SSM - Population Health 19 (2022) 101233

Contents lists available at ScienceDirect

SSM - Population Health

journal homepage: www.elsevier.com/locate/ssmph

Does participating in a long-term cohort study impact research subjects' longevity? Experimental evidence from the Wisconsin Longitudinal Study

John Robert Warren [a,*], Andrew Halpern-Manners [b], Jonas Helgertz [c]

[a] Department of Sociology, Minnesota Population Center, 909 Social Sciences, 267 19th Ave N, University of Minnesota, USA
[b] Department of Sociology, Indiana University, USA
[c] Department of Economic History and Centre for Economic Demography, Lund University, Sweden

ARTICLE INFO

Keywords:
Panel conditioning
Mortality
Cohort study

ABSTRACT

There is considerable evidence that the act of participating in a survey can alter participants' attitudes, behaviors, and other outcomes in meaningful ways. Considering findings that this form of panel conditioning also impacts health behaviors and outcomes, we investigated the effect of participating in an intensive half-century-long cohort study on participants' longevity. To do so, we used data from a 1957 survey of more than 33,000 Wisconsin high school seniors linked to mortality records. One third of those people were selected at random to participate in the Wisconsin Longitudinal Study (WLS); the other two thirds were never again contacted. Our survival models show no evidence of panel conditioning effects on longevity: People selected at random to participate in the WLS had the same mortality outcomes as their peers who were not selected. This finding holds for the full sample, for women, for men, for population subgroups defined by family socioeconomic origins and educational experiences, and for treatment compliers.

1. Introduction

Cohort studies—research projects in which the same people participate on multiple occasions—are essential sources of information for scientific advances in public health, medicine, the social sciences, economics, and beyond. The resulting longitudinal data play a sustaining role in building knowledge about health, human behavior, social relationships, inequality, and human development—but they also pose several complicated methodological problems. Among the least well-understood of these problems is "panel conditioning," or the bias that results when the very act of participating in the cohort study itself alters participants' outcomes or attributes.

Evidence for substantial panel conditioning effects is now abundant (see, e.g., Struminskaya, 2020; Struminskaya & Bosnjak, 2021; Warren & Halpern-Manners, 2012). Recent research has demonstrated that this form of bias can significantly alter respondents' *answers* to questions about drug, alcohol, and tobacco use (Torche, Warren, Halpern-Manners, & Valenzuela, 2012); physical activity (Williams, Block, & Fitzsimons, 2006); diet and weight loss (Hollywood, Ogden, & Hashemi, 2015); drinking and driving (Halpern-Manners, Warren, & Torche, 2014); condom use (Axinn, Jennings, & Couper, 2015); awareness of health conditions (Wilson & Howell, 2007); labor force participation (Halpern-Manners & Warren, 2012; Krueger, Mas, & Niu, 2017; Solon, 1986); saving for retirement (Crossley, de Bresser, Delaney, & Winter 2017); social networks (Eagle & Proeschold-Bell, 2015); voting (Bartels, 1999); marital satisfaction (Veroff, Hatchett, & Douvan, 1992); health care utilization (Zwane et al., 2011); and willingness to pursue different medical treatment options (Duan, Alegria, Canino, McGuire, & Takeuchi, 2007), to name just a few topical domains. Whether these effects extend to the *behaviors and statuses* that underlie respondents' answers—and whether the behavioral effects are homogeneous across individuals—are open and important questions in the survey methods literature (Bach & Eckman, 2019).

In this article, we describe an experimental study of panel conditioning effects on an outcome that has clear and well-known social and behavioral antecedents: adult mortality. Given the strong potential for panel conditioning effects on a variety of health-related outcomes, we ask whether long-term participation in an intensive longitudinal survey of health behaviors and outcomes impacts the timing of participants' death. To this end, we address two specific questions: First, to what extent does repeatedly answering survey questions—including questions about health behaviors/outcomes and/or aspects of people's lives that are known to correlate strongly with health—influence the timing of respondents' death? Second, are mortality effects from panel

* Corresponding author.
E-mail address: warre046@umn.edu (J.R. Warren).

https://doi.org/10.1016/j.ssmph.2022.101233
Received 21 July 2022; Received in revised form 7 September 2022; Accepted 15 September 2022
Available online 28 September 2022

conditioning distributed evenly across the population or do they vary systematically depending on respondents' background characteristics (e.g., sex, family socioeconomic background, educational performance)?

2. Theory

2.1. Conceptual model

Cohort studies are complex social experiences that have a well-demonstrated potential to change participants' beliefs, behaviors, and knowledge in subtle but consequential ways (Struminskaya & Bosnjak, 2021; Tourangeau, Rips, & Rasinski, 2000). Reflecting on one's behaviors (during a survey or clinic visit or in their aftermath) may prompt individuals to take stock of their lifestyle choices, to seek out additional information, to catalyze or reevaluate their beliefs, and/or to take steps towards improving or modifying some aspect(s) of their lives (Struminskaya & Bosnjak, 2021; Warren & Halpern-Manners, 2012). This may be especially true in cases in which the substantive focus of the cohort study deals with topics that are especially salient or are of some potential utility (Warren & Halpern-Manners, 2012), when the experience of participating conveys new and useful pieces of information (Das, Toepoel, & Soest, 2007), when they prompt future oriented thinking (Oh, Yeatman, & Trinitapoli, 2019), or when the experience provokes a sharp emotional response (Baumeister, Vohs, DeWall, & Zhang, 2007). If the subsequent changes induced by this process impact people's health—either *proximately* (*P*) by altering their health behaviors or *distally* (*D*) by influencing behaviors or statuses that predict health—they could translate into changes in longevity (*M*). We summarize this argument graphically in Fig. 1.

Research by Zwane et al. (2011) provides a useful case in point. Their results—based on a series of carefully designed field experiments in several developing countries—suggest that more frequently participating in surveys can lead to immediate changes in respondents' health behaviors ($S \rightarrow P$ in Fig. 1). In one experiment, they found that participants who received questions about health insurance coverage and various health-related risks (the treatment) were significantly more likely to take up insurance, as compared to members of the control group who received no such questions. In another experiment, they showed that receiving questions about their health and health conditions led to increases in respondents' use of products designed to treat those conditions and to improve health more generally. Both results were validated using external data. Zwane et al. (2011) speculate that these findings could be due to the subtle but important behavioral cues provided by surveys, their useful informational content, or a combination thereof.

As we have already noted, the act of participating in a cohort study can set in motion a series of cognitive processes that could change respondents' attitudes or behaviors in important ways (Struminskaya, 2020; Sturgis, Allum, & Brunton-Smith, 2009). If these changes are themselves predictive of health or more proximate variables that are thought to affect health, indirect pathways connecting exposure to the cohort study and mortality could also emerge ($S \rightarrow D \rightarrow M$ and $S \rightarrow D \rightarrow P \rightarrow M$ in Fig. 1). The methodological literature on panel conditioning anticipates a number of ways this might occur, including indirect effects that arise through changes in marital satisfaction (Veroff et al., 1992), changes in employment status (Halpern-Manners & Warren, 2012), changes in family size preferences (Oh et al., 2019), and/or changes in fertility (Yeatman, Sennott, & Culpepper, 2013). Importantly for us, these variables are all correlates of health and mortality.

