A Methodological Framework for the Analysis of Panel Conditioning Effects

Ruben L. Bach
*University of Mannheim**

Abstract

Panel conditioning refers to the phenomenon whereby respondents' attitudes, behavior, reporting of behavior and/or knowledge are changed by repeated participation in a panel survey. Uncovering such effects, however, is difficult due to three major methodological challenges. First, researchers need to disentangle changes in behavior from changes in the reporting of behavior as panel conditioning may result in both, even at the same time and in opposite directions. Second, the identification of the *causal* effect of panel participation on the various forms of change mentioned above is complicated as it requires comparisons of panel respondents with control groups of people who have not been interviewed before. Third, other sources of error in (panel) surveys may easily be mistaken for panel conditioning if not properly accounted for. Such error sources are panel attrition, mode and interviewer effects. I review the challenges mentioned above in detail and provide a methodological framework for the analysis of panel conditioning effects by identifying the strengths and weaknesses of the various designs that researchers have developed to address the challenges. I conclude with a discussion of a future research agenda on panel conditioning effects in longitudinal surveys.

*Address: Lehrstuhl für Statistik und sozialwissenschaftliche Methodenlehre, Universität Mannheim, 68131 Mannheim, Germany. E-mail: r.bach@uni-mannheim.de Phone: (+49) 621 181 3288

1 Introduction

Researchers working with data collected through surveys usually assume that the act of measuring does not affect what is being measured. Yet, as early as in 1940, Lazarsfeld (1940:128) noted that "the big problem, as yet unsolved, is whether (...) interviews are likely, in themselves, to influence a respondent's opinion". Since then, researchers have spent decades analyzing how repeated measurement can result in unintended changes in respondents' attitudes, knowledge, behavior and reports of behavior. These changes are unintended insofar as researchers usually assume that any intra-individual over-time change found in their data reflects a 'true' change in a person's attitudes or behavior, and that these changes would have also occurred had the respondent not participated in the survey (Warren and Halpern-Manners 2012). Yet, when the sheer act of measuring affects what is being measured, this assumption is no longer satisfied and researchers risk mis-characterizing the existence, magnitude, and correlates of changes across survey waves in respondents' attitudes and behaviors (Clinton 2001; Halpern-Manners, Warren, and Torche 2014).

Scholars from various disciplines have spent much time on uncovering these so-called *panel conditioning* effects and considerable progress has been made regarding the theoretical mechanisms leading to panel conditioning effects (e.g., Bergmann and Barth forthcoming). Still, there are three methodological challenges that have plagued researchers of panel conditioning ever since and many studies fail to address them adequately. First, panel conditioning may influence both respondents' reporting of behavior and the actual behavior itself, even at the same time and in opposite directions. Therefore, researchers need to disentangle changes in respondents' behavior from changes in respondents' reporting of behavior to get unbiased estimates of either one.[1] For example, respondents may remember from participation in prior waves of a panel survey how the interview is structured and how they can speed through the interview by taking shortcuts. I call this form of panel conditioning changes-in-reporting panel conditioning. At the same time, respondents' actual behavior may be conditioned by participation in the survey. For example, repeatedly answering questions may work as a stimulus that affects respondents' subsequent behavior. I call this form of panel conditioning changes-in-behavior panel conditioning. To get unconfounded estimates of either form of panel conditioning, it is crucial to clearly distinguish between the two.

The second challenge refers to the availability of control group data. That is, researchers of panel conditioning (at least implicitly) wish to estimate the *causal effect* of panel survey participation on respondents' attitudes, knowledge, reporting of behavior or actual behavior (Bach and Eckman 2018). To estimate a causal or treatment effect, one usually compares treated cases against untreated, i.e., control cases in (quasi) experimental designs. In the panel conditioning framework, the former are those cases of individuals who responded to two or more waves of a panel survey, while the latter are cases who have not been interviewed or have been interviewed only once. To identify an unbiased treatment effect, assignment to treatment and control is, ideally, random because randomization balances (in expectation) all differences (e.g., socio-demographic differences) between the treatment and the control group. As a result, the only remaining difference between the two groups is that one receives the treatment, and the other one does not. In large-scale social science panel surveys, the kind of surveys that social scientists frequently use for substantive research, methodological experiments with random assignment of cases to treatment or control (i.e., participation in a survey several times or only once) are hardly ever (intentionally) implemented (Warren and Halpern-Manners 2012). That is, data for treated cases (panel survey respondents) are usually easily available from the panel survey itself; control group data, however, are much harder to find. As a result, the identification of the causal effect is difficult without further assumptions.

The third challenge is that, even when experimental manipulations can be implemented, researchers need to account for confounding sources of error. Error sources that may confound estimates of panel conditioning are longitudinal nonresponse (i.e., panel attrition), mode effects and interviewer effects. While attrition affects the composition of a longitudinal survey sample over time, mode effects and interviewer effects confound the level of measurement error. All of them may result in changes in an outcome variable over time that are not due to a 'true' change in a person's attitudes or behavior and may thus easily be mistaken for panel conditioning. As a result, estimates of panel conditioning, even when based on experimental manipulations, will be biased if these error sources are not properly accounted for.

Many previous studies did not address these challenges adequately, calling for a systematic review of the challenges *and* methods developed to tackle them. In the remainder of this article, I briefly summarize the different forms of panel conditioning effects and hypotheses why panel conditioning can occur in social science longitudinal surveys. In the main part of this paper, I elaborate on the methodological challenges mentioned above. I demonstrate for each challenge why it requires researchers' special attention when studying panel conditioning. By systematically discussing the strengths and weaknesses of the different research designs developed to tackle the challenges, I provide a methodological framework for the analysis of panel conditioning effects in social science longitudinal surveys. I conclude with a discussion of research questions that the literature has not answered satisfactorily yet and questions that only recently emerged due to new developments in data collection techniques.

2 Various forms of panel conditioning

Panel conditioning has been studied for some decades. However, the literature often lacks a clear distinction between the various forms of panel conditioning. Moreover, the underlying mechanisms as well as their theoretical foundations are often unclear, although recent work has made considerable progress regarding the circumstances under which panel conditioning is likely to affect respondents' attitudes, knowledge, behavior and reporting of behavior (Cantor 2010; Warren and Halpern-Manners 2012) and the psychological processes leading to panel conditioning (Bergmann and Barth forthcoming). In this section, I distinguish between the different forms of panel conditioning, briefly review hypotheses that explain when the different forms may arise and present empirical evidence.

2.1 Changes-in-reporting panel conditioning

Changes-in-reporting panel conditioning refers to the phenomenon whereby panel survey participation influences the way respondents report over time. This form of conditioning can result in respondents becoming either better or worse reporters over the waves of a panel survey. Respondents can become better reporters because they become more trusting of the survey experience. For example, they might feel more comfortable and more motivated giving less socially desirable, but more accurate answers. That is, increasing trust in the confidentiality of their responses and reduced suspicion towards the interviewer leads to a reduction in measurement error over time as respondents report more honestly. Similarly, respondents may gain a better understanding of the meaning of the questions or may become more convinced of the importance of their answers for the survey and report fewer "don't know"-responses. Thus, respondents become more comfortable with the interviewing process leading them to answer the questions more accurately (Bailar 1989; Waterton and Lievesley 1989; Van der Zouwen and Van Tilburg 2001; Warren and Halpern-Manners 2012).

The literature documents several examples of changes-in-reporting panel conditioning resulting in respondents becoming better reporters. Halpern-Manners et al. (2014) find that respondents report more honestly regarding having previously driven drunk or having stolen something of little value in later waves of a panel survey. Similarly, Waterton and Lievesley (1989) report declines in social desirability bias (reporting racial prejudice) over time and fewer reports of "don't knows". Van Landeghem (2012) and Chadi (2013) both find that increased trust towards the interviewer leads to more honest reporting of life satisfaction. Furthermore, Kroh, Winter, and Schupp (2016) find that the reliability of person fit measures increases with every wave of a panel survey because respondents get a better understanding of the interview process. Angel, Heuberger, and Lamei (2017) report that both over- and under-reporting of household income decreases over panel waves as respondents feel more comfortable reporting their income honestly and prepare for the interviews to provide more accurate responses to the survey.

