
*Panel Conditioning Effects in Longitudinal Social Science Surveys:
Implications for Measuring Social and Economic Well-Being*

John Robert Warren
Department of Sociology
Minnesota Population Center
University of Minnesota

Andrew Halpern-Manners
Department of Sociology
Minnesota Population Center
University of Minnesota

Version: September, 2008

WORKING DRAFT: PLEASE DO NOT CITE OR QUOTE

*Paper prepared for presentation at the annual meetings of the Population Association of America, Detroit, April 2009. Support for this project has been provided by the National Science Foundation (SES-0647710) and by the University of Minnesota's Life Course Center, Department of Sociology, College of Liberal Arts, and Minnesota Population Center. However, errors and omissions are the responsibility of the authors. Please direct correspondence to John Robert Warren, Department of Sociology, University of Minnesota, 909 Social Sciences, 267 ~ 19th Ave. South, Minneapolis, MN 55455 or email warre046@umn.edu

Panel Conditioning Effects in Longitudinal Social Science Surveys: Implications for Measuring Social and Economic Well-Being

ABSTRACT

Does participating in one wave of a longitudinal survey affect respondents' reports of their social and economic well-being in follow-up survey waves? If survey participation does alter respondents' subsequent answers to questions about their social and economic well-being, then this calls into question the validity of information derived from any number of widely used data resources. In this paper, we estimate the magnitude of what methodologists have called panel conditioning or "time-in-survey" effects. Previous efforts to estimate the magnitude of panel conditioning effects have utilized methodologically weak designs and have focused on consequences for a limited range of measures. We use a stronger research design and use large-scale survey data on a wide range of measures of individuals' social and economic well-being to test a series of theoretically derived hypotheses about the circumstances under which panel conditioning effects should be most severe. Preliminary results generally support these hypotheses.

Panel Conditioning Effects in Longitudinal Social Science Surveys: Implications for Measuring Social and Economic Well-Being

Longitudinal surveys provide great methodological leverage in making causal inferences and in understanding processes that unfold over time. Despite their tremendous value for research in the social sciences, public health, and other academic and applied fields, longitudinal surveys also present a variety of unique methodological problems. Among the least well understood of these problems is known as “panel conditioning,” or bias introduced when answering questions in one wave of a longitudinal study alters respondents’ answers to parallel questions in subsequent waves.

Does participation in one wave of a longitudinal survey affect respondents’ subsequent reports of their social and economic well-being? For example, are people’s labor force behaviors—or at least their self-reports of their labor force behaviors—altered by repeated participation in a longitudinal survey that asks detailed questions about labor force participation? Do people’s descriptions of their educational attainments change as a result of periodically answering questions about whether they successfully made various educational transitions? If partaking in one wave of a longitudinal survey does alter participants’ responses to subsequent questions about their social and economic well-being—either because their actual well-being changes or because the quality of their reports of their well-being changes—then this calls into question the quality of information derived from any number of widely-used longitudinal data resources.

Unfortunately, prior attempts to identify panel conditioning (or “time-in-survey”) effects have suffered from three important limitations. First, as described below, most existing research has drawn on methodologically weak or flawed research designs that conflate biases resulting from panel conditioning with those resulting from panel attrition. Second, previous efforts to estimate the magnitude of panel conditioning effects have generally focused on consequences for a narrow range

of measures of individuals' attitudes, behaviors, or characteristics. Third, few investigators have systematically tested theoretically-derived hypotheses about the circumstances under which panel conditioning effects will be large, small, or non-existent. As a result, it is not clear how pervasive or how serious panel conditioning effects are, and there are no established guidelines available to researchers that might help them to decide when panel conditioning is worth worrying about.

In this paper, we will address three research questions. First, does answering questions about social and economic well-being in one wave of a longitudinal survey alter respondents' answers to parallel questions in subsequent survey waves? Is respondents' social and economic well-being—or at least their subsequent reports of their social and economic well-being—influenced by the act of responding to related questionnaire items in previous survey waves? Second, under what conditions might users of longitudinal survey data on social and economic well-being expect panel conditioning effects to be serious and under what conditions might they expect such effects to be weak or non-existent? Does the magnitude of panel conditioning biases vary by the nature of the aspects of social and economic well being under consideration and/or by the manner in which they are measured? Third, what substantive implications might panel conditioning have for our understanding of inequalities in social and economic well being? By way of example, is the presence of panel conditioning enough to alter conclusions about the relationship between gender and labor force participation? Or race/ethnicity and educational attainment?

In this project, we will take advantage of unique design features of the U.S. Current Population Survey (CPS) and the German Socioeconomic Panel (SOEP) to test a series of hypotheses concerning the presence and degree of panel conditioning in measures of social and economic well being. The strong research design and the breadth and variety of survey items under consideration, as we elaborate in the sections that follow, allow for a better understanding of the significance of panel

conditioning and the circumstances under which its effects may be more or less serious. Again, this is no small matter for users of longitudinal survey data. If participation in a panel study substantially alters respondents' subsequent reports of their social and economic well-being then it is imperative to develop techniques to avoid panel conditioning and/or to account statistically for its effects. Although scholars have carefully developed methods to deal with other problems associated with longitudinal research—for example, panel attrition—much less attention has been paid to what may be an equally serious source of bias.

The remainder of this paper is organized into four sections. In the section that follows, we appraise prior research on panel conditioning effects. In the second section, we outline a series of theoretically-derived hypotheses concerning the conditions under which panel conditioning should be more or less severe. Next, we provide a brief discussion of our research design, including a description of our technique for minimizing complications related to panel attrition. Finally, we present an abbreviated sketch of our results. In addition to expanding and elaborating on the sections included in the present draft (particularly the sections describing our research design and results), the final version of this paper will also include an empirical exercise illustrating the substantive implications of panel conditioning for research on social and economic well-being.

