Sociological Spectrum
Mid-South Sociological Association

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/usls20

Effects of nonresponse and coverage problems on survey estimates of physical activity

Philip S. Brenner

To cite this article: Philip S. Brenner (2021) Effects of nonresponse and coverage problems on survey estimates of physical activity, Sociological Spectrum, 41:4, 338-351, DOI: 10.1080/02732173.2021.1926379

To link to this article: https://doi.org/10.1080/02732173.2021.1926379

Published online: 25 May 2021.

SOCIOLOGICAL SPECTRUM
2021, VOL. 41, NO. 4, 338–351
https://doi.org/10.1080/02732173.2021.1926379

Effects of nonresponse and coverage problems on survey estimates of physical activity

Philip S. Brenner

Department of Sociology and Center for Survey Research, University of Massachusetts Boston, Boston, USA

ABSTRACT
To what extent do survey nonresponse and coverage problems bias estimates of physical activity? Research has focused on social desirability as a cause of observed bias but there are relatively few studies of nonresponse and coverage. I analyze data from a survey designed to allow estimation of nonresponse bias using a two-phase sampling design that resamples nonrespondents from an initial wave of telephone and IVR (interactive voice response) interviewing for follow-up face-to-face interviews. Both initial and nonresponse interview waves included measures of physical activity frequency and duration. Estimates are compared between first-round respondents and those from follow-up interviews of nonrespondents, accounting for mode and other design elements. Telephone, but not IVR, interviews were found to include bias from two sources. Findings suggest that coverage is a cause of bias in the measure of frequency of physical activity but nonresponse may bias the measure of physical activity duration.

Introduction

Obesity is a serious public health problem in the United States, the United Kingdom, and elsewhere, increasing individuals' risk of chronic illnesses and premature mortality (Wang et al. 2011). While obesity is a complex phenomenon with multiple intersecting social and biological causes, physical inactivity is an important risk factor (Lee et al. 2012; De Rezende et al. 2016). Effective public health interventions address obesity by encouraging behavioral and lifestyle change, such as increasing physical activity (Wadden, Tronieri, and Butryn 2020). Assessing the effectiveness of public health interventions depends on good data which depend on surveys that validly and reliably measure the state of the population (Washburn, Heath, and Jackson 2000).

Yet, surveys commonly produce biased estimates that misrepresent the level of physical activity in the population. Research has primarily focused on measurement bias as a cause of inflated estimates (Sallis and Saelens 2000; Shephard 2003). Indeed, research has found that survey respondents over-report their behavior (Durante and Ainsworth 1996; Shephard 2003) relative to objective measures such as percentage of body fat and energy intake (Ainsworth, Jacobs, and Leon 1992; Albanes et al. 1990) and exercise performance (Adams et al. 2005; Leenders et al. 2001; Matthews et al. 2000; Matthews and Freedson 1995). Social science approaches confirm these findings, comparing survey reports to measures from flexible interviews (Rzewnicki et al. 2003), direct observation (Klesges et al. 1990), diaries (Brenner 2017), and reverse record checks (Brenner and DeLamater 2014, 2016; Harvey et al. 2018).

CONTACT Philip S. Brenner philip.brenner@umb.edu Department of Sociology and Center for Survey Research, University of Massachusetts Boston, 100 Morrissey Blvd., Boston, MA 02125, USA.

These studies clearly demonstrate that measurement error is an important problem for survey estimates of physical activity. Yet, the research literature has not as comprehensively investigated the problem that survey nonresponse may cause for estimates of physical activity (Jordan et al. 2011). Survey nonresponse arises when some of those people invited to participate fail to do so. While nonresponse may indeed be a source of bias, it is not *inherently* so (Groves et al. 2006). Nonresponse causes bias if individuals who can be contacted and agree to participate are different in key ways from those who cannot be contacted or who refuse to participate. Thus, if physically active individuals are more likely to participate than sedentary individuals, survey estimates may be positively biased. In most circumstances, however, we don't know why we fail to reach some individuals or why others do not participate. Therefore, we also typically do not know if or how respondents and nonrespondents differ, including on the frequency and duration of their physical activity (Bethlehem, Cobben, and Schouten 2011; Groves 2006; Groves and Couper 1998).

As this suggests, nonresponse bias is difficult to diagnose. Absent a criterion, the sample survey itself may be used to estimate the potential for nonresponse bias. A sample of nonrespondents is recontacted and measured using a more rigorous—and typically more expensive—data collection method. This approach, called a two-phase sampling design, compares estimates from the nonrespondent subsample to those from the original achieved sample to assess the potential for nonresponse bias in the original round of data collection.