On the other hand, respondents may become worse reporters over the waves of a panel survey (Bailar 1989; Waterton and Lievesley 1989; Williams, Block, and Fitzsimons 2006; Warren and Halpern-Manners 2012). Respondents can learn from participation in previous waves of a survey how the interview is structured and how its burden or length can be reduced by taking shortcuts in the interview. For example, respondents of a labor market survey might learn that reporting to be employed leads to additional follow-up questions regarding the employment and may report to be unemployed in later waves to skip the follow-up questions. As a result, measurement error will increase over the waves of a panel survey. Declining data quality over time may also result from increasing levels in social desirability bias. That is, when questions refer to socially non-normative behavior, respondents may be confronted with a conflict between their behavior and society's norms. To avoid such cognitive dissonance, respondents may resort to reporting behavior closer to society's norms.

Examples of respondents becoming worse reporters due to panel conditioning are numerous. Schonlau and Toepoel (2015), for example, show that straightlining, i.e., the tendency to give the same responses to a series of questions with identical answer choices, increases with respondents' panel experience. Several other studies find increases in social desirability bias over time, e.g., in reports of children's vaccination status (Battaglia, Zell, and Ching 1996) or in reports of exercising (Williams et al. 2006). Furthermore, it is well documented that the Current Population Survey (CPS) underestimates unemployment rates because respondents misreport unemployment to skip follow-up questions in later waves of the survey (Hansen et al. 1955; Bailar 1975; 1989; Shack-Marquez 1986; Solon 1986; Shockey 1988; Halpern-Manners and Warren 2012).[2] Other studies find respondents taking shortcuts in later waves of a panel survey regarding reports of home alteration and repair jobs (Neter and Waksberg 1964), functional limitations among elderly people (Mathiowetz and Lair 1994), substance abuse (Torche, Warren, and Halpern-Manners 2012), every day personal hygiene product use (Nancarrow and Cartwright 2007) and consumption of purchased and own-produced goods (Schündeln 2017). Similar effects, called *survey conditioning*, may also occur in cross-sectional surveys (Duan et al. 2007). For example, respondents learn to misreport to filter questions in order to skip follow-up questions and speed through an interview (e.g., Duan et al. 2007; Kreuter et al. 2011; Eckman et al. 2014; Eckman and Kreuter forthcoming). However, there are also studies that find no increases or decreases in misreporting over time in panel surveys (e.g., Cohen and Burt 1985; Halpern-Manners et al. 2014; Struminskaya 2016; Bach and Eckman forthcoming).

2.2 Changes-in-behavior panel conditioning

Panel participation may also result in changes in actual behavior over time. The common explanation for this type of panel conditioning is the cognitive stimulus approach. It holds that repeatedly being asked the same questions causes respondents to become more aware of the topic of the survey, raises their consciousness of the issues and motivates them to engage in the behavior under study (Sturgis, Allum, and Brunton-Smith 2009; Waterton and Lievesley 1989; Warren and Halpern-Manners 2012). For example, the participation in a pre-election poll may increase voter turnout because being asked about voting intentions increases the likelihood of actual voting.

Furthermore, being asked knowledge questions might stimulate respondents who do not know the answer to look it up after the interview. E.g., respondents of a political science survey who are not aware of a new law might look it up after the survey. In a follow-up interview, they would then know the correct answer. Their knowledge would not have changed, however, had they not participated in the interview. Similarly, survey questions can serve to provide information about behaviors that respondents were not aware of. If respondents had not participated in a survey, they would not have known of the behavior under study and could not have engaged in it (Halpern-Manners and Warren 2012). A survey that asks for participation in a specific cancer screening measure, for example, might inform a respondent who had not known of the measure before about the existence of this measure. Therefore, had the respondent not participated in the survey, she would not have known of that measure.

Changes in behavior may also arise from survey questions dealing with socially non-normative or stigmatized behavior or attitudes. As respondents are confronted with a conflict between their attitudes or behavior and society's norms, they bring their future behavior or attitudes in line with society's norms (Williams et al. 2006).[3] In a survey about alcohol consumption, for example, dissonance between respondents' heavy-drinking habits and society's norms of modest alcohol consumption may stimulate respondents to reconsider the amount and frequency of alcohol consumption and thereby lead them to drink less by the time of a follow-up survey (Warren and Halpern-Manners 2012).

Panel participation might also lead to 'real' changes in an attitude. When attitudes are less crystallized, responding to questions about an attitude can sometimes change it. Moreover, when respondents lack crystallized attitudes about a specific topic, they will nonetheless offer a response to a question about that attitude. This response may start a cognitive process that will lead to a change in the attitude by the time of the next wave (Waterton and Lievesley 1989; Sturgis et al. 2009; Warren and Halpern-Manners 2012). For example, respondents of a survey on same-sex marriage may not have an opinion regarding this topic, but might provide an opinion nonetheless. Moreover, responding to a question regarding their views on this topic may also start a cognitive process that will lead them to form an opinion.

Changes-in-behavior panel conditioning is reported in studies of voting behavior, where participation in a pre-election survey leads to increases in voter turnout in upcoming elections (Clausen 1969; Kraut and McConahay 1973; Yalch 1976; Traugott and Katosh 1979; Granberg and Holmberg 1992). However, not all studies detect this effect (Smith, Gerber, and Orlich 2003). Other behaviors affected by panel conditioning are water treatment product use (Zwane et al. 2011); purchases of health insurance (Zwane et al. 2011), automotive services (Borle et al. 2007), automobiles (Morwitz, Johnson, and Schmittlein 1993; Chandon, Morwitz, and Reinartz 2005) and computers (Morwitz et al. 1993); saving for retirement (Crossley et al. 2017); cheating in exams (Spangenberg and Obermiller 1996), use of contraceptives (Axinn, Jennings, and Couper 2015); perceptions of marital quality (Veroff, Hatchett, and Douvan 1992) and participation in labor market programs (Bach and Eckman 2018). Regarding panel conditioning leading to changes in knowledge, increases in knowledge are reported for bacteria and for pension schemes (Das, Toepoel, and van Soest 2007), contraception methods (Coombs 1973) and vaccination programs (Battaglia et al. 1996). Real changes in attitudes due to panel conditioning are reported for about half of all items tested by Warren and Halpern-Manners (2012) in the General Social Survey in the U.S. and the Socio-Economic Panel in Germany. The authors suspect that responding to attitudinal questions may have created attitudes even among those who did not hold attitudes before. Similar results are reported by Sturgis et al. (2009) for the British Household Panel Survey.

To sum up, theory and evidence from numerous studies suggest that repeated participation in panel surveys can lead to changes in respondents' reports of behavior and attitudes as well as to changes in their actual behavior, attitudes and knowledge. Furthermore, all forms of panel conditioning are more likely to occur the shorter the interval between the waves of a panel study is and the more often respondents are interviewed (Halpern-Manners et al. 2014; Van Landeghem 2012). Identifying a panel conditioning effect of any form, however, is difficult, as we will see in the next section.

3 Methodological challenges

Three major methodological challenges (disentangling the different forms of panel conditioning, finding control group data and accounting for confounding sources of error) complicate the identification of panel conditioning effects in longitudinal data. I review them in detail in this section. Moreover, I discuss the various designs that previous studies have developed to address them. Table 1 summarizes the methodological challenges and the approaches developed to tackle them.

3.1 Challenge 1: Disentangling the different forms of panel conditioning

Insert Table 1 about here.