PRIOR EVIDENCE

The potential impact of panel conditioning has been recognized since at least 1940, when Lazarfeld (1940: 128) noted that “the big problem yet unsolved is whether repeated interviews are likely, in themselves, to influence a respondent’s opinions.”

The earliest research to address Lazarfeld’s “big problem” focused not on opinions but instead on voting behaviors. Initially spurred on by Clausen’s (1968) serendipitous finding that participation in a pre-cursor to the National Election Study (NES) appeared to increase voter turnout in the 1964

U.S. presidential election, researchers have repeatedly concluded that participation in the NES and other political opinion polls increases voter turnout (Anderson, Silver, and Abramson 1988; Bartels 1999; Kraut and McConahay 1973; Traugott and Katosh 1979; Yalch 1976). For example, using an experimental design, Kraut and McConahay (1973) found that being interviewed before a primary election doubled respondents' probability of voting in that election (as validated by precinct records). Likewise, as part of an effort to validate respondents' self-reports of their voting behavior against administrative records Traugott and Katosh (1979: 376) found that "some respondents were stimulated to vote by the very act of being interviewed."

There is also evidence that panel conditioning affects responses to more widely-used major national U.S. surveys. It has long been understood, for example, that the panel design of the CPS (described in more detail below) has the consequence of leading to some degree of panel conditioning bias in estimates of unemployment rates. Bailer (1975; 1989), Shockey (1988), and others have demonstrated that for any particular calendar month unemployment rates for respondents in their initial month in the CPS rotation are higher than those for respondents who have been in the panel for longer. Indeed the issue has made its way into documentation about the design of the CPS (U.S. Bureau of Labor Statistics 2000), where Table 16-10 shows that in September 1995 the unemployment rate for CPS respondents in their first month in the sample was 6.6 percent higher than for CPS respondents as a whole in that month. The implication is that CPS respondents are less likely to be unemployed, or less willing to report unemployment, after their initial CPS interview.

On the other hand, there do not appear to be similar biases in the Survey of Income and Program Participation (SIPP) (McCormick, Butler, and Singh 1992; Pennell and Lepkowski 1992). Indeed, research on the impact of survey participation on a number of other behaviors—including consumer spending (Bailer 1989; Silberstein and Jacobs 1989), medical care expenditures (Corder and

Horvitz 1989), and news media consumption (Clinton 2001)—shows few such effects. For example, Sen (1976) found no evidence that participating in a survey about bird hunting affected a group of Canadian hunters' subsequently reported bird hunting behaviors as compared to an equivalent group that did not participate in the baseline survey.

Researchers in the U.S. and England have demonstrated effects of panel conditioning on survey respondents' attitudes on a wide variety of subjects (Sturgis, Allum, and Brunton-Smith Forthcoming; Veroff, Hatchett, and Douvan 1992; Waterton and Lievesley 1989; Wilson, Kraft, and Dunn 1989). For example, in an effort to estimate the magnitude of panel conditioning effects Veroff et al. (1992) randomly assigned newlywed couples to one of two groups: one that participated in frequent and intensive interviews about marital satisfaction and well-being over the course of four years and another that participated in minimal and infrequent interviews over that period. The authors concluded that “compared to the control group, the study group had significantly higher variance in their reported marital satisfaction in the second year of marriage, reflecting the fact that a greater number of study group respondents reported low satisfaction. By the fourth year, however, the marriages of the study group couples appeared to be better adjusted on several dimensions of marital quality” (p.315).

Not all researchers seeking to estimate the impact of panel conditioning on attitudes, however, come to similar conclusions. Wang, Cantor, and Safir (2000), for instance, compare members of the National Survey of America's Families panel who were interviewed in 1997 and 1999 to a randomly selected fresh cross-section of respondents who were only interviewed in 1999. A comparison of responses by members of the longitudinal and cross-sectional samples revealed statistically significant differences for only 4 of 32 behavioral and attitudinal items after controls for compositional differences between the two groups.

Critique of the existing literature

These generally mixed results about the existence and magnitude of panel conditioning effects may have to do with weaknesses of the research designs used to detect such effects. Most researchers have proceeded by comparing survey responses from members of a longitudinal panel to those from members of an independent cross-sectional sample (Bartels 1999; Clinton 2001; Sobol 1959; Wang, Cantor, and Safir 2000; Waterton and Lievesley 1989). Details of the problems with this design have been carefully laid out elsewhere (e.g., Holt 1989; Sturgis, Allum, and Brunton-Smith Forthcoming). Most importantly, such a design confounds panel conditioning effects with panel attrition effects (Sturgis, Allum, and Brunton-Smith Forthcoming; Williams and Mallows 1970). Whereas the cross-sectional sample may be representative of a particular population, the panel sample has suffered from attrition over time. Unless adequate steps are taken to account for the resulting selectivity of the panel, differences in responses between the two samples cannot be clearly attributed to panel conditioning.

Other researchers have estimated the magnitude of panel conditioning effects using data from rotating and/or overlapping panel survey designs such as the CPS or SIPP (Bailar 1975; Bailar 1989; McCormick, Butler, and Singh 1992; Pennell and Lepkowski 1992; U.S. Bureau of Labor Statistics 2000). The evidence for panel conditioning effects in the CPS, for example, is derived from a comparison of reported unemployment rates in a given calendar month by respondents in their first month in the panel to those in later months in the panel. Unless the sample weighting procedures fully account for panel attrition, however, any such differences may also be partly due to selective attrition across waves of the CPS. Sturgis et al. (Forthcoming) utilize a different strategy, estimating the magnitude of panel conditioning effects on attitudes by analyzing change over time in responses from a single panel of respondents. As they note, however, this design does not allow them to distinguish

panel conditioning effects from endogenous change over time in respondents' attitudes.