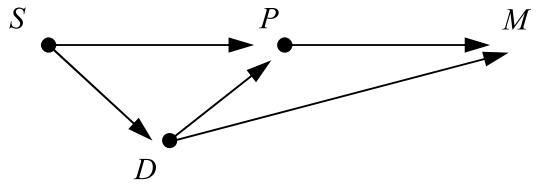

Fig. 1. Conceptual model linking survey exposure to mortality.
Note: *S* represents level of exposure to the cohort study; *D* and *P* represent *distal* (e.g., social participation or income) and *proximate* (e.g., access to health information or awareness of health problems) causes of adult health; and *M* represents mortality (i.e., timing of death).

2.2. Implications for substantive research on mortality and its correlates

Panel conditioning effects on longevity could impact scholarship on mortality and its correlates, and those impacts may vary depending on how the effects of panel conditioning are distributed. If the experience of participating in a cohort study leads (through distal or more proximate channels) to behavioral changes that are *consistent* across participants—and that distinguish them (and the timing of their eventual death) from other non-participants in the target population—*then researchers' ability to make out-of-sample generalizations based on survey data may be diminished.* This is primarily an issue of external validity.

An additional problem would arise if the effects of panel conditioning *vary* across sub-groups, in either direction or magnitude. This could occur if certain segments of the population invest more heavily in the experience of participation (Cantor, 2008; Tourangeau, 2000), reflect more deeply on their survey answers and other activities, or are more likely to retain and use the information they acquire as a result of being participants (Warren & Halpern-Manners, 2012). If these processes prompt behavioral changes within some subgroups (e.g., more highly educated respondents) more so than within others, and if these behavioral changes are consequential with respect to future health outcomes, they could lead to *biased estimates of multivariate relationships* (e.g., the association between education and mortality). This would raise concerns about internal validity and inference in longitudinal survey-based research. Indeed there is some evidence—albeit limited—suggesting that different demographic and socioeconomic subgroups invest to differing degrees in the survey process and may thus be differentially susceptible to panel conditioning (e.g., Battaglia, Zell, & Ching, 1996; Cantor, 2008; Clausen, 1968).

These are high-stakes problems. If partaking in a cohort study alters participants' long-term mortality-risk profile—by inducing health-relevant behavioral changes, by conveying new pieces of health information or knowledge that the participant did not previously possess, or by influencing more distal causes of different health-related outcomes and behaviors—then our understanding of mortality based on these kinds of data, as well as its causes and correlates, could be altered. Despite its importance across a wide variety of scholarly and applied fields, we are not aware of any study that has examined this issue using a strong population-based experimental design like the one we employ here.

3. Research design

3.1. Data

We utilize a unique population level dataset that allows for a direct comparison between (1) a group of individuals who—by chance—completed a survey just once in adolescence and (2) their cohortmates who—also by chance—have been in an intensive and high-retention cohort study for more than half a century after completing that same survey in adolescence. The substantive content of the surveys and other interventions that comprise the cohort study focus on many topics directly and proximally related to health behaviors and outcomes.

In the spring of 1957, the state of Wisconsin administered an in-school survey to all graduating high school seniors to understand their

educational experiences, achievements, plans, and aspirations. After the 1957 survey, a one-third sample of the seniors was selected at random for subsequent follow-up surveys; this one-third sample became the Wisconsin Longitudinal Study (WLS; Herd, Carr, & Roan, 2014). WLS respondents were re-interviewed in 1975, 1992, 2004, and 2011 (respondents who missed one or more waves due to non-response remained eligible for subsequent rounds of the study). The 1992 through 2011 surveys were lengthy and featured extensive survey and anthropometric measures of health behaviors, health outcomes, and cognitive functioning. Sample retention in the WLS has been exceptionally high, with response rates ranging from 84% (in 2011) to 90% (in 1975) of surviving respondents (Herd et al., 2014). The two-thirds of the original group of graduating seniors that was not randomly selected to participate in the WLS has never been part of the cohort study.

The sampling strategy underlying the WLS allows for a comparison between those who were randomly selected into the longitudinal WLS panel—the treatment group, which was subsequently invited to participate in what became an intensive cohort study—and those who were not. If we observe significant differences between the two groups with respect to timing of death, we can infer that these differences are due to behavioral changes induced by panel conditioning. Because we have demographic, socioeconomic background, and educational characteristics of individuals from their 1957 surveys, we can also assess heterogeneity in the effects of panel conditioning across diverse groups of individuals.

All surveys from the spring of 1957 are available in PDF format and were keypunched under our supervision in 2020 in such a way that keypunchers did not know which records were in the 1/3 (WLS, or treatment) group and which were in the 2/3 (non-WLS, or control) group. All surveys were independently keypunched twice; when there were discrepancies or entry errors, a third independent person resolved them.

The content of the WLS cohort study makes it an ideal testing ground for panel conditioning effects on timing of death. Members of the WLS panel have been asked questions about their social background, educational outcomes, labor market experiences, occupational attainments, job characteristics, asset accumulation, fertility, marital status and marital history, social participation, social support, civic participation, internet use, religion and spirituality, intergenerational relationships, living arrangements, retirement planning, and end-of-life preparations through a series of lengthy face-to-face, mail, and telephone interviews. Most of these variables—which we refer to above as *distal* predictors—are known correlates of health behaviors, health outcomes, and mortality.

Although the time between waves makes panel conditioning somewhat less likely (Halpern-Manners, Warren, and Torche 2012), the intensity of the survey experience and the considerable buy-in required of respondents could have a countervailing effect. Besides the broad set of items listed above, respondents have been asked an extensive battery of questions that deal directly with health and well-being as a part of mail-back surveys of at least 50 pages in length, hour-long phone surveys, and 2-hour long in-person interviews that also featured anthropometric measures. Survey questions from 1992 onward have covered issues like heart problems, cancers and malignant tumors, weight gain and obesity, diabetes and high blood sugar, arthritis and rheumatism, mental illness and depression, high blood pressure, stress, sleep behaviors, nutrition and eating habits, tobacco and alcohol use, prescription drug use, levels of physical activity, ambulation, vision and hearing, disabilities, surgeries and other medical procedures, health literacy, access to health resources and information, health care utilization, relationships with doctors, and health care coverage. Many of the survey batteries are extensive. For example, in 1992 WLS women were asked to complete a five-page paper survey on menopause, and in 2004 all sample members were asked to complete a 50+ page questionnaire (in additional to a telephone survey); the first 10 pages were entirely measures of health behaviors and outcomes. These are all more *proximate* predictors of mortality (Cutler & Lleras-Muney, 2010; Rogers, Hummer, & Nam, 1999).

3.2. Measuring timing of death

Records from the 1957 survey include each respondent's first and last name (and in some cases middle name or initial), age, a parent or guardian's name, name of high school, and residential addresses. This information—in combination with reasonable assumptions about the state (Wisconsin) where respondents' social security numbers were most likely issued—is sufficient to locate death records in the Social Security Death Master File (SSDMF), in the public version of the NUMIDENT (the Social Security Administration's Numerical Identification System), and in online genealogy databases (e.g., Ancestry.com or Findagrave.com). Because respondents to the 1957 survey were all high school seniors at the time, their years of birth have a narrow range—77.4% were born in 1939 and 16.3% were born in 1938, with only 0.2% born outside the range 1937–1940. This means that the distribution of *year* of death quite closely resembles the distribution of *age* at death.

3.3. Linking 1957 surveys to mortality records

To link the 1957 surveys to mortality records we relied on the following information about students from the surveys: first and last name, year of birth (derived from reported age in 1957), sex, residential location, and parent's/guardian's name (including surname). All linkages were carried out after blinding the records so that we did not know whether students were in the treatment group or the control group. None of the linkages to mortality records used information other than what was contained in the 1957 surveys (i.e., nothing from the 1975 through 2011 WLS surveys was used since that information is unavailable for people in the control group). Both steps are meant to prevent the introduction of differences between the treatment and control groups that could bias or otherwise compromise our results.