Repeated participation in a panel survey can result in both changes-in-reporting and changes-in-behavior as I have laid out above.[4] Surveys, however, usually measure behavior through respondents' self-reports only. Thus, working with survey data only, i.e., with respondents' self-reported behavior, it is difficult for researchers to tell whether panel conditioning affects respondents' behavior or reports of behavior. To further complicate things, panel conditioning may even affect both at the same time. Bach and Eckman (2018), for example, discuss a hypothetical example where respondents of a panel survey on recycling and environmental behavior over-report recycling in early waves due to social desirability, but report more honestly in later waves as they become more comfortable and trusting of the interview process. At the same time, however, repeatedly asking respondents about recycling and environmental behavior may work as a stimulus and increase respondents' awareness of the importance of recycling and thereby lead to changes in their behavior. Researchers seeking to understand panel conditioning need to be aware of the various ways that panel conditioning might affect behavior and the reporting of it at the same time. It is therefore crucial to clearly distinguish between the two forms (Waterton and Lievesley 1989; Van der Zouwen and Van Tilburg 2001).

Luckily, disentangling changes-in-reporting from changes-in-behavior is straightforward when records that are unaffected by respondents' reporting are available, for example, from administrative records. Using such data, researchers can study changes-in-behavior by linking both data sources and analyzing respondents' behavior in the administrative records only. Such administrative data (e.g., tax records or insurance records), are themselves not free of error (e.g., Oberski et al. 2017). Yet, they can be considered the gold standard for analyzing changes-in-behavior panel conditioning because they are usually generated independently of respondents' reporting.

Linked survey-administrative records can also be used to study changes-in-reporting, for example, using them to validate survey responses. If measurement error, the deviation of a survey report from the true value (i.e., the value recorded in the validation data) changes over time in a panel survey, we can interpret such results as evidence for changes-in-reporting panel conditioning.

Unfortunately, however, in most scenarios, researchers will not have external validation records at hand, thereby making disentanglement of the two forms of panel conditioning difficult or impossible. In a few cases, theoretical considerations may allow researchers to conclude that only one of the two forms of panel conditioning is possible. Warren and Halpern-Manners (2012), for example, conclude that respondents of the CPS who report to be unemployed in the first wave, but report to be out of the labor force in subsequent waves, do so to avoid follow-up questions and *not* because participating in the survey made them actually more likely to leave the labor force. Thus, some scenarios may allow researchers to preclude one or the other form of panel conditioning.[5]

It is likely that the lack of administrative records explains why many studies do not consider differences between changes-in-reporting and changes-in-behavior panel conditioning. If researchers do not clearly disentangle the two forms, they risk making flawed statements about either form. In the worst case, the two forms cancel each other out, leading researchers to conclude that panel conditioning is not present.

Only a few studies explicitly acknowledge that panel survey participation can lead to both forms of change. Angel et al. (2017), studying the development of misreports of household income over time (see Section 2.1), for example, link respondents' survey reports of their household income to register data containing the same information. Using the income from the register as validation records, they are able to study how panel survey participation leads to changes-in-reporting. Similarly, Yan and Eckman (2012) link survey responses of a large-scale labor market panel survey to administrative labor market records to disentangle the two forms of panel conditioning, demonstrating that both take place at the same time. Crossley et al. (2017) and Bach and Eckman (2018) are two examples of studies that explicitly study changes-in-behavior panel conditioning by observing the behavior of respondents of large-scale social science panel surveys in administrative records. Similar approaches are applied by the other studies cited in Section 2.2, though many of them rely on small datasets collected among students (e.g., Spangenberg and Obermiller 1996) or marketing data (e.g., Chandon et al. 2005) and do not use the kind of longitudinal data that builds the basis for substantive research in the social sciences.

To sum up, disentangling the two forms of panel conditioning is difficult when administrative data or validation records of reported behavior do not exist or cannot be linked. In these cases, no clear statement can be made about either form of panel conditioning. Recent studies (e.g., Crossley et al. 2017; Angel et al. 2017), however, have made considerable methodological advances in disentangling the different forms of panel conditioning.

3.2 Challenge 2: Finding control groups

The second challenge (finding control group data) can be best understood in terms of the counterfactual causal model (e.g., Holland 1986). Researchers analyzing panel conditioning usually wish to study the *causal* effect of panel survey participation. Thus, we can think of panel conditioning as the effect of a treatment, $D \in [0, 1])$ (panel survey participation) on some outcome, Y (e.g., the value of a reported survey variable). To define the treatment effect, τ, define two potential outcomes, following Rubin (1974; 1978): $Y_{i,1}$ is the outcome that occurs when a case i receives treatment (participates in the survey), $Y_{i,0}$, by contrast, is the outcome when a case does not receive treatment, i.e., is in the control condition (does not participate in the survey). Using these potential outcomes, define the individual treatment effect as $\tau_i = Y_{i,1} - Y_{i,0}$. The *fundamental problem of causal inference*, however, is that we observe only $Y_i = D_i Y_{i,1} + (1 - D_i) Y_{i,0}$ for any individual (Holland 1986). In other words, we observe only one of the two potential outcomes because a person either participates in a survey or does not. While this problem retains us from estimating an individual treatment effect, we may still estimate an average treatment effect, such as the average treatment effect on the treated (ATT), $\tau_{ATT} = E(\tau|D = 1) = E(Y_1|D = 1) - E(Y_0|D = 1)$.

In the panel conditioning case, the expected value of the outcome of the treated cases, $E(Y_1|D = 1)$, can be directly observed from the data: it is simply the expected value of Y among respondents who responded to several waves of a panel. The counterfactual outcome of the treated, $E(Y_0|D = 1)$, by contrast, cannot be observed. When treatment assignment is random, however, we can replace $E(Y_0|D = 1)$ with $E(Y_0|D = 0)$, the expected value of the outcome among the cases that were not treated (that is, the expected value of Y of the control group), and estimate an unbiased treatment effect. Using this framework to study panel conditioning effects, the challenge is to find suitable estimates of the control group outcome, i.e., an estimate of Y from persons who did not participate or participated only once in a survey, but who would participate in all waves of the panel, had they been selected.

Approaches to replace the counterfactual outcome of the treated can broadly be grouped into two categories. The first approach relies on within-person comparisons and the second on between-person comparisons. A review of the literature suggests that the majority of studies applies a variant of the between-person approach. Each approach has its own specific strengths and weaknesses in identifying panel conditioning effects. One challenge that all of them face is eliminating or adjusting for panel attrition (one aspect of the third methodological challenge, see Section 3.3). The initial review of the different ways to come up with an estimate of the counterfactual does not include a discussion of panel attrition. Instead, I discuss panel attrition and other sources of confounding error only after all approaches haven been reviewed.

3.2.1 The within-person approach

The within-person approach uses panel respondents as their own control group and simply compares survey outcomes among the same people at different waves of a panel. That is, the control group outcome is usually defined as responses in the first wave of a panel (where respondents are not yet conditioned), while the outcome of the treatment group is defined as responses by the same respondents to subsequent waves of the same panel survey. A few studies have applied this design, e.g., by comparing coefficients of variation at different waves of a panel diary (Toh, Lee, and Hu 2006), by comparing responses from a baseline survey with follow-up reports of the same respondents collected one week later (Sharpe and Gilbert 1998), by studying trends in attitudes across eleven waves of the British Household Panel Survey (Sturgis et al. 2009) or by analyzing person fit measures in several dozen waves of the German Socio-Economic Panel (Kroh et al. 2016).