To our knowledge only two studies have randomly assigned individuals to treatment (survey participant) and control (non-participant) groups in order to estimate the impact of panel conditioning (Kraut and McConahay 1973; Veroff, Hatchett, and Douvan 1992). Both found substantial evidence for panel conditioning effects, but each focused on a limited set of outcomes (voting in a primary election in one case, marital quality in the other). However, as with other research designs, even these experiments confounded panel attrition effects with panel conditioning effects. Indeed, neither study makes mention of the extent of panel attrition in their data.

Beyond their methodological problems, no prior research has systematically tested hypotheses about the conditions under which answering survey questions might alter a respondent's attitudes, behaviors, or attributes (or at least their responses to questions about those things). It is true, as described below, that some researchers have speculated about the processes through which responding to survey questions conditions respondents' answers to subsequent questions about particular attitudes or behaviors. For example, Clausen (1968) hypothesized that pre-election surveys "stimulate" respondents' interest and engagement in the election process and thus make them more likely to vote. On the other hand, as noted above, Sen (1976) found no evidence that surveys of bird hunters "stimulated" respondents' interest or engagement in bird hunting. What, theoretically, distinguishes these two examples? Why might panel conditioning effects be more pronounced for measures of some types of respondent attributes and less pronounced for others? Looking beyond their methodological problems in estimating the magnitude of panel conditioning effects, prior research has done little to theorize more broadly about the circumstances under which panel conditioning effects might present a serious threat to the validity and reliability of longitudinal data.

THEORETICAL BACKGROUND AND HYPOTHESES

Under what circumstances might we expect to observe panel conditioning effects? Under what circumstances might we not expect such effects? In this discussion, we begin by distinguishing between two distinct types of panel conditioning. The first type represents actual changes in respondents' attributes between baseline and follow-up surveys that occur as the result of answering the baseline survey questions. The second type of panel conditioning represents situations in which responding to baseline survey questions affects responses to parallel questions on follow-up surveys, even though the underlying attributes do not actually change. When might responding to survey questions actually change respondents' attitudes, behaviors, or other characteristics? When might responding to survey questions change respondents' subsequent responses despite underlying stability in the attributes being described?

Real changes in attitudes as a result of panel conditioning

Most people conceive of attitudes as “crystallized” or fixed in people’s minds. Schuman and Presser (1981: 271) define “crystallized attitudes” as those “that exist independently of our measurement, and that when appropriately measured show high reliability.” In contrast, social psychologists, political scientists, and others view attitudes as varying a great deal in their degree of crystallization (e.g., Tourangeau, Rips, and Rasinsk 2000). Sturgis et al. (Forthcoming: 4), for example, note that “opinions are often weakly held, easily influenced, and founded on a rather thin informational base.” The implication is that respondents’ attitudes about some issues may be quite fluid and malleable, particularly when the topic is one about which respondents’ have given little thought or are not particularly knowledgeable. For instance, many people probably have quite well-established and fixed attitudes about the president’s job performance but less well-established and more malleable

attitudes about their county commissioners' job performance.

With this conceptualization of attitudes in mind, research in survey methodology and experimental social psychology has shown that thinking about an attitude as a result of responding to a survey question has the potential to change that attitude (Millar and Tesser 1986; Sudman, Bradburn, and Schwarz 1996; Tesser 1978). Respondents who lack crystallized attitudes about a topic will nonetheless offer a response to a question about that attitude in a baseline survey. However, the act of participating in the baseline survey may set in motion a series of thoughts or actions that change that attitude by the time of a follow-up survey (Sturgis, Allum, and Brunton-Smith Forthcoming; Waterton and Lievesley 1989; Wilson, Dunn, Bybee, Hyman, and Rotondo 1984; Wilson, Kraft, and Dunn 1989; Wilson, LaFleur, and Anderson 1996). As a result, when asked about that attitude again in the follow-up survey some people's responses may differ from their responses in the baseline survey. For example, the act of responding to a survey question about their attitudes about racial segregation may cause some people to reflect more deeply on that topic, to discuss the topic with others, or to otherwise reconsider their attitudes about that topic. As a consequence, some people's attitudes about racial desegregation may be different the next time they are surveyed.

HYPOTHESIS 1.—Real changes in a particular attitude will occur as the result of responding to questions about that attitude when respondents' initial attitudes are less crystallized and when the issue at hand is salient for respondents.

Put another way, we expect that responding to attitudinal survey questions about a topic will most strongly affect attitudes when respondents have not already developed strongly held or carefully-considered attitudes about that topic. Furthermore, this effect will only come about when the topic is an important or salient one for respondents. If the topic is trivial or irrelevant to respondents, they are less likely to reflect on or reconsider their attitude or discuss it with others as a result of reporting that attitude in a survey.

Real changes in behaviors as a result of panel conditioning

Under what circumstances might we expect to see panel conditioning effects on respondents' actual behaviors? That is, under what circumstances might responding to survey questions about a particular behavior make respondents more or less likely to engage in that behavior?

To explain why participating in a pre-election survey increased respondents' chances of voting, Clausen (1968) speculated that the survey "stimulated" respondents' interest in the election; this increase in interest in the election led more respondents to turn out to the polls. In replicating and re-evaluating Clausen's (1968) findings, Kraut and McConahay (1973) developed two alternative hypotheses. First, their "alienation reduction" hypothesis suggests that the personal, one-on-one contact with the survey interviewer changed respondents' sense of alienation from the political process, reducing psychological barriers toward that process in many respondents' minds. Second, their "self concept" hypothesis suggests that participating in the pre-election survey led many respondents to view themselves as more politically involved than they were prior to the pre-election survey; respondents who saw themselves as more politically involved were thus more likely to vote. Yalch (1976) evaluated each of these hypotheses, found little support for either of Kraut and McConahay's (1973) hypotheses, and confirmed Clausen's (1968) original "stimulation" hypothesis.