An intuitive approach to linking 1957 survey records to mortality data would be to have a team of research assistants search, by hand, using genealogical resources like Ancestry.com. Although this approach can produce large numbers of good matches, it (1) is not reproducible and (2) is prohibitively costly in terms of time and money (Feigenbaum, 2016). Instead, we trained a machine learning algorithm similar to one used by Feigenbaum (2016) and elaborated by others (see, e.g., Fu, Boot, Christen, & Zhou, 2014; Goeken, Huynh, Lynch, & Vick, 2011; Ruggles, 2014). As has been recommended in more recent methodological research on record linkage, our algorithm considers a more extensive set of linking attributes in order to enhance both accuracy of linkages and overall linkage rates (Bailey et al., 2022; Helgertz et al., 2022). The basic procedure, which we describe below, proceeded in a series of steps.

First, we only compared each of the 1957 survey records to a plausible set of records in the SSDMF and the NUMIDENT. That is, instead of comparing each one of the 1957 survey records to each one of the tens of millions of mortality records, we compared each one of the 1957 survey records to records for people in the mortality databases who are similar with respect to age (plus or minus 2 years) and first and last name; for linking to the NUMIDENT, we also considered the similarity of the name of the guardian from the WLS to the father, as reported in the NUMIDENT. The similarity of names was based on their Jaro-Winkler distance

(Winkler, 2004).

Second, we trained a computer linking algorithm to determine which—if any—of the plausible set of matches was correct. The first step was to produce "training data"—a set of records in which the truth is known. For 1,000 randomly selected cases, we carefully trained a team of research assistants to use online resources (e.g., Ancestry.com) to determine which—if any—of the plausible set of mortality records was the focal Wisconsinite; this determination was made when one and only one mortality record logically matched. Each case was linked by more than one hand linker, and we only declared a match to mortality records when the majority of hand linkers agreed.

Next, we fit a logistic regression model to the training data, where the [0,1] variable containing manually declared matches in the training data was regressed onto a selection of individual- and pair-level predictors (e.g., age, sex, discrepancy in age, individual's first and last name Jaro-Winkler distance, father's or guardian's Jaro-Winkler distance, state of birth). The results of this model informed the algorithm as to which, if any, of the plausible set of matches should be considered a true match. In training the algorithm, we repeatedly split the training data into two, fitting the model on one half and testing how it performed on the other half, where the "truth" was known. The algorithm declared a unique match based on (1) the greatest similarity between any 1-to-1 match within any given individual (technically the predicted probability based on the logistic regression estimates) and (2) the relative difference between the best and the second-best possible match within any given individual. By looping multiple times over a range of realistic values on both parameters, and provided that the training data resembles the population it is meant to represent, we were able to choose values on both that optimized the overall performance of the algorithm. Here, our main objectives were to minimize the presence of false positives (incorrectly matched cases) while at the same time maximizing the number of true positives (correctly matched cases) and true negatives (correctly unmatched cases).

In selecting thresholds for declaring matches in our data, we used the Matthew's Correlation Coefficient (MCC), which is an especially useful measurement for two-class data where the classes are not very balanced (Chicco, 2017). This is definitely the case in our situation, since the vast majority of the Wisconsinites in the training data had multiple plausible matches in the mortality records but only (at most) one actual match. The MCC, in Eq. (1) below, compares the predictions of the algorithm to all possible outcomes (true/false positives/negatives) and provides a single metric (ranging from 0 to 1) to be used to select which thresholds to use. The formula is as follows, where *TP* represents the number of true positives, *TN* represents true negatives, *FP* represents false positives, and *FN* represents false negatives:

$$\mathrm{MCC} = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad (1)$$

After optimizing the machine learning algorithm, we applied the final prediction equation to the complete set of 1957 survey records and their sets of plausible matches. We performed separate machine linkages to the SSDMF and NUMIDENT, and then merged the results. In the few cases where a WLS record was linked to different individuals between the two datasets, we opted for the NUMIDENT link due to the greater amount of linking information available.

3.4. Measures

Excluding a handful of cases in which either (1) the 1957 survey respondent was not between the ages of 16 and 19 in 1957 or (2) gender was not recorded, a total of 33,686 people's 1957 surveys were keypunched and linked to mortality records. As shown in Table 1, we identified 1,664 people in the treatment group and 3,653 people in the control group as deceased in SSDMF or NUMIDENT files; this equates to 16.3% of the treatment group and 15.6% of the control group. In Fig. 2, we plot the cumulative number of surviving 1957 survey respondents by year and by gender, separately for the treatment and control groups. Whereas Table 1 shows that the final percentage of deceased survey respondents was very similar across treatment and control groups, Fig. 2 shows that age patterns of death were also quite similar.

Table 1
Descriptive statistics, by treatment group and mortality status.

	Treament Group (in WLS)		Control Group (Not in WLS)	
	Alive (%)	Dead (%)	Alive (%)	Dead (%)
Father's Education				
Missing	8.0	8.1	7.8	9.2
No HS Diploma	52.5	52.7	51.4	52.6
Just HS Diploma	17.2	18.7	18.1	17.1
Some College or More	22.4	20.6	22.8	21.1
Mother's Education				
Missing	6.8	8.5	6.4	8.6
No HS Diploma	47.0	45.4	45.2	44.2
Just HS Diploma	26.5	27.8	27.8	29.5
Some College or More	19.7	18.3	20.7	17.7
Family Income				
Missing	2.9	3.1	3.0	3.1
Below Average	6.9	7.0	6.6	6.5
Average	70.1	69.3	70.8	70.0
Above Average	20.1	20.6	19.6	20.3
Plan to Go to 4-Year College				
Missing	52.0	57.4	50.9	56.7
No	19.4	15.4	20.1	17.2
Yes	28.6	27.1	29.0	26.1
Took College Prep Classes				
Missing	12.6	14.1	12.4	15.0
No	52.1	53.9	52.0	54.3
Yes	35.3	32.0	35.6	30.7
Gender				
Male	44.7	76.7	44.7	76.4
Female	55.3	23.3	55.3	23.6
Row % Within Group	83.7	16.3	84.4	15.6
N	8,572	1,664	19,797	3,653

Note: Sample restricted to the 33,686 Wisconsin high school seniors in 1957 who completed surveys, who were between ages 16 and 19 at that time, and for whom gender is known.

Note that the flattening of survival curve for women in Fig. 2 is the result of the fact that relatively fewer women are linked to the NUMIDENT after 2007; the NUMIDENT includes last name at birth, thereby increasing linkage rates for women as compared to the SSDMF. However, because of our experimental design, there is no reason to expect differences between the treatment and control groups with respect to relative likelihood that women are linked.

In Table 1 we report descriptive statistics for variables used in our analyses, separately for people in the treatment and control groups and—within those groups—based on whether they were identified as deceased. All measures come from the 1957 survey. We include three measures of family socioeconomic background: mother's educational attainment, father's educational attainment, and family economic circumstances. The latter is measured using a question that asked: "In terms of income or wealth of families in my community, I think my family is: (1) considerably above average, (2) average, (3) somewhat above average, (4) somewhat below average, or (5) considerably below average." We observe very few differences in socioeconomic background across the treatment or control groups.