Given the cost and difficulty of this procedure, there are relatively few rigorous investigations of the effect of nonresponse on survey measures of physical activity and their findings are mixed. Most studies of nonresponse and physical activity focus on panel attrition: estimating nonresponse bias as the difference between baseline and subsequent estimates in a panel survey of before-and-after treatment design. Of these, many studies find a lower propensity for nonresponse for physically active respondents (Boshuizen et al. 2006; Brownstein et al. n.d.; Chantala, Kalsbeek, and Andraca 2005; Gray et al. 1996; cf. Batty and Gale 2009; van Loon et al. 2003). While each of these studies operationalizes physical activity as playing sports (Caspersen, Powell, and Christenson 1985), additional analyses using time spent biking and walking find no association between physical activity and the propensity of nonresponse in later waves (Boshuizen et al. 2006). A similar approach using health clinic records as a baseline demonstrated that non-exercisers have a higher likelihood of nonresponse on subsequent surveys (Macera et al. 1990)

Findings from the few cross-sectional studies investigating nonresponse and physical activity estimates are also generally mixed similar to those using panel data. Mail survey nonrespondents followed up by telephone (Hill et al. 1997) and personal interview survey nonrespondents followed up by mail or phone (Shahar, Folsom, and Jackson 1996) reported lower levels of physical activity. However, a telephone survey comparing a standard 10-call contact protocol with an experimental contact protocol—increasing the number of types of contacts, including mailed letters, increased incentives, and refusal conversion—doubled its response rate from 25% to 50% but did not alter estimates of physical activity (Keeter et al. 2000, 2006).

Where nonresponse is suspected or expected to bias estimates, adjustment procedures have been used to correct for bias. Post-stratification weighting alters the distribution of sample characteristics to resemble what is known about the population, typically relying on demographic population parameters provided by a federal statistical agency. Multivariate models also typically employ similar sets of demographic characteristics as independent variables in a regression model to control for differential nonresponse that may be caused or related to these characteristics. Yet, these procedures to adjust for demographic and health differences between respondents and nonrespondents may fail to eliminate bias in physical activity estimates (Boshuizen et al. 2006; Gray et al. 1996; Hill et al. 1997; van Loon et al. 2003) or may even exacerbate bias (Harvey et al. 2018). Thus, nonresponse may bias estimates of physical activity and typical correctives such as weighting and modeling, may not be an effective solution.

The current study contributes to this literature by investigating the potential for nonresponse bias in survey estimates of physical activity in a general population. We address two limitations of the extant research. First, because the mechanism that generates attrition in panel and clinical designs may differ from that which generates nonresponse in cross-sectional designs, we conduct follow-up interviews of a sample of nonrespondents from an initial wave of interviewing. Nonresponse analyses of clinical and panel data are typically more interested in individual-level nonresponse as a potential confounder that may harm internal validity. Causal inferences, such as the effect of the efficacy of a treatment in a clinical design or net-change estimates using panel data, may be weakened if the mechanism causing attrition is related to the clinical treatment or concepts measured over time. Although important, the threat of nonresponse to internal validity at the individual-level is not the focus of this research. Probability-based sample surveys, like that used here, rely on representativeness to make inferences to a population. Differential nonresponse related to concepts being measured on a survey may harm representativeness, weakening external validity.

Second, because residents with listed telephone numbers may differ from those without (Blumberg and Luke 2007), we sample from an address-based frame that includes all residential addresses in the target population. We conduct telephone interviews with residents at addresses to which telephone numbers could be matched and personal interviews with residents at addresses without a matched telephone number. This design allows estimation of the contribution of coverage error attributable to systematic differences between households to which a phone number could be matched and those without a matched phone number. Follow up personal interviews are then conducted with telephone nonrespondents to allow estimation of potential nonresponse bias.

The design is structured to test two main hypotheses. First, in the face of survey nonresponse, we hypothesize that a difference will emerge between respondents who respond to the original request to participate and those who do not but are resampled and measured during follow-up interviews. Second, given imperfect coverage, we hypothesize that coverage bias will generate a difference between cases for which a telephone number is available and those cases for which a telephone number is not available.