The ability of this method to estimate an unbiased panel conditioning effect, however, heavily depends on the object of study. If one is interested in behavioral or attitudinal changes due to panel conditioning, one needs to account for the fact that there are many other factors (besides participating in the panel survey) that may cause a change in behavior or attitudes. Disentangling such 'true' change (i.e., change, that would have happened even in the absence of participation in the survey) from change that is caused by survey participation alone is impossible when relying on a single set of respondents only (see, for example, Shadish, Cook, and Campbell 2002:ch.4, for a general discussion of this approach in the causal inference framework). Thus, this method will lead to biased estimates of panel conditioning in many settings. Sturgis et al. (2009), for example, conclude from discovering changes over time in political attitudes among the same respondents of the British Household Panel Survey that these attitudes changed due to panel conditioning. However, many factors that influence political attitudes changed over the same course of time (e.g., new bills being passed or new governments being elected). Thus, change in political attitudes over several waves of the panel is not only caused by panel survey participation, but also by change in other (external) factors over time. Disentangling these two forms of change, however, is impossible using only one set of respondents. That is, Sturgis et al. (2009)'s estimates of panel conditioning are likely biased because they do not separate change in attitudes over time that would have happened irrespective of panel participation from change over time that is only due to panel participation.

In other scenarios, relying on a single set of respondents can work well, however. Angel et al. (2017), for example, study changes in the level of measurement error in income reports over time. Other than attitudes or behavior, which can change due to external influences (see above), measurement error in respondents' survey reports is unlikely to covary with external factors. That is, if the level of measurement error is smaller (or larger) in the first wave than in subsequent waves, respondents became worse (or better) reporters over time due to panel conditioning. Thus, when the objective of a study is to analyze changes-in-reporting panel conditioning, relying on a single set of respondents can produce unbiased estimates of panel conditioning effects.

3.2.2 The between-person approach

The second method taking a different approach is more powerful in this regard because it can be applied to study *all* forms of panel conditioning. Instead of within-person comparisons, it relies on between-person comparisons, usually implemented in an experimental design with random assignment of cases to either the treatment or the control group (see the potential outcomes framework introduced above). That is, the between-person approach compares outcomes between respondents with varying levels of exposure to the treatment of survey participation. There are many different variants of this design, which we review below.

The first variant randomly assigns people to either repeated participation in a panel survey or to one-time participation in one wave of the same survey only. As a result, some respondents are interviewed several times (treatment condition), whereas other respondents are interviewed only once (control condition). Kruse et al. (2009), for example, compare attitudes and (reported) behavior of respondents from eleven subsequent waves of the Knowledge Networks Panel with attitudes and behavior of respondents of three independent cross-sectional samples of the same panel that were interviewed with the same instrument at waves three, six or nine. Using a similar design, Axinn et al. (2015) compare reported contraception use between women who were assigned to a baseline survey followed by weekly journal keeping for twelve months and women who were assigned to the baseline survey and a follow-up survey after twelve months only. Pushing the variant described in this paragraph to the most extreme, three studies (Zwane et al. 2011; Crossley et al. 2017; Bach and Eckman 2018) randomly assign people to participation in (several waves of) a survey (treatment) or to no participation at all (control). To measure the outcome, they identify survey participants and people not assigned to survey participation in administrative records that contain measures of the outcome and are available for both groups. In the absence of nonresponse and panel attrition, any differences between the sample selected for survey participation and the control group sample that is observed in the administrative records only are then due to panel conditioning.

The second variant is similar to the first, with one important difference. In variant one, the treatment group is interviewed several times with the same instrument, whereas the control group is interviewed only once (or not at all). In the second variant, however, both the treatment and the control group are interviewed several times, but with different instruments. For example, a study that analyzes the (panel conditioning) effect of repeatedly answering a set of questions A may split respondents into two groups. One group is interviewed with questions A for several waves and the other group is interviewed with questions B. Only in the last wave is the latter group also interviewed with questions A. As a result, the first group is treated with A, but the second one not. Such a design is implemented, for example, by Struminskaya (2016). The main reason for interviewing the control group with other or 'placebo' questions instead of not interviewing them at all is that many panel surveys simply cannot afford to *not* interview some respondents.

The third variant of the between-respondent approach builds on a somewhat different idea. Instead of assigning people to different levels of exposure to the same survey (or variants of it), this approach uses data from two different surveys, where one is a panel survey and the other is a cross-sectional survey. Wilson and Howell (2005), for example, compare trends in the prevalence of arthritis in the U.S. between 1992 and 2002 derived from panel respondents of the Health and Retirement Survey (HRS) with arthritis trends for the same years derived from cross-sectional respondents of the National Health Interview Survey (NHIS). Because data from the two surveys are comparable, as the authors argue, any differences between trends in the HRS and the NHIS data (in the absence of panel attrition in the HRS) are due to panel conditioning in the former. In other words, because respondents of the HRS respond to the same survey multiple times (the treatment group) and the NHIS is a repeated cross-sectional survey that interviews a new group of respondents in each round (the control groups), any differences in the prevalence of arthritis trends between the two is due to panel conditioning in the HRS.

The fourth variant exploits experimental manipulations that occur unintentionally in some panel surveys. Because the sample size of panel surveys decreases over time due to panel attrition, some surveys introduce refreshment samples from time to time to renew the respondent pool to approximately its original sample size (e.g., the General Social Survey or the Current Population Survey in the U.S. or the panel study Labor Market and Social Security in Germany). To ensure comparability between the original sample and the refreshment sample(s), refreshment samples are usually drawn from the same population and with the same sampling design. Thus, people who join a panel survey as part of a refreshment sample form an ideal control group: When they join the panel as novice respondents, respondents of the original sample have already participated in several waves of the panel. Thus, the only difference between the two groups (absent of panel attrition in the original sample) is that one group has been treated and the other one has not been treated (yet). Warren and Halpern-Manners (2012), for example, use this design to study panel conditioning effects in several outcomes measured in the General Social Survey and the German Socio-Economic Panel.

A slightly different variant of this approach has been applied extensively to study panel conditioning effects among respondents of the Current Population Survey (Hansen et al. 1955; Bailar 1975; Shack-Marquez 1986; Solon 1986; Shockey 1988; Bailar 1989; Halpern-Manners and Warren 2012). This survey uses a rotating panel structure, that is, respondents are only interviewed for a certain number of waves before being removed from the panel. At each wave, a new rotation group joins the panel. Comparing outcomes between respondents of one rotation group with another rotation group (with different panel tenure) thus offers an ideal setting to study panel conditioning effects.

All of the four variants reviewed above can provide suitable estimates of the counterfactual outcome $E(Y_0|D=1)$ (had the respondents not participated in the survey). Variants one and four seem the most powerful as they do not require as many assumptions as variants two and three. At the same time, however, they are also the most difficult to implement. Variant one requires researchers to intentionally implement experimental manipulations in a survey. Such manipulations, especially in large-scale social science surveys, are usually expensive. Moreover, many well-established social science longitudinal surveys may simply not be willing to allow researchers to hold out some respondents to be interviewed only occasionally. Principal investigators may fear that such experiments interfere with their historically grown panels. In fact, I am not aware of any intentionally implemented panel conditioning experiments in large-scale longitudinal social science surveys such as the Panel Study on Income Dynamics, the German Socio-Economic Panel or the British Household Panel. Only recently, with the emergence of online panels (e.g., the Longitudinal Internet Study for the Social Sciences in the Netherlands or the German Internet Panel), have some surveys opened their panels for experimental manipulations and methodological experiments.

Variant four can overcome the difficulties discussed above because researchers rely on experimental manipulations that are already (unintentionally) implemented in surveys some other way. The drawback of this variant is that researchers can only work with what is available. That is, while some surveys use refreshment samples or rotating designs (see above) that can be used for uncovering panel conditioning effects, others simply do not apply such designs. Thus, this variant, although very powerful in producing unbiased control group outcomes, completely depends on the availability of refreshment samples or rotation groups.

Variant two also depends on the question whether researchers can implement experimental manipulations in a panel survey. While it may be easier to implement variant two than variant one because *all* respondents are interviewed, the drawback of variant two is that researchers need to be sure that asking some other (placebo) questions does not influence the outcome in some other way.