The commonalities in Clausen (1968) and Kraut and McConahay's (1973) hypotheses are more broadly useful for understanding the conditions under which participating in a survey might affect respondents' behaviors. In their own way each hypothesis suggests that the experience of responding to survey questions about a behavior alters respondents' motivation to engage in that behavior. Based on this and other research, we suggest that there are two basic ways in which answering survey questions about a behavior may affect respondents' motivation to engage in that behavior.

First, answering survey questions about a behavior may increase respondents' interest in, awareness of, or information about that behavior. Beyond Clausen's (1968) "stimulation" hypothesis—which focuses on the role of a survey in heightening respondents' interest in a behavior—it seems clear that the survey experience may simply provide respondents with information that facilitates their ability to engage in a behavior. A survey of low-income individuals that focuses on their utilization of the State Children's Health Insurance Program (SCHIP) may serve to inform many respondents that the SCHIP exists and is something for which their children may be eligible. Likewise a survey of people's end of life legal preparations may have the consequence of informing (or reminding) some respondents that having a living will and assigning durable power of attorney are things that they should consider doing.

HYPOTHESIS 2.—Real changes in a particular behavior will occur as the result of responding to questions about that behavior when the questions serve to increase respondents' interest in, awareness of, or information about that behavior. This effect will only come about when respondents are inclined to view that behavior as having some positive utility.

A survey of respondents' blood donation behaviors may make them more aware of the benefits of blood donation and may make them better informed about the process of blood donation—thereby motivating some respondents to donate blood who otherwise would not have. The same may not be true for behaviors that have little positive utility for respondents. For example, a survey of respondents' cocaine use may make them more aware of the risks of cocaine use and may make them better informed about the methods of using cocaine—but it is unlikely to motivate respondents to begin to use cocaine.

Second, answering questions about a behavior may lead to self-evaluation by some respondents that subsequently change their chances of engaging in that behavior. For example, a survey of respondents' alcohol use may serve as an opportunity for respondents to reflect on that behavior; respondents who view their alcohol use as excessive may subsequently reduce their alcohol

use. More generally, answering questions about a behavior will affect respondents' subsequent behavior when respondents are forced to give answers to questions that are negatively non-normative or that cause respondents to portray themselves in a way that does not fit with their self-conceptions (Fowler 1995).

HYPOTHESIS 3.—Real changes in a particular behavior will occur as the result of responding to questions about that behavior when the questions cause respondents to give negatively non-normative answers or to portray themselves in a way that does not fit with their self-image.

Responding to a survey in such a way that forces respondents to acknowledge that their behaviors are non-normative—and not in a positive way—or that does not accord with how respondents would like to view themselves may have the consequence of changing subsequent behaviors. A survey of charitable giving behaviors, for example, may lead some respondents to increase their charitable giving if the survey causes them to view themselves as being less generous than other people or less generous than they would like themselves to be.

Panel conditioning effects when the underlying attribute remains constant

In the preceding sections we developed hypotheses about the conditions under which participation in a baseline survey may actually change respondents' attitudes or behaviors prior to a follow-up survey. Although there is evidence—as reviewed above—that attitudes and behaviors may actually be changed as a result of participating in a survey, there is good reason to suppose that another form of panel conditioning is at least as pervasive: respondents' actual characteristics may remain unchanged between waves of a longitudinal survey, but the act of answering questions in a baseline survey may affect the answers that respondent provide in follow-up surveys. Previous research suggests three types of reasons for this form of panel conditioning.

First, improvements in the relationship between interviewers and respondents across waves of a longitudinal survey may influence respondents to indicate that their attitudes or behaviors have

changed even when no such change has actually occurred. Survey research methodologists have found that respondents' judgments about the relative benefits and risks associated with answering survey questions are significantly related to the chances that respondents provide truthful answers (Willis, Sirken, and Nathan 1994). As respondents become more familiar with the survey process and with interviewers and interviewing organizations, respondents may become less suspicious of interviewers and their confidence in the confidentiality of their responses may grow (Fowler 1995). Furthermore, the tendency for survey respondents to give "socially desirable" answers (Tourangeau, Rips, and Rasinsk 2000) may diminish as respondents' trust in interviewers and survey research projects grows (Waterton and Lievesley 1989). For all of these reasons we might expect respondents' answers to questions about their attitudes and behaviors to become more truthful across survey waves.

Whereas respondents may be unwilling to report non-normative, socially inappropriate, illegal, or embarrassing attitudes or behaviors in early waves of a longitudinal study, they may be more willing to do so in later waves. Panel conditioning in this context comes about as respondents come to trust interviewers and the survey research enterprise.

HYPOTHESIS 4.—Apparent changes across survey waves in respondents' attitudes and behaviors will occur as the result of improvements in the relationship between interviewers and respondents. These changes will be most evident in respondents' willingness to report non-normative, socially inappropriate, illegal, or embarrassing attitudes or behaviors.

Second, longitudinal survey respondents may learn to manipulate the survey experience in such a way that minimizes their burden but that also leads to the false impression that their attitudes or behaviors have changed across survey waves. Bailar (1989: 18) notes that "respondents learn that some responses mean additional questioning, so they may avoid giving certain answers." Likewise, Hernandez et al. (1999: 122) note that respondents learn in early waves of longitudinal surveys that certain "trigger" responses lead to additional questions and to a longer interview. They speculate that respondents are thus biased against giving those "trigger" responses in subsequent survey waves. For

example, respondents may learn in the baseline of a longitudinal telephone survey that they are subjected to five minutes of questioning about each sexual relationship that they had in the previous year. As a result, and in order to shorten the duration of their follow-up survey, respondents may underreport the number of sexual relationships that they had in the year preceding the follow-up survey. Although respondents' actual number of sexual relationships per year may remain stable, this form of panel conditioning may give the false impression that respondents reduced their number of sexual relationships over time.