Data and methods

Data collection started in September 2015 and lasted until April 2016. In the first phase of this two-phase sampling design, a proportionate-to-size stratified random sample of 3000 residential addresses was drawn from an address-based sampling frame covering five neighborhoods (Dorchester, Jamaica Plain, and Mattapan) and suburbs (Milton and Quincy), of Boston, Massachusetts, selected purposively based on their demographic diversity. The sample was randomly divided into two subsamples (see Figure 1). Addresses in the first subsample received a letter, accompanying a two-dollar cash incentive, inviting residents to call a toll-free number to complete an IVR interview (an interviewing method that uses a synthesized or prerecorded voice to read questions to respondents over the telephone which respondents answer by speaking or pressing a number on their phone's keypad). Invitation letters were printed in English on University of Massachusetts letterhead and addressed to the household by name with "or current resident" added to the addressee in case the name was out-of-date. These materials described the topic of the survey as focused on "health and our community" and advised residents that it would take approximate 10–15 min to complete. Invitation letters included wording that randomly selected either the youngest or oldest adult 18 years of age or older in the household. In this subsample, 148 respondents (AAPOR Response Rate [RR] #3 = 10%) completed the survey.

Addresses in the second subsample that were matched to a telephone number (approximately 60% of the sample) received a nearly identical letter informing them that an interviewer would

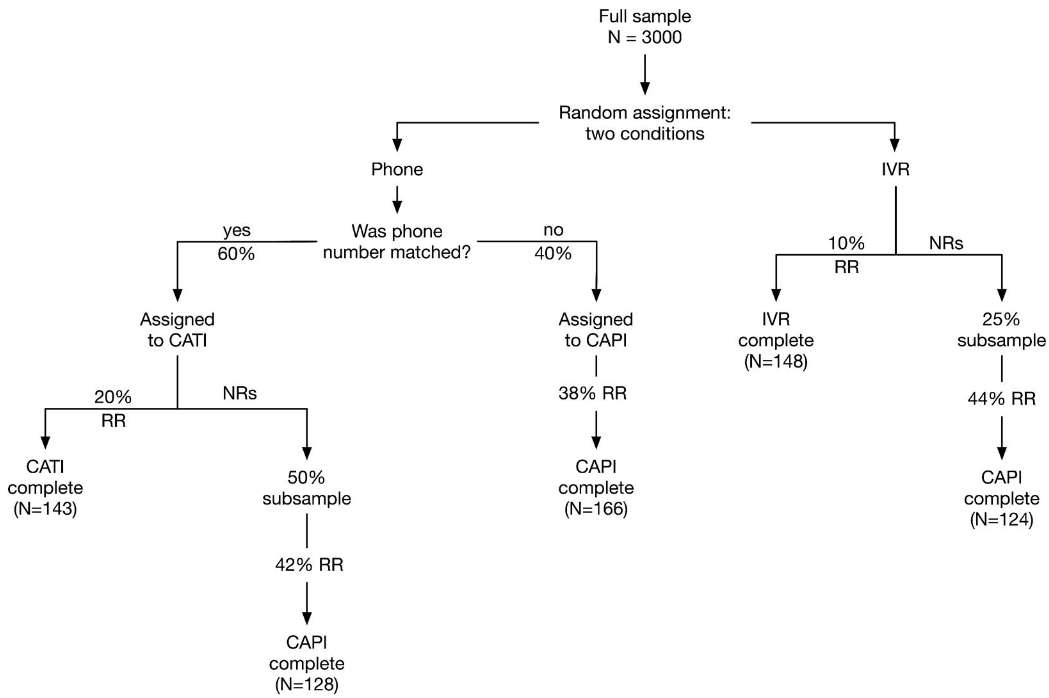

Figure 1. Study design, sample sizes, and response rates.

call in the next few days. Each telephone number was called a maximum of 12 times; the median number of calls was 6. Calls were placed on various days of the week, primarily during evenings and on weekends. All telephone interviews were conducted in English. In this part of the subsample, 143 respondents (20% RR) completed the survey. Addresses in the second subsample that were not matched to a telephone number received a nearly identical letter informing them that an interviewer would be visiting their home to complete a personal interview and promising twenty-dollar postpaid incentive. Each address was visited a maximum of 13 times; the median number of visits was 6. All personal interviews were completed in English. In this part of the subsample, 166 respondents (38% RR) completed the survey.

In the second phase of our two-phase sampling design, approximately a quarter of the nonresponding cases from the IVR subsample ($N = 335$) and half of the nonresponding cases from the telephone subsample ($N = 350$) were randomly selected for nonresponse follow-up interviews. They were sent a second letter informing them that an interviewer would be visiting their home to complete a personal interview and promising a twenty-dollar postpaid incentive. Each address was visited a maximum of 12 times; the median number of visits was 6. In these nonresponding segments, 124 and 128 respondents (44 and 42% RR) completed the survey.

