J. J., and W. E. Saris. 2003. "To Participate or Not to Participate: The Link between Survey Participation, Electoral Participation, and Political Interest." *Political Analysis* 11 (2): 164–179.

Voogt, R. J. J., and H. Van Kempen. 2002. "Nonresponse Bias and Stimulus Effects in the Dutch National Election Study." *Quality and Quantity* 36 (4): 325–345.

Warren, J. R., and A. Halpern-Manners. 2012. "Panel Conditioning in Longitudinal Social Science Surveys." *Sociological Methods & Research* 41: 491–534.

Watson, N., and M. Wooden. 2009. "Identifying Factors Affecting Longitudinal Survey Response." In *Methodology of Longitudinal Surveys*, edited by Peter Lynn, 157–183. Chichester, UK: John Wiley & Sons.

Wawro, G. 2002. "Estimating Dynamic Panel Data Models in Political Science." *Political Analysis* 10 (1): 25–48.

Weisberg, H. F. 2005. *The Total Survey Error Approach: A Guide to the New Science of Survey Research*. Chicago: University of Chicago Press.

Westat. 2009. "SIPP Sample Design and Interview Procedures." In *Survey of Income and Program Participation Users' Guide*, 1–25. Rockville, MD. Retreived from .

Yalch, R. F. 1976. "Pre-election Interview Effects on Voter Turnout." *Public Opinion Quarterly* 40 (3): 331–336.

Yeager, D. S., J. A. Krosnick, L. Chang, H. S. Javitz, M. S. Levendusky, A. Simpser, and R. Wang. 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.

Zabel, J. E. 1998. "An Analysis of Attrition in the Panel Study of Income Dynamics and the Survey of Income and Program Participation with application to a model of labor market behavior." *Journal of Human Resources* 33 (2): 479–506.

Notes:

([1]) Although most online survey panels are nonprobability panels, in which panelists have opted-in to the panel, there are limited examples of online probability survey panels, such as the RAND American Life Panel, AmeriSpeaks, and GfK Knowledge Networks.

([2]) A related difference is in the definition of attrition. Some designs allow individuals who fail to respond to one wave to return to subsequent waves (temporary attrition), while other designs would consider those individuals permanent attriters. Internet panel studies that rely on an online panel of respondents are especially likely to use the former design, as it is nearly costless to invite former attriters into subsequent waves.

([3]) To be sure, the exact nature of the relationship between the intervention and the data collection can affect the strength of the causal claims. Generally speaking, data collected closer to the intervention give greater confidence that any observed changes are the result of the intervention rather than confounding factors.

([4]) The weights provided often account for both unequal probabilities of selection in the sampling design as well as unit nonresponse. As such, new weights are typically provided for each wave to account for sample attrition.

([5]) To be sure, some researchers have found minimal attrition bias (Bartels 1999; Clinton 2001; Kruse et al. 2009). Most critical, of course, is that such an evaluation be conducted, since the extent of attrition bias can vary across different outcomes.

([6]) In using any alternative approach to panel attrition correction, it remains important to account for the sampling design in making inferences. If the survey firm does not provide all variables related to the sampling design (e.g., geographic clusters), researchers can use the sampling design weights or wave 1 survey weights to make the necessary adjustments.

([7]) Of course, even nonpanel studies must also confront the possibility that simply the act of measuring social phenomena can sometimes change the object under investigation—the classic Hawthorne effect (e.g., Landsberger 1958).

([8]) Das, Toepoel, and van Soest (2011) offer one such approach that relies on a nonparametric test for estimating separate attrition and conditioning effects.

([9]) No doubt it does not help that researchers tend to ask the exact same political knowledge questions across different studies.

([10]) Notable exceptions include Warren and Halpern-Manners (2012); Sturgis, Allum, and Brunton-Smith (2009); and Das, Toepoel, and van Soest (2011).

([11]) Interested readers may want to consult the resources available at http://dism.ssri.duke.edu/question_design.php.

D. Sunshine Hillygus

D. Sunshine Hillygus (Ph.D., Stanford University) is an Associate Professor of political science and Director of the Initiative on Survey Methodology at Duke University. Her research and teaching specialties include public opinion, political behavior, survey research, campaigns and elections, and information technology and society. She is co-author of *The Hard Count: The Social and Political Challenges of the 2000 Census* (Russell Sage Foundation, 2006) and *The Persuadable Voter: Wedge Issues in Political Campaigns* (Princeton University Press, 2008).

Steven Snell

Postdoctoral Fellow at the Duke Initiative on Survey Methodology, Duke University