HYPOTHESIS 5.—Apparent changes across survey waves in respondents' attitudes and behaviors will occur as respondents learn to manipulate the survey experience in such a way as to minimize their burden. These changes will be most evident in respondents' willingness to report attitudes or behaviors that lead to additional questioning.

Third, respondents' ability to provide accurate answers may improve across waves of a longitudinal survey, resulting in evidence of change across survey waves when no such change has occurred. Waterton and Livesly (1989) argued that repeated interviewing of the same respondents "leads to improved understanding [by respondents] of the rules that govern the interview process." While this may lead to undesirable manipulation of the survey process (as described above), it may also lead to more accurate responses (Bailar 1989). For example, Hernandez et al. (1999: 122) contend that "participants might learn from earlier waves what is the information required for the survey; thus, they are more capable of compiling and processing the information required to provide accurate responses to the survey questions."

HYPOTHESIS 6.—Apparent changes across survey waves in respondents' attitudes and behaviors will occur as respondents learn to provide more accurate responses. These changes will be most evident in respondents' ability to respond to attitudinal and behavioral questions that involve skip patterns or that require respondents to elaborate on their answers.

Overview

Previous investigations of the magnitude of panel conditioning effects have suffered from three basic

shortcomings. First, as described above, the typical research design utilized in those investigations has been methodologically weak—in particular, prior research has tended to conflate panel conditioning effects with panel attrition effects. Second, previous research has been limited in scope, considering a very limited range of substantive measures. Third, prior research has not systematically tested theoretically-derived hypotheses about the circumstances under which panel conditioning effects might arise. As a result, it is not clear how pervasive or how serious panel conditioning effects are, and there are no guidelines for users of longitudinal survey data that might help them to decide when panel conditioning effects are worth worrying about. As we outline in the next section, in our analyses we use a stronger research design and data on a wide range of attitudinal and behavioral measures to test the hypotheses spelled out above. The results of these analyses will improve our understanding of the magnitude of panel conditioning effects, as well as our knowledge about the conditions under which users of longitudinal survey data should be concerned about this form of bias.

RESEARCH DESIGN AND METHODS

Our analytic strategy for testing the above hypotheses is straightforward. To empirically identify and measure the effects of panel conditioning we compare members of a longitudinal panel (the “treatment” group) to *statistically equivalent* members of a fresh cross-sectional sample (the “control” group).¹ After making the appropriate adjustments for panel attrition, which we describe in substantial detail below, any differences that obtain between the two groups may be solely attributed to differences in their exposure (or lack thereof) to the survey instrument.

Unique design components in the CPS and the SOEP allow for just such a comparison. Although different in nature and content, both data sets are similar in one important respect—respondents are surveyed repeatedly *and* new members are periodically introduced or rotated into the panel. In the case of the CPS, panel members (or rotation groups) are enumerated on eight separate

occasions, spread across two equal time frames. After a rotation group enters the sample its members are surveyed for four consecutive months, left unenumerated during the subsequent eight months, and finally resurveyed for another four months. This design element guarantees that at any point in time, one-eighth of the sample is in the first month of enumeration (rotation group 1), one-eighth is in the second month (rotation group 2), and so forth. It also ensures differences across rotation groups in respondents' familiarity with yearly CPS supplements like the Annual Social and Economic Supplement, such that in any given supplement half of the sample is responding to the add-on questions for the first time (rotation groups 1-4) whereas the other half (rotation groups 5-8) is not. Thus, we are able to test our hypotheses by comparing responses among individuals with previous exposure to a particular CPS item—either on the basic monthly survey or on one of the annual supplementary files—to those who were answering the question for the first time.

In the case of the SOEP, original sample members have been interviewed on an annual basis since 1984 (Wagner, Frick, and Schupp 2007). The survey instruments are comprised of wide array of behavioral and attitudinal items, including questions concerning respondents' personality traits, physical and mental health, employment and professional mobility, earnings, education, social participation, biographical information, and sense of personal satisfaction. In order to maintain sufficient statistical power and ensure continued representativeness, new subsamples were added to the study in 1990, 1994, 1998, 2000, 2002, and, most recently, in 2006. Of these six additional subsamples, three are of particular interest for present analysis—the 1998 “E” subsample, the 2002 “F” subsample, and the 2006 “H” subsample, all of which were drawn using the same sampling scheme that was used to select members of the original longitudinal panel. This feature provides the basis for our treatment-control comparison: a subset of respondents was selected at random to be interviewed longitudinally in 1984 (the treatment group), and then in subsequent years new

respondents (the control group) were added—at random and using strictly comparable sampling methods—to the panel.

As we noted earlier, these comparisons are only instructive insofar as the treatment and control groups differ only with respect to their exposure to the survey instrument (or certain items on the survey instrument). We can imagine a number of scenarios in which this stipulation would be violated. It may be that patterns of non-response vary substantially and systematically across treatment and control groups. For example, original base-year members of the SOEP may be more likely to respond to the survey than individuals who were only just recently added to the panel—they are more heavily invested, they are more experienced respondents, and so forth. Even if response rates are identical for treatment and control group members, however, it is not obvious that non-responders among the treatment group will be the same, statistically, as non-responders among the control group. It may be, for instance, that non-responders in the CPS comprise a random 15% of the control group, but make up a more highly selective 15% of the treatment group. Recovering unbiased estimates of the magnitude of panel conditioning effects, therefore, depends in part on our ability to account for panel attrition in particular and differential response more generally.

To minimize the chances of conflating effects associated with panel conditioning with those attributable to panel attrition, we employ a post-stratification weighting technique (Wu and Sitter 2001). The methodology, which is often termed raking or sample-balancing (Deming and Stephan 1940; Little 1993; Stephan 1942), uses an iterative proportional fitting (IPF) algorithm to generate weights (Izrael, Hoaglin, and Battaglia 2000; Izrael, Hoaglin, and Battaglia 2004). These weights can, in turn, be used to correct for known discrepancies between a sample and a target population, which arise as a result of non-response or related coverage issues (see, e.g., Little 1993).