We also include two measures of educational performance: An indicator of whether the respondent said they planned to go to a four-year college or university and an indicator of whether they took college preparatory courses in high school. Again, we observe very few differences between people in the treatment and control groups.

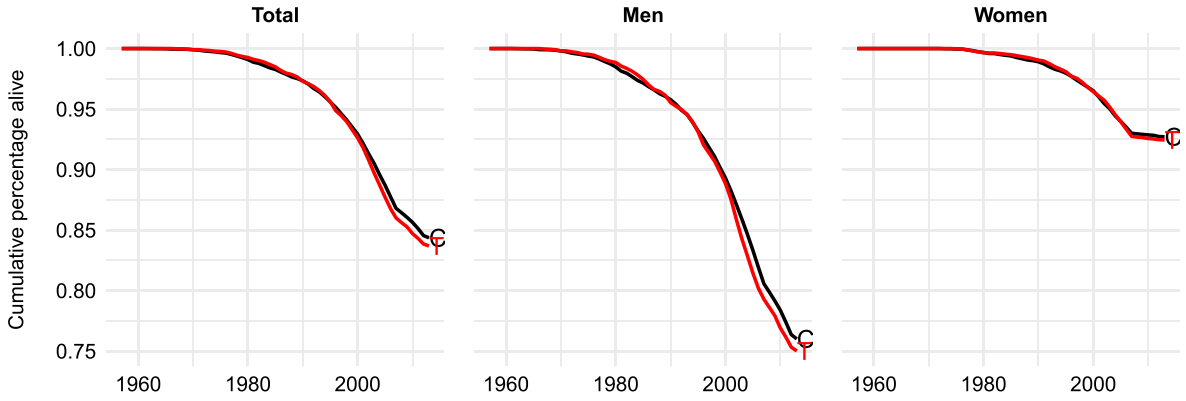

Fig. 2. Cumulative percentage alive, by year, treatment group, and gender. Sample restricted to the 33,686 Wisconsin high school seniors in 1957 who completed surveys, who were between ages 16 and 19 at that time, and for whom gender is known. Members of the treatment group (whose survival curve is shown in red and labeled with a T) were a part of the 1/3 random sample that was selected to participate in the Wisconsin Longitudinal Study (WLS). Members of the control group (whose survival curve is shown in black and labeled with a C) were not. See text for more details.

3.5. Modeling strategy

We fit a series of multivariate regression models to obtain information about panel conditioning effects on mortality. In this section, we specify our models and link them to the research questions we set forth earlier.

To what extent does repeatedly answering survey questions—including questions about health behaviors/outcomes and/or aspects of people's lives that are known to correlate strongly with health—influence the timing of respondents' death? To evaluate the *main effect* of being in a cohort study, we fit a series of Cox proportional hazard models, with age serving as the underlying time metric (Kom, Graubard, & Midthune, 1997). The baseline model can be written as:

$$\log h_i(t) = \log h(t) + \alpha P_i + \sum_j \beta_i X_{ji} \quad (2)$$

where $h(t)$ is the unspecified baseline hazard function; t measures age; i refers to individuals ($i = 1, \ldots, N$); P to their treatment status (i.e., as panel members in the WLS); and X to other covariates ($j = 1, \ldots, J$) that were measured at the time of the 1957 survey. The key parameter of interest, α, captures the relationship between random assignment to the longitudinal sample and an individual's hazard of dying. Although not strictly necessary due to the experimental design, additional controls are included in the model to account for residual variation and increase the precision of our estimates.

Estimates obtained from Eq. (2) provide information about "intent-to-treat" (ITT) effects. This approach ensures that we are not making comparisons between long-time participants in the WLS (who represent a non-random, and potentially healthier, subset of the respondents who were originally randomized into the longitudinal panel) and members of the control group (who have not been subset in this way). Later, we describe a modified version of this approach that allows us to explore issues related to non-compliance—i.e., non-participation in one or more follow-up waves and/or attrition from the study—among members of the treatment group.

Are mortality effects from panel conditioning distributed evenly across the population or do they vary systematically depending on respondents' background characteristics (e.g., sex, family socioeconomic background, and educational performance)? The estimates we obtain from Eq. (2) provide information about the average treatment effect (ATE) associated with assignment to the longitudinal WLS panel. To determine whether panel conditioning effects vary across different segments of the population, we also estimated conditional average treatment effects (CATEs) using a modified version of Eq. (2) that contained interactions between our treatment indicator, P, and key covariates. In these analyses, we examine a series of treatment-by-covariate interactions, where the covariates consist of measures of (1) social background—including mother's education, father's education, and subjective family income; (2) educational activities—including college plans and college preparatory course taking; and (3) gender. All of these variables were measured in the 1957 survey, prior to treatment assignment.

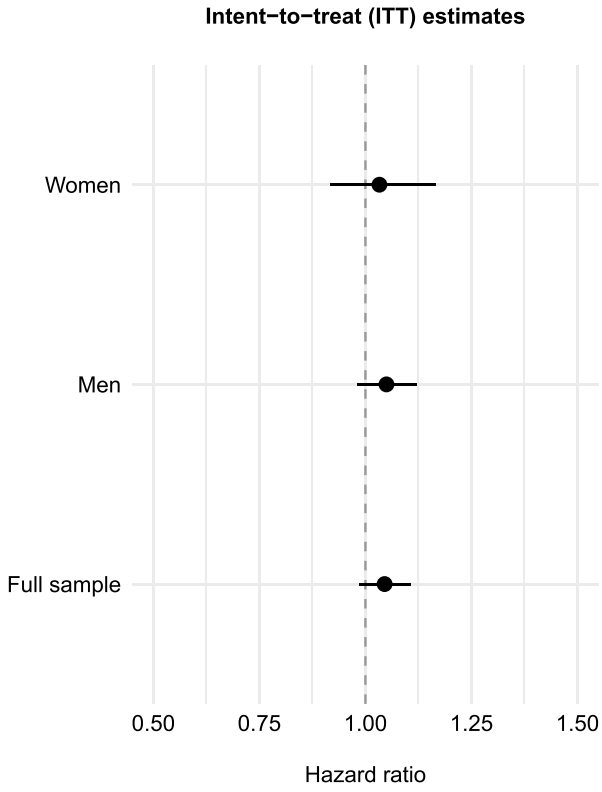

Fig. 3. Estimated effect of treatment from intent to treat analysis.
Note: Sample restricted to the 33,686 Wisconsin high school seniors in 1957 who completed surveys, who were between ages 16 and 19 at that time, and for whom gender is known. Results from Cox proportional hazard models which express time to death as a function of treatment group membership net of the family socioeconomic background, education, and gender variables shown in Table 1. Figure shows point estimates and 95% confidence intervals.

Table 2
Selected results from models of time to death as a function of treatment, socioeconomic background, educational attributes, gender, and interactions between treatment and covariates.