Variant three requires somewhat different assumptions in order to produce an unbiased control group outcome. Researchers need a cross-sectional survey that asks the same question. Furthermore, this survey should be conducted in the same mode (e.g., CATI or CAPI) to avoid mixing up mode effects with panel conditioning effects (see also the discussion in Section 3.3). Data should be collected at the same time because attitudes and behavior may follow seasonal trends and change over time for other external reasons (see examples above). Samples of both surveys should be drawn with the same design and from the same population to avoid selection bias. Thus, in order for this variant to produce unbiased results, researchers need to find two almost identical surveys with the only difference between the two being that one is conducted as a longitudinal survey, i.e., interviews respondents more than once with the same questionnaire, and the other one as a cross-sectional survey. Given these requirements, variant three seems difficult to implement.

To sum up, the within-person approach and the four variants of the between-person approach all provide helpful designs for the analysis of panel conditioning effects. The between-person variants are more powerful in most settings as they are not restricted to certain forms of panel conditioning and may require fewer assumptions to produce unbiased control group outcomes. Both approaches (even if based on careful experimental manipulations), however, can produce biased estimates of panel conditioning if other confounding sources of error in (panel) surveys are not properly accounted for.

3.3 Challenge 3: Accounting for confounding sources of error

The third challenge is that researchers need to account for confounding sources of error when analyzing panel conditioning effects. The most common are longitudinal nonresponse (i.e., panel attrition), mode effects and interviewer effects.

3.3.1 Panel attrition

The main confounder in studies of panel conditioning is panel attrition, i.e., nonresponse in follow-up waves of a panel survey (e.g., Das, Toepoel, and van Soest 2011). Panel attrition is usually highly selective: it is related to certain observable and/or unobservable characteristics of a respondent, leading to compositional differences between all respondents of the first wave and those who also participate in subsequent waves of a panel. For example, respondents may differ from the full sample regarding socio-demographics or personality traits (e.g., Lepkowski and Couper 2002; Lugtig 2014). In these cases, attrition can easily be mistaken for panel conditioning because changes in the composition of the panel respondents over time can also affect the outcome where panel conditioning is suspected. Consider variant four of the between-person approach, for example. The distribution of socio-demographic information in the panel respondent group will be different from the distribution of these characteristics in the incoming sample due to nonrandom attrition.[6] If attrition is also correlated with the outcome, then comparisons of the mean of this variable between panel respondents and the refreshment sample could reveal (or hide) a panel conditioning effect that, in fact, is only due to nonrandom dropout over time among panel respondents. As a result, panel attrition bias would be mis-characterized as panel conditioning.

Studies of panel conditioning implement various approaches to tackle panel attrition. A few studies recognize attrition as a confounding source of error, but do not adjust for it in their empirical analysis (e.g., Bailar 1975) or subsume conditioning *and* attrition under 'panel bias' (e.g., Bartels 1999). Many more studies address it using one of the adjustment techniques reviewed below.

Broadly speaking, we can categorize approaches to adjust for attrition into three classes. The first class comprises approaches that simply disregard those respondents who do not participate in all waves of a panel. These approaches do not require any assumptions about the form of panel attrition and are therefore a powerful way to eliminate any bias in panel conditioning effects due to panel attrition. A drawback of these approaches, however, is that they can be implemented in combination with the within-person approach only (e.g., Sturgis et al. 2009:see Section 3.2.1) and variants two and four of the between-person approach only (e.g., Halpern-Manners and Warren 2012:see Section 3.2.2). Disregarding attriters cannot be used in combination with variants one and three of the between-person approach because only the treatment groups are affected by attrition. Thus, limiting the treatment groups to respondents who participate in all waves would not account for attrition at all. Instead, it would exacerbate error due to attrition because additional cases would be excluded from the group that already suffers from nonrandom dropout. Therefore, this approach is limited to certain scenarios.

Insert Table 2 about here.

The study by Halpern-Manners and Warren (2012), implementing this approach in combination with variant four of the between-person approach, deserves a closer look as the authors implement one of the most elaborate research designs to the analysis of panel conditioning effects. In the simplest case, their design consists of two rotation groups, as shown in Table 2. The treatment group is first interviewed Wave 1 and interviewed for a second time in Wave 2. The control group is interviewed for the first time in Wave 2 and for the second time in Wave 3. Restricting both groups to those respondents who participate in two waves (Wave 1 and Wave 2 for the treatment group and Wave 2 and Wave 3 for the control group) and comparing the Wave 2 outcome of the treatment group (their second interview) with the Wave 2 outcome of the control group (their first interview) eliminates any confounding error due to panel attrition without having to impose any further assumptions. Thus, this design is very powerful in eliminating any confounding error due to panel attrition from panel conditioning and in providing unbiased control group data. Yet, as noted before, it is limited to those surveys that apply the refreshment or rotation group design described in Section 3.2.2.

The second class of approaches comprises all methods that adjust for attrition by conditioning on observable determinants of attrition (e.g., by including them as control variables in regression functions) or functions of the determinants (e.g., by using them as weights). These approaches usually assume that attrition is missing at random (MAR), i.e., attrition is random conditional on some fully observed covariates (Rubin and Little 2002). In other words, any confounding bias due to attrition in estimates of panel conditioning is adjusted for when all variables that determine whether a respondent drops out over time or not are correctly accounted for. The difficulty with this approach, however, is that the MAR assumption cannot be tested or verified in a statistical way. That is, researchers can only *assume* that attrition is MAR by, for example, conditioning on all determinants identified as predictors of attrition in previous research or all covariates thought to be related to attrition based on theory. If researchers fail to include relevant variables or if attrition is missing not at random (MNAR), then panel conditioning effects will be confounded by attrition.

Empirical implementations of this class of approaches are numerous. Kruse et al. (2009), for example, adjust for attrition by including determinants of attrition as control variables in their regression model. Other studies (e.g., Pennell and Lepkowski 1992; Dennis 2001; Nancarrow and Cartwright 2007), by contrast, condition on determinants of attrition by including nonresponse weights supplied by the panel survey (e.g., longitudinal nonresponse weights or poststratification weights) or calculate their own weights based on propensity scores (e.g., Struminskaya 2016). Regardless, these studies cannot test whether the MAR assumption actually holds.

Approaches based on the MAR assumption are often used because they can be combined with *each* of the approaches presented in Section 3.2. Preference should be given to the first class of approaches (restricting the sample to respondents of all waves) however, because they do not require relying on an untestable assumption.

A third approach to account for attrition uses instrumental variables. That is, the third approach does not require researchers to assume that they observe all determinants of attrition. Instead, this method exploits random allocation of people to a treatment and a control group. I am aware of only one study that has applied this approach to account for attrition. I therefore describe the approach with the example of this study.[7] Bach and Eckman (2018) observe the behavior of two groups of people (a treatment group and a control group) in administrative records to study a changes-in-behavior panel conditioning effect. The treatment group consists of respondents and nonrespondents of a panel survey, and the control group is observed in the administrative data only. To account for attrition and initial nonresponse, the authors instrument the (endogenous) participation in the survey among members of the treatment group with the random allocation of people to the treatment or the control group. Because the allocation of people to the two groups is random and correlated with the endogenous treatment (actual participation in the survey), it is a valid instrument by definition. That is, this method exploits the random variation from the allocation of people to treatment or control to overcome the endogeneity resulting from members of the treatment group self-selecting into panel response or nonresponse. Thus, this method is a powerful tool to eliminate bias due to attrition and initial nonresponse among members of the treatment group. A major drawback of the approach, however, is that it works only when outcome information can be observed for respondents, nonrespondents, and the control group. Moreover, instrumental variables will likely underestimate the true panel conditioning effect because all respondents who participate in at least one wave of a panel have to be treated as panel respondents.

All three classes of approaches to account for confounding error due to panel attrition can be used to uncover unbiased panel conditioning effects. The choice of the method, however, is often limited by the availability of data and the design applied by the panel study of interest. Even when attrition is successfully accounted for, other sources of error still need to be considered in order to estimate unbiased panel conditioning effects.