The raking procedure proceeds as follows. First, we tabulate marginals for the control group on a specified set of raking variables (hereafter referred to as marginal control totals). Our choice of raking variables was informed by two primary considerations: (1) they must be plausibly related to sample attrition; and (2) they cannot themselves be susceptible to panel conditioning effects. Both data sets include a wealth of sociodemographic indicators that satisfy these conditions, including age, gender, urban/rural status, race/ethnicity, nativity, geographic region, and marital status.² Next, the IPF algorithm uses the marginal control totals to compute the appropriate weights. The weights are first adjusted to be consistent with control totals from the marginal distribution of the first raking variable. The resulting weights are then recalibrated to the control totals for the second marginal distribution, a process that is repeated for each of the raking variables. One sequence of adjustments through all of the raking variables represents a single iteration. The algorithm iterates until the weighted marginals converge to the control totals (within a specified tolerance) for all of the marginal distributions simultaneously.

After implementing the weights neither the treatment nor the control group will necessarily be representative of any meaningful population; but that is relatively unimportant for our purposes. What is key is that the two groups are effectively equated to one another along a variety of dimensions so that comparison of responses across treatment and control groups can be more clearly attributable to panel conditioning effects.

PRELIMINARY RESULTS

In this section we provide a *preliminary* sketch of our results. For the sake of time and ease of exposition, we focus on a single hypothesis (Hypothesis 3) and limit our analysis to data from the CPS. In the paper version that we prepare for the PAA meetings in April, we will derive empirical expectations for each of our hypotheses; test appropriate attitudinal and behavioral

items from both the CPS and SOEP; and provide a fuller description and discussion of the results, as well as their implications for substantive research on social and economic well-being.

The logic of Hypothesis 3 suggests that respondents in the treatment group who are presented with behavioral questions that force them to provide non-normative answers or answers that do not fit with their own self-images will be *less* likely to report engaging in those behaviors in subsequent surveys. Two empirical expectations follow:

EXPECTATION 3.1—*In a follow-up survey, the treatment and control groups will not differ in their responses to behavioral items that have not previously been presented to the treatment group.*

EXPECTATION 3.2—*In a follow-up survey, the treatment and control groups will differ in their responses to behavioral items that have previously been presented to the treatment group; treatment group members will be less likely to engage in non-normative or stigmatizing behaviors.*

These expectations are best verified by examining responses to follow-up survey questions—some of which appeared on baseline surveys, others of which did not—about behaviors that may be deemed non-normative, embarrassing, or shameful.

The CPS contains a number of behavioral items that fit this description. Since January of 1998, for example, the basic monthly survey has asked respondents to report the means by which they obtained their high school leaving credential. Respondents are instructed to choose between two options, “graduation from high school” or “GED or equivalent”. Although the acquisition of equivalency degrees has become increasingly commonplace in recent decades (Warren and Halpern-Manners 2007), it by no means constitutes a normative educational outcome. Thus, if Hypothesis 3 is correct, we would expect to observe a relatively *lower* rate of GED acquisition among members of the treatment group in those months in which they had previously responded to the item in question (e.g., all months subsequent to January, 1998), and *equal* rates in the month in which the GED item was first added to the survey.

Figure 1 gives the percent completing high school via a GED, disaggregated by respondents' month in sample and the calendar month of the survey. In order to maintain the basic integrity of our research design, the sample is restricted to CPS reference individuals who were not members of "replacement" households, for whom information was not collected via proxy, and for whom values on the variable of interest were not allocated or otherwise imputed. Because the administration of the CPS varies slightly depending on respondents' month in sample (field operatives are permitted to interview respondents in months 2-4 and 5-8 via telephone), the sample was further restricted to respondents who completed the survey via computer-assisted personal interviewing (CAPI). Solid dots represent point estimates and the attached line segments give 95% confidence intervals. Post-stratification weights are applied throughout in order to adjust for panel attrition in months 2-8.

We find clear support for both of our empirical expectations. As indicated by the two leftmost estimates in Figure 1, we do not observe a statistically significant difference between the estimates obtained for the treatment and control groups in the month in which the GED question was first added to the CPS. We do, however, observe statistically significant differences in subsequent months. In January of 2003, respondents who had previously encountered the GED question were 1.5 percentage points less likely to report obtaining a GED than respondents who were encountering the item for the first time. The observed difference is even more pronounced in January of 2008—respondents in months 2-8 of the CPS were roughly 3.5 percentage points less likely to say that they completed high school via a GED, a difference that amounts to a nearly 30 percent decline. To put this latter percentage into better context, a 30 percent decline in the number of GEDs issued to 16-19 year olds in 2007—and a corresponding increase in regular diplomas—would imply that an additional 60,000 young people graduated

from high school, as opposed to obtaining an equivalency degree (authors' own calculations using data from GED Testing Service 2008).

CONCLUSION

We find preliminary support for Hypothesis 3 using this particular measure from the CPS. Whether we will also find support for this or other hypotheses using other measures from the CPS and German SOEP remains to be seen. By PAA we will have carried out analyses using both data sets sufficient to test each hypotheses, usually with multiple indicators for each hypothesis and resulting empirical expectation.

The methodological value of our work is potentially large. We stand to gain a great deal of basic information about both the magnitude of panel conditioning effects as well as the circumstances under which we might most be concerned about this form of bias. Beyond this, the substantive focus of our test measures---virtually all of which concern some central aspect of social and economic well being---mean that our work will have particular salience for research in this area.