Model 1	Full Sample			Men			Women		
	Hazard	s.e.	95% C.I.	Hazard	s.e.	95% C.I.	Hazard	s.e.	95% C.I.
Treatment	0.94	0.10	(0.77 - 1.15)	0.83	0.10	(0.66 - 1.05)	1.51	0.33	(0.99 - 2.31)
Father's Education (vs Missing)									
No HS Diploma	1.01	0.07	(0.88 - 1.16)	0.92	0.07	(0.79 - 1.07)	1.43	0.23	(1.04 - 1.96)
Just HS Diploma	0.95	0.07	(0.81 - 1.10)	0.88	0.08	(0.75 - 1.04)	1.26	0.22	(0.89 - 1.78)
Some College or More	0.98	0.07	(0.85 - 1.14)	0.91	0.08	(0.77 - 1.07)	1.35	0.23	(0.96 - 1.89)
Treatment x Father's Education (vs Missing)									
No HS Diploma	1.09	0.12	(0.88 - 1.36)	1.28	0.16	(1.00 - 1.63)	0.62	0.14	(0.40 - 0.98)
Just HS Diploma	1.25	0.15	(0.98 - 1.59)	1.36	0.19	(1.03 - 1.80)	0.88	0.23	(0.53 - 1.46)
Some College or More	1.08	0.13	(0.85 - 1.37)	1.26	0.17	(0.96 - 1.65)	0.60	0.16	(0.36 - 1.00)

Model 2	Full Sample			Men			Women		
	Hazard	s.e.	95% C.I.	Hazard	s.e.	95% C.I.	Hazard	s.e.	95% C.I.
Treatment	0.97	0.10	(0.80 - 1.19)	0.96	0.11	(0.77 - 1.19)	1.05	0.25	(0.65 - 1.69)
Mother's Education (vs Missing)									
No HS Diploma	0.87	0.06	(0.75 - 1.00)	0.89	0.07	(0.76 - 1.04)	0.81	0.13	(0.59 - 1.11)
Just HS Diploma	0.89	0.07	(0.77 - 1.03)	0.93	0.08	(0.79 - 1.09)	0.78	0.13	(0.56 - 1.09)
Some College or More	0.82	0.06	(0.70 - 0.96)	0.87	0.08	(0.73 - 1.03)	0.69	0.12	(0.48 - 0.97)
Treatment x Mother's Education (vs Missing)									
No HS Diploma	1.06	0.12	(0.86 - 1.32)	1.11	0.14	(0.87 - 1.41)	0.91	0.23	(0.55 - 1.51)
Just HS Diploma	1.05	0.12	(0.84 - 1.32)	1.05	0.13	(0.82 - 1.35)	1.03	0.28	(0.60 - 1.75)
Some College or More	1.19	0.15	(0.93 - 1.51)	1.19	0.16	(0.91 - 1.56)	1.14	0.32	(0.66 - 1.97)

Model 3	Full Sample			Men			Women		
	Hazard	s.e.	95% C.I.	Hazard	s.e.	95% C.I.	Hazard	s.e.	95% C.I.
Treatment	1.11	0.19	(0.80 - 1.54)	1.10	0.21	(0.76 - 1.58)	1.19	0.45	(0.57 - 2.48)
Family Income (vs Below Average)									
Below Average	1.10	0.13	(0.88 - 1.38)	1.09	0.14	(0.85 - 1.41)	1.14	0.30	(0.68 - 1.91)
Average	1.13	0.11	(0.94 - 1.37)	1.14	0.12	(0.93 - 1.41)	1.13	0.26	(0.72 - 1.77)
Above Average	1.12	0.12	(0.92 - 1.38)	1.09	0.12	(0.87 - 1.36)	1.29	0.31	(0.81 - 2.07)
Treatment x Family Income (vs Below Average)									
Below Average	0.98	0.20	(0.66 - 1.46)	1.03	0.23	(0.66 - 1.61)	0.81	0.36	(0.34 - 1.94)
Average	0.93	0.16	(0.67 - 1.30)	0.94	0.18	(0.64 - 1.36)	0.90	0.34	(0.42 - 1.90)
Above Average	0.95	0.17	(0.67 - 1.35)	1.00	0.20	(0.67 - 1.48)	0.75	0.30	(0.34 - 1.66)

Model 4	Full Sample			Men			Women		
	Hazard	s.e.	95% C.I.	Hazard	s.e.	95% C.I.	Hazard	s.e.	95% C.I.
Treatment	1.04	0.04	(0.96 - 1.12)	1.04	0.05	(0.96 - 1.14)	1.03	0.09	(0.88 - 1.21)
Planning to Go to 4-Year College (vs Missing)									
No	0.94	0.04	(0.85 - 1.02)	0.92	0.05	(0.82 - 1.02)	0.97	0.08	(0.82 - 1.15)
Yes	0.82	0.04	(0.75 - 0.90)	0.83	0.04	(0.75 - 0.91)	0.80	0.08	(0.66 - 0.97)
Treatment x Planning to Go to 4-Year College (vs No)									
No	0.94	0.08	(0.80 - 1.11)	0.97	0.10	(0.80 - 1.19)	0.88	0.14	(0.65 - 1.19)
Yes	1.06	0.07	(0.92 - 1.21)	1.03	0.08	(0.89 - 1.20)	1.14	0.17	(0.85 - 1.54)

Model 5	Full Sample			Men			Women		
	Hazard	s.e.	95% C.I.	Hazard	s.e.	95% C.I.	Hazard	s.e.	95% C.I.
Treatment	1.01	0.08	(0.87 - 1.18)	1.01	0.09	(0.86 - 1.20)	1.00	0.19	(0.69 - 1.44)
Took College Preparatory Classes (vs No)									
No	0.96	0.05	(0.87 - 1.06)	0.93	0.05	(0.84 - 1.04)	1.07	0.12	(0.86 - 1.34)
Yes	0.89	0.05	(0.80 - 0.99)	0.85	0.05	(0.76 - 0.97)	1.02	0.13	(0.80 - 1.30)
Treatment x Took College Preparatory Classes (vs No)									
No	1.02	0.09	(0.86 - 1.22)	1.02	0.10	(0.84 - 1.23)	1.03	0.21	(0.69 - 1.53)
Yes	1.08	0.10	(0.90 - 1.30)	1.08	0.11	(0.88 - 1.33)	1.06	0.23	(0.70 - 1.61)

Model 6	Full Sample			Men			Women		
	Hazard	s.e.	95% C.I.	Hazard	s.e.	95% C.I.	Hazard	s.e.	95% C.I.
Treatment	1.05	0.04	(0.98 - 1.12)						
Gender (vs Male)									
Female	0.28	0.01	(0.26 - 0.30)						
Treatment x Gender (vs Male)									
Female	0.99	0.07	(0.86 - 1.13)						

Note: Sample restricted to the 33,686 Wisconsin high school seniors in 1957 who completed surveys, who were between ages 16 and 19 at that time, and for whom gender is known.

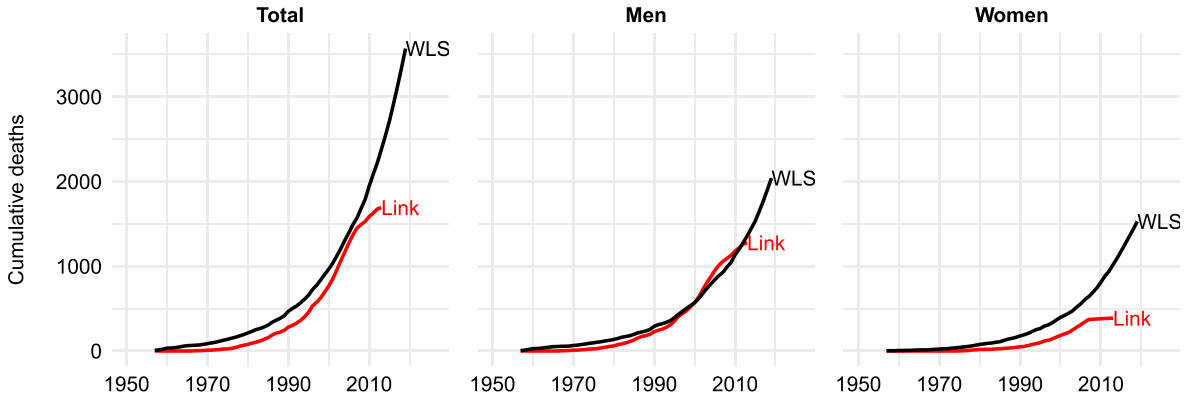

Fig. 4. Cumulative number dead, by year, sample, and gender. Mortality curves for members of the treatment group, as ascertained from internal WLS records, are shown in black. Mortality curves for the same group, as ascertained from probabilistic links to the SSDMF and NUMIDENT, are shown in red. See text for more details.