Measures

All three data collection modes (IVR, telephone interviews, and personal interviews) used the same questionnaire, asking the same questions in the same order, including the two operationalizations of physical activity analyzed here. The first question measures *frequency* of physical activity in a typical week: "In a typical week, how many days do you do any physical activity or exercise of at least moderate intensity, such as brisk walking, bicycling at a regular pace, swimming at a regular pace, or heavy gardening?" Responses ranged from zero to seven days, and respondents reported a mean of 3.4 days of physical activity (SD = 2.4). This question serves as a filter question for the second which measures *duration* of physical activity episodes: "On the days that you do any physical activity or exercise of at least moderate intensity, how many minutes are you typically doing these activities?" Responses ranged from zero to 600 min with a mean of 47.5 min overall (SD = 57.3) and 56.9 min (SD = 58.3) for those respondents reporting at least one day of physical activity in the filter question. The variable analyzed here includes those reporting no days of physical activity in a typical week from the filter question, coded as zero minutes of physical activity. Both questions were originally asked on the Health Information National Trends Survey (HINTS), a nationally representative survey sponsored by the National Cancer Institute and are similar to those asked in other federally sponsored surveys.

The key independent variable is included as a set of five indicators for the three segments of the study within the telephone subsample ([1] telephone interviews; [2] personal interviews with residents from addresses to which a phone number could not be matched; and [3] personal interviews of phone nonrespondents) and the two segments within the IVR subsample ([4] IVR interviews and [5] personal interviews of IVR nonrespondents). Indicators for the five surveyed neighborhoods and cities are included as covariates along with a set of demographic variables: race/ethnicity in four categories (White; Black or African American; Asian and other races; and Latino/a of any race); educational attainment in four categories (less than a high school diploma; high school diploma or GED; some college completed; completed college degree or higher); age in years; employment status (employed/not employed); marital status (married/not married); and sex (male/female). Self-rated health, from excellent to poor, is also included as a covariate. Finally, an indicator of a matched phone number is used in analyses of the IVR subsample data to control for potential differences between addresses in this subsample with and without a matched phone number.

Weights

The proportionate-to-size stratified sampling design is self-weighting to the household level. A sampling design base weight was computed to adjust the propensity of respondent selection within households of varying size. Post-stratification weights adjust sample estimates to parameters from the American Community Survey on age (in three categories: 18–39, 40–64, 65 years and older), race, ethnicity, educational attainment, sex, and marital status for each of the five strata. These adjustments are raked to the margins, combined with base weights, and trimmed for extreme values. Weights are computed for each segment and for the full sample adjusting for unequal probabilities of selection across the two-phase sampling design.

Analysis

First, segments within the telephone subsample are compared. Estimates of physical activity from the telephone interviews are compared to two other estimates from this segment: personal interviews with respondents living at addresses to which phone numbers cannot be matched; and estimates from follow-up personal interviews with telephone nonrespondents. Second, estimates of physical activity from the IVR subsample are compared between respondents interviewed by IVR and nonrespondents who completed a follow-up personal interview. This comparison allows for the production of estimates for a realistic hypothetical: how much bias would have been incurred had nonrespondents in both subsamples not been interviewed or had residents from addresses lacking a matched telephone number in the telephone subsample not been pursued?

Measures of physical activity from initial modes of data collection are compared with those from the nonresponse interviews. Mean frequency and duration of physical activity are compared across study segments using base weights and post-stratification weights. Least squares regression

Table 1. Comparing demographic groups, between segments and the population.