REFERENCES

- Anderson, Barbara A., Brian D. Silver, and Paul R. Abramson. 1988. "The Effects of Race of the Interviewer on Measures of Electoral Participation by Blacks in SRC National Election Studies." *The Public Opinion Quarterly* 52:53-83.
- Bailar, Barbara A. 1975. "The Effects of Rotation Group Bias on Estimates from Panel Surveys." *Journal of the American Statistical Association* 70:23-30.
- . 1989. "Information Needs, Surveys, and Measurement Errors." Pp. 1-24 in *Panel Surveys*, edited by D. Kasprzyk, G. J. Duncan, G. Kalton, and M. P. Singh. New York: Wiley.
- Bartels, Larry M. 1999. "Panel Effects in the American National Election Studies." *Political Analysis* 8:1-20.
- Clausen, Aage R. 1968. "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." Unpublished manuscript, retrieved from http://www.princeton.edu/~clinton/WorkingPapers/C_WP2001.pdf on April 5, 2006. Department of Political Science, Stanford University.
- Corder, Larry S. and Daniel G. Horvitz. 1989. "Panel Effects in the National Medical Care Utilization and Expenditure Survey." Pp. 304-318 in *Panel Surveys*, edited by D. Kasprzyk, G. J. Duncan, G. Kalton, and M. P. Singh. New York: Wiley.
- Deming, W. Edwards and Frederick F. Stephan. 1940. "On a Least Squares Adjustment of a Sampled Frequency Table When the Expected Marginal Totals are Known." *The Annals of Mathematical Statistics* 11:427-444.
- Fowler, Floyd J. Jr. 1995. *Improving Survey Questions: Design and Evaluation*. Thousand Oaks, CA: Sage Publications.
- GED Testing Service. 2008. "2007 GED Testing Program Statistical Report." American Council on Education, Washington, DC.
- Hernandez, Lyla M., Jane S. Durch, Dan G. Blazer II, and Isabel V. Hoverman. 1999. *Gulf War Veterans: Measuring Health*. Committee on Measuring the Health of Gulf War Veterans, Division of Health Promotion and Disease Prevention, Institute of Medicine. Washington, D.C.: National Academies Press.
- Holt, D. 1989. "Panel Conditioning: Discussion." Pp. 340-347 in *Panel Surveys*, edited by D. Kasprzyk, G. J. Duncan, G. Kalton, and M. P. Singh. New York: Wiley.

- Izrael, David, David C. Hoaglin, and Michael P. Battaglia. 2000. "A SAS Macro for Balancing a Weighted Sample." in *Proceedings of the Twenty-Fifth Annual SAS Users Group International Conference*. Cary: SAS Institute Inc.
- . 2004. "To Rake or Note to Rake is Not the Question Anymore with the Enhanced Raking Macro." in *Proceedings of the Twenty-Ninth Annual SAS Users Group International Conference*. Cary: SAS Institute Inc.
- Kraut, Robert E. and John B. McConahay. 1973. "How Being Interviewed Affects Voting: An Experiment." *The Public Opinion Quarterly* 37:398-406.
- Lazarsfeld, Paul F. 1940. "'Panel' Studies." *The Public Opinion Quarterly* 4:122-128.
- Little, R. J. A. 1993. "Post-Stratification: A Modeler's Perspective." *Journal of the American Statistical Association* 88:1001-1012.
- McCormick, Maryann K., Derrick M. Butler, and Rajendra P. Singh. 1992. "Investigating Time in Sample Effect for the Survey of Income and Program Participation." Pp. 554-559 in *Proceedings of the Survey Research Methods Section of the American Statistical Association*.
- Millar, M.G. and A Tesser. 1986. "Thought-Induced Attitude Change: The Effects of Schema Structure and Commitment." *Journal of Personality and Social Psychology* 51:259-269.
- Pennell, Steven G. and James N. Lepkowski. 1992. "Panel Conditioning Effects in the Survey of Income and Program Participation." Pp. 566-571 in *Proceedings of the Survey Research Methods Section of the American Statistical Association*.
- Schuman, Howard and Stanley Presser. 1981. *Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording, and Context*. New York: Academic Press.
- Sen, A.R. 1976. "Developments in Migratory Game Bird Surveys." *Journal of the American Statistical Association* 71:43-48.
- Shockey, James. 1988. "Adjusting of Response Error in Panel Surveys, A Latent Class Approach." *Sociological Methods & Research* 17:478-488.
- Silberstein, Adriana R. and Curtis A. Jacobs. 1989. "Symptoms of Repeated Interview Effects in the Consumer Expenditure Survey." Pp. 289-303 in *Panel Surveys*, edited by D. Kasprzyk, G. J. Duncan, G. Kalton, and M. P. Singh. New York: Wiley.
- Sobol, Marion. 1959. "Panel Mortality and Panel Bias." *Journal of the American Statistical Association* 54:52-68.
- Stephan, Frederick F. 1942. "An Iterative Method of Adjusting Sample Frequency Tables When Expected Marginal Totals are Known." *Annals of Mathematical Statistics* 13:166-178.