4. Results

To what extent does repeatedly answering survey questions—including questions about health behaviors/outcomes and/or aspects of people's lives that are known to correlate strongly with health—influence the timing of respondents' death? Fig. 3 depicts the key coefficient estimate—α from Eq. (2) above—which reflects the relationship between random assignment to the longitudinal sample and an individual's hazard of dying; the figure also includes 95% confidence intervals around that coefficient estimate, and depicts separate results for the full sample, for men, and for women. The models include the covariates described in Table 1; coefficients for those covariates are not shown.

In short, Fig. 3 shows no evidence that assignment to the treatment group had any effect on a respondent's hazard of death. Those people randomly assigned to the treatment group—and thus participating for half a century in the Wisconsin Longitudinal Study—look no different with respect to hazard of death as compared to their peers who were assigned to the control group.

Are mortality effects from panel conditioning distributed evenly across the population or do they vary systematically depending on respondents' background characteristics (e.g., sex, family socioeconomic background, and educational performance)? Table 2 reports results from models that allow for statistical interactions between treatment group assignment and father's education (Model 1), mother's education (Model 2), family income (Model 3), college plans (Model 4), college preparatory coursework (Model 5), and gender (Model 6). The table reports separate results for the full sample, men, and women.

The key results in Table 2 are those pertaining to coefficients for (1) treatment group assignment and (2) interactions between treatment group assignment and the focal covariate. In no models—for the full sample, men, or women—are any of those key coefficients statistically significant. Whereas Fig. 3 showed that the ATE estimates do not differ statistically from zero, Table 2 shows that none of the CATE estimates do either. In other words, we find no effect of treatment group assignment for subgroups defined by family socioeconomic background, educational experiences, or gender.

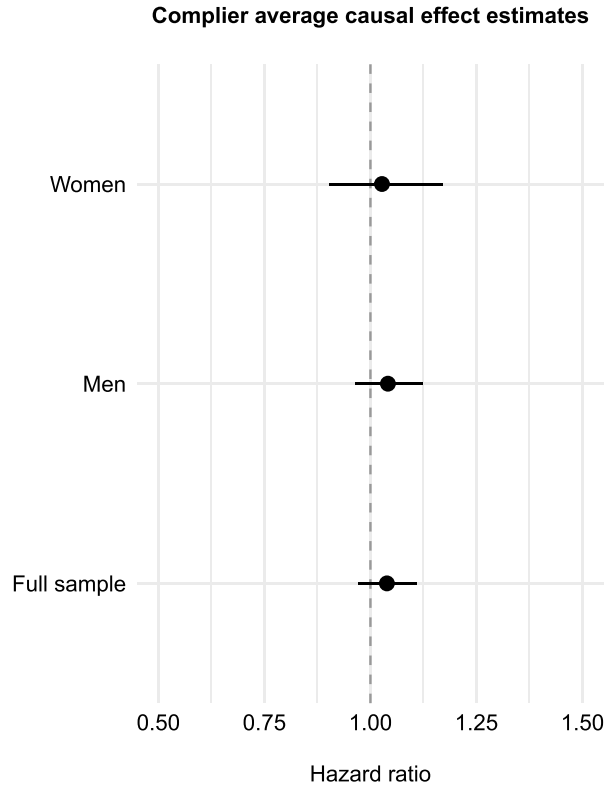

Fig. 5. Estimated effect of treatment from complier average causal effects analysis.
Note: Sample restricted to the 33,686 Wisconsin high school seniors in 1957 who completed surveys, who were between ages 16 and 19 at that time, and for whom gender is known.. Results from a complier average causal effect analyses with a Weibull family link function. Figure shows point estimates and 95% confidence intervals.

4.1. Assumptions and robustness checks

One potential limitation with the results shown in Fig. 3 and Table 2 is that our possible failure to link all deceased survey respondents to mortality records could bias our results. To explore this possibility, we compared the distribution of year of death in the treatment group—the 1/3 sample linked by us to SSDMF and NUMIDENT mortality records—to the same distribution as recorded by the WLS. As noted above, the WLS routinely merges their records to the National Death Index and supplements that information with obituary records and information provided by key informants. Fig. 4 shows those distributions, separately for the full sample, men, and women.

Among men, the results of our efforts to link the treatment group to mortality records closely resembles information in the WLS data (except that our linked records only observe deaths occurring through 2013). Among women, we were less successful in linking; this is mainly because the SSDMF does not contain information about name at birth (NUMIDENT provides father's surname). However, the important

consideration is that our linking methods—which were more successful among men than among women—were performed the same way (and blinded) for the treatment and control groups. There is no reason to believe that the treatment and control groups differ with respect to their propensity to appear in these administrative mortality records—or to be linked to the 1957 surveys. While the external validity of our findings (especially for women) may thus be threatened, internal validity should be intact.

Another possible threat to the validity of our results concerns the fidelity of random assignment. Identifying panel conditioning effects requires an appropriate counterfactual. Because members of the WLS's longitudinal sample (the "treatment group") were selected at random from the broader population of graduating high school seniors in 1957, we can obtain an estimate of the counterfactual by examining the mortality outcomes of respondents who participated in the original 1957 survey *but who were not randomly sampled into the WLS's longitudinal panel and thus were not re-interviewed in subsequent waves* (the "control group"). Due to random assignment, the two groups should, within sampling error, be equivalent in all respects except for their levels of WLS survey exposure. As shown in Table 1, this appears to be the case.

A third possible threat to the validity of our results is treatment noncompliance. Although rates of attrition from the study have been relatively low over time, not everyone who was assigned to the treatment group participated in all waves of the WLS (Herd et al., 2014). If panel conditioning (or the magnitude of its effects) is dependent on *repeated* exposure to health-relevant questions on survey instruments, a failure to participate in one or more waves (i.e., noncompliance or partial compliance with the assigned treatment protocol) could dilute our estimates of the treatment effect, resulting in overly conservative estimates (Greenland, Lanes, & Jara, 2008; White & Pocock, 1996). To explore this possibility, we calculated estimates of the complier-average causal effect (CACE) where compliance was defined among treatment group members as participating in all of the telephone, mail, and in-person WLS surveys administered after 1957, conditional on survival to that wave (Little & Rubin, 2000).

These analyses draw on methodological advances in modeling time-to-event data when there is nonrandom noncompliance (or partial compliance) in the treatment arm. Although space constraints prevent us from presenting a full derivation of the model, the intuition is straightforward: to make valid comparisons between compliers in the treatment arm (e.g., panelists who participated in all of the 1957 through 2011 surveys) and "would-be compliers" in the control arm (e.g., members of the control group who would have participated in those same waves had they been assigned to the treatment group), we need a way to identify compliers in each group (Imbens and Ruben 1997; Skrondal and Rabe-Hesketh 2004). This is trivial for members of the treatment group (whose compliance is fully observed) but requires special steps for members of the control group (whose potential compliance, had they been assigned to the treatment, is unobserved). In our analyses, we use the latent variable approach proposed by Muthén (2002), which produces an estimate of the "true" compliance status for members of the control group using known information about those individuals (see, also, Troncoso and Morales-Gomez 2022).