3.3.2 Other confounding sources of error

Mode and interviewer effects are two other sources of error in surveys that require adjustment (Halpern-Manners and Warren 2012; Chadi 2013). Both phenomena can lead to changes in measurement error in panel surveys over time and may therefore easily be mistaken for panel conditioning. Halpern-Manners and Warren (2012), for example, report that many studies of panel conditioning in the Current Population Survey (CPS) likely estimate biased conditioning effects because they confound conditioning with mode effects as the survey uses mostly in-person interviews in a respondent's first round and telephone interviews in subsequent rounds. Thus, changes in an outcome between the first round and subsequent rounds of the CPS may (partly) be due to face-to-face and telephone interviews resulting in different levels of measurement error. Halpern-Manners and Warren (2012) demonstrate, however, that the CPS is still affected by panel conditioning, even when mode effects have been properly accounted for.

Regarding interviewer effects, Chadi (2013) shows that self-reported life satisfaction decreases steadily over time due to increased trust in the interviewer (a changes-in-reporting effect). Once panel respondents are interviewed by a new interviewer, however, he finds an abrupt rise in life satisfaction (an interviewer effect). Without acknowledging changes in interviewer allocation, such abrupt rises in life satisfaction may be falsely attributed to panel conditioning. Similarly, Van der Zouwen and Van Tilburg (2001) report that changes in reported network size over time are due to interviewer behavior and not panel conditioning. Moreover, other substantial changes in panel surveys, such as a change of the data collection agency or the introduction of dependent interviewing, might lead to similar changes over time that can easily be mistaken for panel conditioning. However, we are not aware of any published research regarding these other sources of error. To sum up, researchers need to make sure that any changes found over time are only due to repeated interviewing of the same people and not due to other elements of a survey that might change between waves and cause change in respondents' behavior and/or reporting.

4 Discussion and a look forward

The purpose of this study was to give an overview of the current state of methodological research on panel conditioning effects in social science longitudinal surveys. I have identified three methodological challenges that have plagued research on panel conditioning for a long time. First, panel conditioning can result in changes-in-behavior and/or changes-in-reporting. To make statements about either form, it is essential to clearly disentangle the two forms (at minimum, researchers should acknowledge that panel conditioning can take various forms). Second, obtaining control group data of people who have not been interviewed or have been interviewed only once is crucial to the identification of the causal effect of panel survey participation on inter-wave changes. Third, panel attrition and other sources of error in (panel) surveys have to be accounted for to estimate unbiased panel conditioning effects. I have reviewed these challenges and the research designs developed to tackle them and discussed their strengths and weaknesses.

The discussion of these designs and their implementations in the literature have shown that more methodologically sound research is needed that carefully tackles all of the challenges identified in this study. In addition, more studies should be conducted with the kind of large-scale social science longitudinal surveys that social scientists frequently use for substantive research. For example, I am not aware of any research on panel conditioning effects in the Panel Study on Income Dynamics, the world's longest-running household panel study. However, I also acknowledge that the implementation of methodological experiments is expensive, difficult or even impossible. As a result, the choice of research designs may be limited to those ex post designs that require several (untestable) assumptions, making clear statements about the presence of panel conditioning effects difficult or even impossible.

Future research should also go beyond uncovering *average* panel conditioning effects. Little is known about treatment effect heterogeneity, i.e., how panel conditioning effects vary for different subgroups. Panel conditioning due to learning (see Section 2.1), for example, may vary with respondents' cognitive ability. Research should also extend the current focus on univariate panel conditioning to the multivariate context. To date, little is known about the ways panel conditioning biases estimates derived from complex multivariate statistical models. Future work should assess how panel conditioning translates into bias in, e.g., regression coefficients derived from econometric panel data models. Such projects may, for example, simulate data without panel conditioning from real data that is affected by panel conditioning. Comparing estimated panel regression coefficients between the two datasets would, for example, allow to assess bias in multivariate estimates due to panel conditioning.

Recent technological advancements in data collection techniques, such as the use of mobile (smart)phones, create even more opportunities for studies on conditioning effects. Online panel surveys, text message surveys and app-based surveys, for example, offer new means of data collection that are easy to implement, cheap and often allow to interview respondents up to several times per month (online panels) or even per week (text message and app-based surveys). At the same time, however, high-frequency surveying is also most susceptible to conditioning (see Section 2). Thus, when collecting high-frequency data via text message surveys, smartphones or online surveys, special attention must be paid to the possibility that the act of measuring changes what is being measured. If researchers incorporate experimental manipulations in such new data collection designs from the very beginning, obtaining sound and unbiased estimates of panel conditioning will be much easier.

Any experiment on panel conditioning, however, should only be conducted if there is theoretical reason to assume a conditioning effect to be present (for example, if researchers believe that answering to a survey item encourages respondents to change their behavior). Similarly, studies should be restricted to those items routinely used in surveys because the scientific value of studying items that are hardly ever part of a survey may be questionable, a point also raised by Warren and Halpern-Manners (2012).

Furthermore, findings regarding changes-in-behavior panel conditioning may also open new ethical debates about survey research. That is, as soon as questions change behavior, researchers must carefully think about the resulting behavioral change from an ethical point of view. Veroff et al. (1992), for example, demonstrated that asking questions about marriage may lead to decreased marital satisfaction. Similarly, asking questions about suicidal thoughts in a survey could increase the likelihood of suicide among certain respondents. Thus, some questions may have huge negative influences on respondents' behavior. Research on panel conditioning therefore may also provide guidelines for researchers collecting data regarding the ethics of asking certain questions.

Finally, no recommendations exist regarding what to do when panel data are affected by panel conditioning. Are data really "irredeemably biased" as Warren and Halpern-Manners (2012:522) put it? Or is it possible to 'repair' data or account for panel conditioning by, e.g., adjusting statistical models? Can survey designers imagine ways to avoid panel conditioning in the first place? Although this review has shown that scholars have analyzed for decades how repeated interviewing of the same people can change their (reporting of) behavior and attitudes, I have also identified several challenges that many studies fail to address adequately. Similarly, the discussion shows that many important issues related to conditioning effects still need to be addressed.

Notes

[1]Strictly speaking, one would need to disentangle changes in the reporting of attitudes from real changes in attitudes in a similar way because panel conditioning can result in both (real changes in an attitude and changes in the reporting of an attitude). Since attitudes are, to the author's knowledge, always measured through self-reports, disentangling the two seems impossible.

[2]Some of these studies likely over- or underestimate the panel conditioning effect in the CPS due to confounding panel attrition and mode effects (see Section 3.3). However, the general finding that the CPS underestimates unemployment rates seems to hold even when attrition and mode effects are accounted for (Halpern-Manners and Warren 2012).

[3]Note that respondents of surveys containing questions asking for socially non-normative behavior may be subject to both changes-in-reporting as well as changes-in-behavior panel conditioning. This dual effect poses a major challenge to panel conditioning research (see Sections 1 and 3.1).

[4]We do not need to differentiate between different effects of panel survey participation on respondents' knowledge. Panel conditioning either increases respondents' knowledge or does not affect it at all. I am not aware of any research regarding misreporting of knowledge in surveys.

[5]While excluding the possibility that survey participation affects actual behavior seems justified in some scenarios, I cannot think of an example where panel conditioning may result in changes-in-behavior only, i.e., excluding changes-in-reporting.

[6]Initial nonresponse is usually less of a problem because it affects both samples to the same degree. It may affect estimates of panel conditioning only if nonresponse patterns in the first interview of the treatment group are different from nonresponse patterns of the first interview of the control group. For example, nonresponse patterns in a survey of attitudes on same-sex marriage may differ between the treatment group and the control group. If same-sex marriage were to be legalized between the first wave of the treatment group and the first wave of the control group, then nonresponse patterns may differ between the two groups. Some opponents of the new bill who would have responded to the survey before the introduction of the bill may in fact not respond after legalization as they may be afraid to share their opposing opinion in an interview due to social desirability. In such cases, initial nonresponse patterns may differ between the first round of the treatment group and the first round of the control group causing an additional challenge for the analysis of panel conditioning effects. The methods that adjust for panel attrition (discussed below), however, can easily be extended to adjust for different (initial) nonresponse patterns, too.