- Sturgis, Peter, N. Allum, and I. Brunton-Smith. Forthcoming. "Attitudes Over Time: The Psychology of Panel Conditioning." in *Methodology of Longitudinal Surveys*, edited by P. Lynn. New York: Wiley.
- Sudman, Seymour, Norman M. Bradburn, and Norbert Schwarz. 1996. *Thinking About Answers: The Application of Cognitive Processes to Survey Methodology*. San Francisco: Jossey-Bass Publications.
- Tesser, A. 1978. "Self-Generated Attitude Change." Pp. 289-338 in *Advances in Experimental Social Psychology*, vol. 11, edited by L. Berkowitz. New York: Academic Press.
- Tourangeau, Roger, Lance J. Rips, and Kenneth Rasinski. 2000. *The Psychology of Survey Response*. New York: Cambridge University Press.
- Traugott, Michael W. and John P. Katosh. 1979. "Response Validity in Surveys of Voting Behavior." *The Public Opinion Quarterly* 43:359-377.
- U.S. Bureau of Labor Statistics. 2000. *Current Population Survey: Design Methodology*. Technical Paper 63. Washington, D.C.: U.S. Bureau of the Census.
- Veroff, Joseph, Shirley Hatchett, and Elizabeth Douvan. 1992. "Consequences of Participating in a Longitudinal Study of Marriage." *Public Opinion Quarterly* 56:315-327.
- Wagner, Gert G., Joachim R. Frick, and Jürgen Schupp. 2007. "The German Socio-Economic Panel Study (SOEP): Scope, Evolution and Enhancements." *Schmollers Jahrbuch* 127:139-169.
- Wang, Kevin, David Cantor, and Adam Safir. 2000. "Panel Conditioning in a Random Digit Dial Survey." Pp. 822-827 in *Proceedings of the Survey Research Methods Section of the American Statistical Association*.
- Warren, John Robert and Andrew Halpern-Manners. 2007. "Is the Glass Emptying or Filling Up? Reconciling Divergent Trends in High School Completion and Dropout." *Educational Researcher* 36:335-343.
- Waterton, Jennifer and Denise Lievesley. 1989. "Evidence of Conditioning Effects in the British Social Attitudes Panel Survey." Pp. 319-339 in *Panel Surveys*, edited by D. Kasprzyk, G. J. Duncan, G. Kalton, and M. P. Singh. New York: Wiley.
- Williams, W. H. and C. L. Mallows. 1970. "Systematic Biases in Panel Surveys Due to Differential Nonresponse." *Journal of the American Statistical Association* 65:1338-1349.
- Willis, G., M. Sirken, and G. Nathan. 1994. *The Cognitive Aspects of Responses to Sensitive Survey Questions*. Working Paper Series, No. 9. Hyattsville, MD: National Center for Health Statistics.

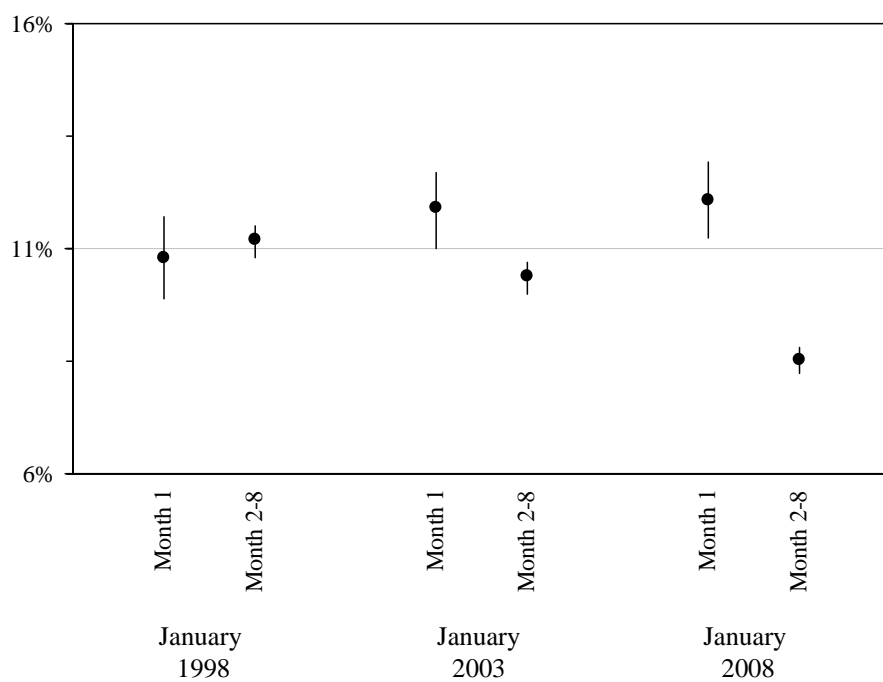
- Wilson, Timothy D., Dana S. Dunn, Jane A. Bybee, Diane B. Hyman, and John A. Rotondo. 1984. "Effects of Analyzing Reasons on Attitude-Behavior Consistency." *Journal of Personality and Social Psychology* 47:5-16.
- Wilson, Timothy D., Delores Kraft, and Dana S. Dunn. 1989. "The Disruptive Effects of Explaining Attitudes: The Moderating Effect of Knowledge about the Attitude Object." *Journal of Experimental Social Psychology* 25:379-400.
- Wilson, Timothy D., Suzanne J. LaFleur, and D. Eric Anderson. 1996. "The Validity and Consequences of Verbal Reports About Attitudes." Pp. 91-114 in *Answering Questions: Methodology for Determining Cognitive and Communicative Processes in Survey Research*, edited by N. Schwarz and S. Sudman. New York: Jossey-Bass Publishers.
- Wu, C. and R.R. Sitter. 2001. "A Model-Calibration Approach to Using Complete Auxiliary Information from Survey Data." *Journal of the American Statistical Association* 96:185-193.
- Yalch, Richard F. 1976. "Pre-Election Interview Effects on Voter Turnout." *The Public Opinion Quarterly* 40:331-336.

ENDNOTES

¹ Note that we use the terms “treatment group” and “control group” for heuristic purposes only. Although respondents are exogenously (and randomly) assigned to the different groups, we recognize that our research design and general analytic strategy does not technically amount to a controlled experiment.

² Our description of the raking procedure, the specific raking variables that we used, coding decisions, and various diagnostic issues, is obviously incomplete. We intend to present a much fuller and more technical discussion of our sample-balancing technique in subsequent drafts, including the draft that we present at the PAA meetings April.

Figure 1. Percent Completing High School via a GED in the CPS Basic Monthly Survey, by Month in Sample and Calendar Month



Note: Sample restricted to CPS reference individuals who were not members of “replacement” households, for whom information was not collected by proxy, for whom the interview was completed via CAPI, and for whom values on the variable of interest were not allocated. Solid dots represent point estimates and lines represent 95% confidence intervals. Post-stratification weights were used to adjust for panel attrition in Months 2-8. See text for further details.