In Fig. 5 we report the results of the CACE analyses. Although the confidence intervals are somewhat wider (because only 3,997 members of the treatment group fully complied with treatment), the point estimates are virtually identical to those from the ITT analyses (Fig. 3). For the full sample, for men, and for women, we find no effect of panel conditioning on timing of death.

5. Discussion

At the outset we drew on substantial evidence of panel conditioning effects on health behaviors and health outcomes to argue that there is a strong theoretical rationale for expecting that participating in an intensive longitudinal study might impact the timing of people's death. Being prompted to repeatedly respond to long sets of questions about health behaviors, health care utilization, health conditions, and other health-relevant topics may force people to think and act differently than they otherwise would have. We suspected that this would lead to panel conditioning such that people who participated in such a cohort study would live longer. We further speculated that such effects might differ across population subgroups.

We tested these ideas using records from more than 33,000 Wisconsin high school seniors who completed a survey just before graduation. One third of that group was selected—strictly at random—to participate in the Wisconsin Longitudinal Study (WLS). The other two thirds were never again contacted. To obtain information on survival, we linked all the seniors to mortality records—making sure that people in both the treatment and control groups were treated equivalently with respect to data entry and record linkage.

In short, we found no evidence of panel conditioning effects on mortality. Any differences that emerged between people selected to participate in the WLS and the counterfactual group with respect to survival tended to be small and non-significant. This finding held for the full sample, for women, and for men; it held across population subgroups differing with respect to socioeconomic background and educational experiences; and it held in supplementary models that focused only on compliers.

Although there was good theoretical reason to suppose that the treatment—participating in a cohort study for more than half a century—would impact health and thus longevity, we found no such effects. What explains this null finding? One possibility is that the long-time lags between survey waves minimized panel conditioning effects on health (and thus longevity); the WLS "treatments" occurred in 1975, 1993, 2004, and 2011. A second possibility is that the treatment "dosages"—an hour or two of survey questions—were too small in comparison to larger structural forces that shape mortality. Another possibility is that there are multiple small effects running in opposite directions that cancel out; for example, the treatment may enhance cardiovascular health but reduce marital satisfaction, potentially offsetting their effects on mortality. Perhaps these possible explanations are all true to some degree. Yet another possibility is that panel conditioning affects people's *reports* of their health-relevant attributes and behaviors more so than their *actual* health-relevant attributes and behaviors. The latter would be more consequential for mortality. A final possibility is that our impression of the literature—which has often documented panel conditioning effects on health behaviors and health outcomes—is misleading because of publication biases against null findings; perhaps disproportionately many articles showing no effects of panel conditioning on such outcomes have never been published.

One key limitation of our work is that the treatment—four intensive health-relevant interviews spread across more than three decades—does not closely resemble the treatment implicitly administered in surveys that interview people more frequently (e.g., the Health and Retirement Study or the Panel Study of Income Dynamics). A second limitation is that we cannot directly observe whether panel conditioning affects health behaviors and health outcomes in our data, because health behaviors and health outcomes (other than mortality) are unobserved for members of the control group. A third limitation is that compared to other cohort studies, the WLS is limited with respect to generalizability. The data we use include only people who were high school seniors in Wisconsin in the late 1950s. Because of geographic and educational selectivity, these individuals are overwhelmingly non-Latinx white (only a few dozen of the respondents sampled into the longitudinal panel were African American and only about a hundred were Native American). Although this is clearly sub-optimal, it is worth noting that nearly two thirds of all Americans from the cohort in question were non-Latinx white high school graduates (Herd et al., 2014). While we would certainly prefer greater racial/ethnic and educational variability, we believe that the virtues of this data source—including the ability to identify instances of panel conditioning using a fully experimental

design, the intensity of the treatment for members of the longitudinal panel, and the relatively large number of observations at our disposal—more than outweigh its downsides.

Author credit statement

John Robert Warren: Conceptualization, Funding acquisition, Formal analysis, Writing - original draft. **Andrew Halpern-Manners**: Conceptualization, Funding acquisition, Methodology, Writing - review & editing. **Jonas Helgertz**: Methodology, Data curation, Writing - review & editing.

Declaration of competing interest

The authors have no financial or personal relationships with other people or organizations that could inappropriately influence or bias their work.

Acknowledgements

Generous support for this project has been provided by the National Science Foundation (SES 1950469). Support has also come from the Minnesota Population Center, University of Minnesota, which receives core funding (P2C HD041023) from the Eunice Kennedy Shriver National Institute for Child Health and Human Development (NICHD). We are also very grateful to Carol Roan, Joseph Savard, and Jason Fletcher for providing helpful assistance. However, errors and omissions are the responsibility of the authors. Please direct correspondence to John Robert Warren at warre046@umn.edu.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ssmph.2022.101233.

References

Axinn, W. G., Jennings, E. A., & Couper, M. P. (2015). Response of sensitive behaviors to frequent measurement. *Social Science Research, 49*, 1–15. https://doi.org/10.1016/j.ssresearch.2014.07.002, 0.

Bach, R. L., & Eckman, S. (2019). *Participating in a panel survey changes respondents' labour market behaviour* (Vol. 182, pp. 263–281). https://doi.org/10.1111/rssa.12367 (1).

Bailey, M., Lin, P. Z., Shariq Mohammed, A. R., Mohnen, P., Murray, J., Zhang, M., et al. (2022). In *The creation of LIFE-M: The longitudinal, intergenerational family electronic micro-database project*. Los Angeles, CA: California Center for Population Research at UCLA.

Bartels, L. M. (1999). Panel effects in the American national election studies. *Political Analysis, 8*, 1–20.

Battaglia, M. P., Zell, E., & Ching, P. (1996). Can Participating in a panel sample introduce Bias into trend estimates?. In *National immunization survey working paper*. Washington, D.C: National Cenrter for Health Statistics. Available at http://www.cdc.gov/nis/pdfs/estimation_weighting/battaglia1996b.pdf. Retrieved from http://www.cdc.gov/nis/pdfs/estimation_weighting/battaglia1996b.pdf.

Baumeister, R., Vohs, K. D., DeWall, N., & Zhang, L. (2007). How emotion shapes behavior: Feedback, anticipation, and reflection, rather than direct causation. *Personality and Social Psychology Review, 11*(2), 167–203.

Cantor, D. (2008). A review and summary of studies on panel conditioning. In S. Menard (Ed.), *Handbook of longitudinal research: Design, measurement, and analysis* (pp. 123–138). Burlington, MA: Academic Press.

Chicco, D. J. B. M. (2017). *Ten quick tips for machine learning in computational biology* (Vol. 10, p. 35). https://doi.org/10.1186/s13040-017-0155-3 (1).

Clausen, A. R. (1968). Response validity: Vote report. *Public Opinion Quarterly, 32*, 588–606.

Crossley, T. F., de Bresser, J., Delaney, L., & Winter, J. (2017). *Can survey participation alter household saving behaviour?* (Vol. 127, pp. 2332–2357). https://doi.org/10.1111/ecoj.12398 (606).