[7]Crossley et al. (2017) apply the same approach to account for nonresponse in their study of the effect of participating in only one wave of a panel survey on respondents' behavior.

Table 1: Summary of methodological challenges and solutions

Challenge	Solutions	Remarks
Challenge 1: Disentangling changes-in-behavior and changes-in-reporting	Records independent of respondents' reporting (e.g., administrative data) to study changes-in-behavior	Requires data independent of respondents' reporting and linkage with respondents
	Validation records independent of respondents' reporting (e.g., administrative data) to study changes-in-reporting	Requires data independent of respondents' reporting and linkage with survey records
	Exclude changes-in-behavior based on theoretical considerations	
Challenge 2: Finding control group data	*Within-person approach*: Use panel respondents as their own control group	Restricted to analysis of some forms of changes-in-reporting (e.g., measurement error)
	Between-person approaches	
	Variant 1: Implement experiment	
	Variant 2: Placebo interviews	Confounding effects of placebo interviews?
	Variant 3: Cross-sectional survey with same content	Cross-sectional survey: same mode, same questions, same sampling design and population as well as same time of data collection
	Variant 4: Panel survey with rotating designs	
Challenge 3: Accounting for confounding sources of error		
Panel attrition	Disregard respondents who do not participate in all waves	Only in combination with within-person approach and variants 2 and 4 of between-person approach
	Condition on observable determinants of attrition	Requires (untestable) missing-at-random assumption
	Instrumental variables	Requires outcome data for both respondents and nonrespondents as well as control group of people not interviewed
Mode and interviewer effects	Restrict to cases interviewed in same mode and by same interviewers	

Table 2: Example of a research design accounting for attrition with a rotating panel survey

	Wave 1	**Wave 2**	**Wave 3**
Rotation group 1	X	X	
Rotation group 2		X	X

Comparing the Wave 2 outcomes between Rotation group 1 and Rotation group 2 accounts for attrition because both rotation groups contain only respondents who participate in all waves.

References

Angel, Stefan, Richard Heuberger, and Nadja Lamei. 2017. "Differences Between Household Income from Surveys and Registers and How These Affect the Poverty Headcount: Evidence from the Austrian SILC." *Sociological Indicators Research* Advance online publication, DOI 10.1007/s11205-017-1672-7.

Axinn, William G., Elyse A. Jennings, and Mick P. Couper. 2015. "Response of Sensitive Behaviors to Frequent Measurement." *Social Science Research* 49:1–15.

Bach, Ruben L. and Stephanie Eckman. 2018. "Participating in a Panel Survey Changes Respondents' Labour Market Behaviour." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* `https://doi.org/10.1111/rssa.12367`.

Bach, Ruben L. and Stephanie Eckman. forthcoming. "Motivated Misreporting in Web Panels." *Journal of Survey Statistics and Methodology* .

Bailar, Barbara A. 1975. "The Effects of Rotation Group Bias on Estimates from Panel Surveys." *Journal of the American Statistical Association* 70(349):23–30.

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

Bartels, L. 1999. "Panel Effects in the American National Election Studies." *Political Analysis* 8:1–15.

Battaglia, Michael P., Elizabeth R. Zell, and Pamela L. Y. H. Ching. 1996. "Can Participating in a Panel Sample Introduce Bias into Trend Estimates?" *Proceedings of the American Statistical Association, Survey Research Methods Section* pp. 1010–1013.

Bergmann, Michael and Alice Barth. forthcoming. "What Was I Thinking? A Theoretical Framework for Analysing Panel Conditioning in Attitudes and (Response) Behaviour." *International Journal of Social Research Methodology* .

Borle, Sharad, Uptal M. Dholakia, Siddharth S. Singh, and Robert A. Westbrook. 2007. "The Impact of Survey Participation on Subsequent Customer Behavior: An Empirical Investigation." *Marketing Science* 26(5):711–726.

Cantor, David. 2010. "A Review and Summary of Studies on Panel Conditioning." In *Handbook of Longitudinal Research. Design, Measurement and Analysis*, edited by S. Menard, pp. 123–138. Amsterdam: Elsevier.

Chadi, Adrian. 2013. "The Role of Interviewer Encounters in Panel Responses on Life Satisfaction." *Economic Letters* 121:550–554.

Chandon, Pierre, Vicki G. Morwitz, and Werner J. Reinartz. 2005. "Do Intentions Really Predict Behavior? Self-Generated Validity Effects in Survey Research." *Journal of Marketing* 69(2):1–14.

Clausen, Aage R. 1969. "Response Validity: Vote Report." *The Public Opinion Quarterly* 32:588–606.

Clinton, Joshua D. 2001. "Panel Bias from Attrition and Conditioning: A Case Study of the Knowledge Networks Panel." *Technical report*, Department of Political Science Technical Report, Stanford University, Stanford, CA.

Cohen, Steven B. and Vicki L. Burt. 1985. "Data Collection Frequency Effect in the National Medical Care Expenditure Survey." *Journal of Economic & Social Measurement* 13(2):125–151.

Coombs, Lolagene C. 1973. "Problems of Contamination in Panel Surveys: A Brief Report on an Independent Sample, Taiwan, 1970." *Studies in Family Planning* 4(10):257–261.

Crossley, Thomas, Jochem de Bresser, Liam Delaney, and Joachim Winter. 2017. "Can Survey Participation Alter Household Saving Behavior?" *Economic Journal* 127:2332–2357.

Das, Marcel, Vera Toepoel, and Arthur van Soest. 2007. "Can I Use a Panel? Panel Conditioning and Attrition Bias in Panel Surveys." *CentER Discussion Paper Series* 2007-56.

Das, Marcel, Vera Toepoel, and Arthur van Soest. 2011. "Nonparametric Tests of Panel Conditioning and Attrition Bias in Panel Surveys." *Sociological Methods & Research* 40(1):32–56.

Dennis, J M. 2001. "Are Internet Panels Creating Professional Respondents? A Study of Panel Effects." *Marketing Research* 13(2):34–39.

Duan, Naihua, Margarita Alegria, Glorisa Canino, Thomas McGuire, and David Takeuchi. 2007. "Survey Conditioning in Self-Reported Mental Health Service Use: Randomized Comparison of Alternative Instrument Formats." *Health Research and Educational Trust* 42(2):890–907.

Eckman, Stephanie and Frauke Kreuter. forthcoming. "Misreporting to Looping Questions in Surveys: Recall, Motivation and Burden." *Survey Research Methods* .

Eckman, Stephanie, Frauke Kreuter, Antje Kirchner, Annette Jäckle, Roger Tourangeau, and Stanley Presser. 2014. "Assessing the Mechanisms of Misreporting to Filter Questions in Surveys." *Public Opinion Quarterly* 78(3):721–733.

Granberg, Donald and Soren Holmberg. 1992. "The Hawthorne Effect in Election Studies: The Impact of Survey Participation on Voting." *British Journal of Political Science* 22(2):240–247.

Halpern-Manners, Andrew and John R. Warren. 2012. "Panel Conditioning in Longitudinal Studies: Evidence From Labor Force Items in the Current Population Survey." *Demography* 49(4):1499–1519.

Halpern-Manners, Andrew, John R. Warren, and Florencia Torche. 2014. "Panel Conditioning in a Longitudinal Study of Illicit Behaviors." *Public Opinion Quarterly* 78(3):565–590.

Hansen, Morris H., William N. Hurwitz, Harold Nisselson, and Joseph Steinberg. 1955. "The Redesign of the Census Current Population Survey." *Journal of the American Statistical Association* 50(271):701–719.

Holland, Paul W. 1986. "Statistics and Causal Inference." *Journal of the American Statistical Association* 81(396):945–960.