Cutler, D. M., & Lleras-Muney, A. (2010). Understanding differences in health behaviors by education. *Journal of Health Economics, 29*(1), 1–28. https://doi.org/10.1016/j.jhealeco.2009.10.003

Das, M., Toepoel, V., & Soest, A.v. (2007). Can I use a panel?. In *Panel conditioning and attrition bias in panel surveys. CentER discussion paper series No. 2007-56*. Tilburg, Netherlands: Tilburg University CentER.

Duan, N., Alegria, M., Canino, G., McGuire, T. G., & Takeuchi, D. (2007). Survey conditioning in self-reported mental health service use: Randomized comparison of alternative instrument formats. *Health Services Research, 42*(2), 890–907. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1955350/.

Eagle, D. E., & Proeschold-Bell, R. J. (2015). Methodological considerations in the use of name generators and interpreters. *Social Networks, 40*, 75–83. https://doi.org/10.1016/j.socnet.2014.07.005

Feigenbaum, J. (2016). *Automated census record linking: A machine learning approach*. Boston: Boston University.

Fu, Z., Boot, H., Christen, P., & Zhou, J. (2014). Automatic record linkage of individuals and households in historical census data. *International Journal of Humanities and Arts Computing, 8*(2), 204–225.

Goeken, R., Huynh, L., Lynch, T. A., & Vick, R. (2011) (1). *New methods of census record linking historical methods* (Vol. 44, pp. 7–14). https://doi.org/10.1080/01615440.2010.517152. Retrieved from.

Greenland, S., Lanes, S., & Jara, M. (2008). Estimating effects from randomized trials with discontinuations: The need for intent-to-treat design and G-estimation. *Clinical Trials, 5*(1), 5–13. https://doi.org/10.1177/1740774507087703

Halpern-Manners, A., & Warren, J. R. (2012). Panel conditioning in the current population survey: Implications for labor force statistics. *Demography, 49*(4), 1499–1519.

Halpern-Manners, A., Warren, J. R., & Torche, F. (2014). Panel conditioning in a longitudinal study of illicit behaviors. *Public Opinion Quarterly, 78*(3), 565–590. https://doi.org/10.1093/poq/nfu029

Helgertz, J., Price, J., Wellington, J., Thompson, K. J., Ruggles, S., & Fitch, C. A. (2022). A new strategy for linking US historical censuses: A case study for the IPUMS multigenerational longitudinal panel. *Historical Methods: A Journal of Quantitative and Interdisciplinary History, 55*(1), 12–29.

Herd, P., Carr, D., & Roan, C. (2014). Cohort profile: Wisconsin longitudinal study (WLS). *International Journal of Epidemiology, 43*(1), 34–41. https://doi.org/10.1093/ije/Dys194

Hollywood, A., Ogden, J., & Hashemi, M. (2015). A randomised control trial assessing the impact of an investment based intervention on weight-loss, beliefs and behavior after bariatric surgery: Study protocol. *BMC Obesity, 2*, 1–5.

Kom, E. L., Graubard, B. I., & Midthune, D. (1997). Time-to-Event analysis of longitudinal follow-up of a survey: Choice of the time-scale. *American Journal of Epidemiology, 145*(1), 72–80. Retrieved from http://aje.oxfordjournals.org/content/145/1/72.abstract.

Krueger, A. B., Mas, A., & Niu, X. (2017). The evolution of rotation group bias: Will the real unemployment rate. *Please Stand Up?, 99*(2), 258–264. https://doi.org/10.1162/REST_a_00630

Little, R. J., & Rubin, D. B. (2000). Causal effects in clinical and epidemiological studies via potential outcomes: Concepts and analytical approaches. *Annual Review of Public Health, 21*(1), 121–145. https://doi.org/10.1146/annurev.publhealth.21.1.121

Oh, J. H., Yeatman, S., & Trinitapoli, J. (2019). *Data collection as disruption: Insights from a longitudinal study of young adulthood* (Vol. 84, pp. 634–663). https://doi.org/10.1177/0003122419859574 (4).

Rogers, R. G., Hummer, R. A., & Nam, C. B. (1999). *Living and dying in the USA: Behavioral, health, and social differentials of adult mortality*. San Diego: Academic Press.

Ruggles, S. J. D. (2014). *Big microdata for population research* (Vol. 51, pp. 287–297). https://doi.org/10.1007/s13524-013-0240-2 (1).

Solon, G. (1986). Effects of rotation group bias on estimation of unemployment. *Journal of Business & Economic Statistics, 4*(1), 105–109. Retrieved from <Go to ISI>://A1986AXF5400015.

Struminskaya, B. (2020). *Panel conditioning*. SAGE Publications Limited.

Struminskaya, B., & Bosnjak, M. (2021). Panel conditioning: Types, causes, and empirical evidence of what we know so far. *Advances in Longitudinal Survey Methodology*, 272–301.

Sturgis, P., Allum, N., & Brunton-Smith, I. (2009). Attitudes over time: The psychology of panel conditioning. In P. Lynn (Ed.), *Methodology of longitudinal surveys* (pp. 113–126). New York: Wiley.

Torche, F., Warren, J. R., Halpern-Manners, A., & Valenzuela, E. (2012). Panel conditioning in a longitudinal study of Chilean adolescents' substance use: Evidence from an experiment. *Social Forces, 90*(3), 891–918.

Tourangeau, R. (2000). Remembering what happened: Memory errors and survey reports. In A. Stone, J. Turkkan, C. Bachrach, J. Jobe, H. Kurtzman, & V. Cain (Eds.), *The science of self-report: Implications for research and practice* (pp. 29–48). Englewood Cliffs: Lawrence Erlbaum.

Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response*. New York: Cambridge University Press.

Veroff, J., Hatchett, S., & Douvan, E. (1992). Consequences of participating in a longitudinal-study of marriage. *Public Opinion Quarterly, 56*(3), 315–327. Retrieved from <Go to ISI>://A1992JT20200004.

Warren, J. R., & Halpern-Manners, A. (2012). Panel conditioning effects in longitudinal social science surveys. *Sociological Methods & Research, 41*(4), 491–534.

White, I. R., & Pocock, S. J. (1996). Statistical reporting of clinical trials with individual changes from allocated treatment. *Statistics in Medicine, 15*(3), 249–262. https://doi.org/10.1002/(SICI)1097-0258(19960215)15:3<249::AID-SIM160>3.0.CO;2-J. Retrieved from.

Williams, P., Block, L. G., & Fitzsimons, G. J. (2006). Simply asking questions about health behaviors increases both healthy and unhealthy behaviors. *Social Influence, 1*, 117–127. Retrieved from http://www.informaworld.com/smpp/content~content=a747971099~db=all~order=page.

Wilson, S., & Howell, B. L. (2007). Disease prevalence and survey design effects: A response to weir and smith. *Social Science & Medicine, 65*(6), 1078–1081. https://doi.org/10.1016/j.socscimed.2007.04.039

Winkler, W. E. (2004). Methods for evaluating and creating data quality. *Information Systems, 29*(7), 531–550. https://doi.org/10.1016/j.is.2003.12.003

Yeatman, S., Sennott, C., & Culpepper, S. J. D. (2013). *Young women's dynamic family size preferences in the context of transitioning fertility* (Vol. 50, pp. 1715–1737). https://doi.org/10.1007/s13524-013-0214-4 (5).

Zwane, A. P., Zinman, J., Van Dusen, E., Pariente, W., Null, C., Miguel, E., … Banerjee, A. (2011). Being surveyed can change later behavior and related parameter estimates. *Proceedings of the National Academy of Sciences, 108*(5), 1821–1826.