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

Kreuter, Frauke, Susan McCulloch, Stanley Presser, and Roger Tourangeau. 2011. "The Effects of Asking Filter Questions in Interleafed Versus Grouped Format." *Sociological Methods & Research* 40(1):88–104.

Kroh, Martin, Florin Winter, and Juergen Schupp. 2016. "Using Person-Fit Measures to Assess the Impact of Panel Conditioning on Reliability." *Public Opinion Quarterly* 80(4):914–942.

Kruse, Yelena, Mario Callegaro, J Michael Dennis, Charles DiSogra, Stefan Subias, Michael Lawrence, and Trevor Tompson. 2009. "Panel Conditioning and Attrition in the AP-Yahoo! News Election Panel Study." *Proceedings of the 64th conference of the American Association for Public Opinion Research (AAPOR)* pp. 5742–5756.

Lazarsfeld, Paul F. 1940. "'Panel Studies'." *The Public Opinion Quarterly* 4:122–128.

Lepkowski, J. and M. Couper. 2002. "Nonresponse in the Second Wave of Longitudinal Household Surveys." In *Survey Nonresponse*, edited by R. Groves, J. Eltinge D. Dillman, and R. Little, pp. 259–271. New York: Wiley.

Lugtig, Peter. 2014. "Panel Attrition Separating Stayers, Fast Attriters, Gradual Attriters, and Lurkers." *Sociological Methods & Research* 43(4):699–723.

Mathiowetz, Nancy A. and Tamra J. Lair. 1994. "Getting Better? Change or Error in the Measurement of Functional Limitations." *Journal of Economic and Social Measurement* 20(3):237–262.

Morwitz, Vicki G., Eric Johnson, and David Schmittlein. 1993. "Does Measuring Intent Change Behavior?" *The Journal of Consumer Research* 20(1):46–61.

Nancarrow, Clive and Trixie Cartwright. 2007. "Online Access Panels and Tracking Research: The Conditioning Issue." *International Journal of Market Research* 49(5):573–594.

Neter, John and Joseph Waksberg. 1964. "Conditioning Effects from Repeated Household Interviews." *Journal of Marketing* 28(2):51–56.

Oberski, Daniel, Antje Kirchner, Stephanie Eckman, and Frauke Kreuter. 2017. "Evaluating the Quality of Survey and Administrative Data through Multi-Trait Multi-Method Models." *Journal of the American Statistical Association* 112(520):1477–1489.

Pennell, Steven G. and James N. Lepkowski. 1992. "Panel Conditioning Effects in the Survey of Income and Program Participation." *Proceedings of the American Statistical Association, Survey Research Methods Section* pp. 566–571.

Rubin, Donald B. 1974. "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies." *Journal of Educational Psychology* 66(5):688–701.

Rubin, Donald B. 1978. "Bayesian Inference for Causal Effects: The Role of Randomization." *Annals of Statistics* 6:34–58.

Rubin, Donald B. and Roderick J. A. Little. 2002. *Statistical Analysis with Missing Data*. New York: Wiley, 2 edition.

Schonlau, Matthias and Vera Toepoel. 2015. "Straightlining in Web Survey Panels Over Time." *Survey Research Methods* 9(2):125–137.

Schündeln, Matthias. 2017. "Multiple Visits and Data Quality in Household Surveys." *Oxford Bulletin of Economics and Statistics* Advance online publication, DOI: 10.1111/obes.12196.

Shack-Marquez, Janice. 1986. "Effects of Repeated Interviewing on Estimation of Labor-Force Status." *Journal of Economic and Social Measurement* 14(4):379–398.

Shadish, Willian R., Thomas D. Cook, and Donald T. Campbell. 2002. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference.* Boston, MA: Houghton Mifflin.

Sharpe, J. Patrick and David G. Gilbert. 1998. "Effects of Repeated Administration of the Beck Depression Inventory and Other Measures of Negative Mood States." *Personality and Individual Differences* 24(4):457–463.

Shockey, James W. 1988. "Adjusting for Response Error in Panel Surveys: A Latent Class Approach." *Sociological Methods & Research* 17(1):65–92.

Smith, Jennifer K., Alan S. Gerber, and Anton Orlich. 2003. "Self-Prophecy Effects and Voter Turnout: An Experimental Replication." *Political Psychology* 24(3):593–604.

Solon, Gary. 1986. "Effects of Rotation Group Bias on Estimation of Unemployment." *Journal of Business & Economic Statistics* 4(1):105–109.

Spangenberg, Eric and Carl Obermiller. 1996. "To Cheat or Not to Cheat: Reducing Cheating by Requesting Self-Prophecy." *Marketing Education Review* 6(3):95–103.

Struminskaya, Bella. 2016. "Respondent Conditioning in Online Panel Surveys. Results of Two Field Experiments." *Social Science Computer Review* 34(1):95–115.

Sturgis, Patrick, Nick Allum, and Ian Brunton-Smith. 2009. "Attitudes Over Time: The Psychology of Panel Conditioning." In *Methodology of Longitudinal Surveys*, edited by P. Lynn, pp. 113–126. New York: Wiley.

Toh, Rex S., Eunkyu Lee, and Michael Y. Hu. 2006. "Social Desirability Bias in Diary Panels is Evident in Panelists' Behavioral Frequency." *Psychological Reports* 99:322–334.

Torche, Florencia, John R. Warren, and Andrew Halpern-Manners. 2012. "Panel Conditioning in a Longitudinal Study of Chilean Adolescents' Substance Use: Evidence from an Experiment." *Social Forces* 90(3):891–918.

Traugott, Michael W. and John P. Katosh. 1979. "Response Validity in Surveys of Voting Behavior." *Public Opinion Quarterly* 43(3):359–377.

Van der Zouwen, Johannes and Theo Van Tilburg. 2001. "Reactivity in Panel Studies and its Consequences for Testing Causal Hypotheses." *Sociological Methods & Research* 30(1):35–56.

Van Landeghem, Bert. 2012. "Panel Conditioning and Subjective Well-Being: Evidence from International Panel Data and Repeated Cross-Sections." *SOEPPaper* No. 484.

Veroff, Joseph, Shirley Hatchett, and Elizabeth Douvan. 1992. "Consequences of Participating in a Longitudinal Study of Marriage." *Public Opinion Quarterly* 56(3):315–327.

Warren, John R. and Andrew Halpern-Manners. 2012. "Panel Conditioning in Longitudinal Social Science Surveys." *Sociological Methods & Research* 41(4):491–534.

Waterton, Jennifer and Denise Lievesley. 1989. "Evidence of Conditioning Effects in the British Social Attitudes Panel." In *Panel Surveys*, edited by D. Kasprzyk, G. J. Duncan, G. Kalton, and M. P. Singh, pp. 319–339. New York: Wiley.

Williams, Patti, Lauren G. Block, and Gavan J. Fitzsimons. 2006. "Simply Asking Questions about Health Behaviors Increases both Healthy and Unhealthy Behaviors." *Social Influence* 1(2):117–127.

Wilson, Sven E. and Benjamin L. Howell. 2005. "Do Panel Surveys Make People Sick? US Arthritis Trends in the Health and Retirement Study." *Social Science & Medicine* 60:2623–2627.

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

Yan, Ting and Stephanie Eckman. 2012. "Panel Conditioning: Change in True Value versus Change in Self-Report." *Proceedings of the American Statistical Association, Survey Research Methods Section* pp. 4726–4736.

Zwane, Alix Peterson, Jonathan Zinman, Eric van Dusen, William Pariente, Clair Null, Edward Miguel, Michael Kremer, Dean S. Karlan, Richard Hornbeck, Xavier Giné, Esther Duflo, Florencia Devoto, Bruno Crepon, and Abhijit Banerjee. 2011. "Being Surveyed Can Change Later Behavior and Related Parameter Estimates." *Proceedings of the National Academy of Sciences* 108(5):1821–1826.