	Phone respondents	Phone nonrespondents	No phone match	IVR respondents	IVR nonrespondents	Population
Race/ethnicity						
White	69.6%	55.6%	38.4%	67.4%	50.4%	42.5%
Black/African American	19.6%	25.8%	30.5%	15.6%	26.4%	27.9%
Asian/other race	9.4%	12.1%	16.5%	9.6%	15.2%	17.9%
Latino, any race	1.4%	6.5%	14.6%	7.4%	8.0%	11.7%
$X^2(p)$	45.2***	10.5*	2.4	34.4***	3.9	
Education						
Less than HS	5.0%	7.0%	12.7%	7.7%	8.8%	16.6%
High school	20.7%	25.8%	23.0%	8.5%	25.6%	26.6%
Some college	23.6%	17.2%	24.2%	15.5%	22.4%	22.4%
College or more	50.7%	50.0%	40.0%	68.3%	43.2%	35.0%
$X^2(p)$	23.1***	16.9***	3.7	72.3***	7.0	
Sex (Female)	60.8%	48.4%	51.8%	62.1%	56.0%	53.1%
$X^2(p)$	3.4	1.1	0.1	4.6*	0.4	
Marital status (Married)	42.9%	50.0%	32.9%	41.4%	36.0%	39.8%
$X^2(p)$	0.3	5.6*	3.7	0.2	0.8	
Age						
18–39 years	12.6%	24.2%	46.3%	32.6%	39.2%	45.5%
40–64 years	31.9%	46.9%	39.4%	45.7%	44.8%	39.7%
65 and older	55.6%	28.9%	14.4%	21.7%	16.0%	14.8%
$X^2(p)$	197.8***	34.5***	0.6	12.4**	4.0	
Employment (employed)	44.8%	63.3%	60.8%	65.5%	68.0%	70.4%
	15.8***	3.1	7.3**	0.2	0.4	
Self-reported health						
Excellent	14.7%	26.6%	24.1%	27.6%	16.8%	NA
N	143	128	166	148	124	

Note: ***$p < .001$, **$p < .01$, *$p < .05$.

models are then estimated, controlling for demographic and geographic independent variables to generate marginal means weighted using the base weights.

Results

First, demographic variables from five sample segments are compared to population parameters. Phone respondents, phone nonrespondents, and IVR respondents differ significantly from the population on most of the demographics (see Table 1). These segments overrepresent whites, those with a college degree, and older adults. Phone respondents are notably older than the population, with over half of phone respondents in the 65 and older age group, and are less likely to be employed. These differences raise the strong possibility that some segments, especially telephone respondents, may be very different from the population on the outcomes of interest: frequency and duration of physical activity (Lim et al. 2013). Notably, respondents without a matched phone number only differ from the population on employment status, being somewhat less likely to be employed, and IVR nonrespondents fail to significantly differ from the population on any of the five demographic variables.

Frequency of physical activity, telephone subsample

First, the mean number of physically-active days is compared between segments in the telephone subsample. Using base weights, personal interview respondents from addresses without a matched telephone number reported 4.1 days of physical activity on average, 0.8 more days of physical activity than telephone respondents and telephone nonrespondents interviewed in-person ($p \leq 0.05$; see Table 2). Post-stratification adjustment does not substantially alter these findings. Personal interview respondents from addresses without a matched telephone number reported

Table 2. Mean and marginal mean physically active days and minutes of physical activity, by segment and estimation type.

	Physical activity days in past week						Minutes spent in physical activity					
	Weighted for sampling design		Post-stratification adjustment		Prediction from multivariable model		Weighted for sampling design		Post-stratification adjustment		Prediction from multivariable model	
	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE
Phone subsample												
Respondents	3.3[a]	0.21	3.0[a]	0.28	3.3[a]	0.24	48.6[b]	3.3	44.2	4.5	50.2[b]	3.7
Nonrespondents	3.3[a]	0.22	3.2[a]	0.24	3.3[a]	0.21	39.2	3.1	39.5	3.5	39.0	2.9
No matched phone	4.1	0.21	3.9	0.25	4.1	0.22	49.0[b]	3.0	49.8[b]	4.1	47.4	3.1
Full subsample	3.5	0.14	3.4	0.15	3.5	0.13	43.7	1.9	44.0	2.4	43.4	1.8
IVR subsample												
Respondents	2.9	0.18	2.9	0.24	2.8[c]	0.23	37.9	3.0	38.2	4.1	38.2	3.9
Nonrespondents	3.4	0.24	3.2	0.37	3.4	0.22	42.0	3.1	39.6	4.8	42.5	2.8
Full subsample	3.2	0.18	3.1	0.29	3.2	0.17	40.7	2.4	39.3	3.9	41.3	2.3

Note: [a]differs from phone subsample respondents without matched phone number, $p \leq .05$; [b]differs from telephone nonrespondents, $p \leq .05$. [c]differs from IVR nonrespondents.

3.9 days of physical activity on average, 0.9 days more than telephone respondents ($p \leq 0.05$) and 0.7 days more than telephone nonrespondents interviewed in-person ($p \leq 0.05$). Findings from the multivariable model reiterate these findings. Personal interview respondents from addresses without a matched telephone number reported 4.1 days of physical activity on average, 0.8 days more than telephone respondents ($p \leq 0.05$) and nonrespondents interviewed in-person ($p \leq 0.01$).

Duration of physical activity, telephone subsample

Next, the mean amount of time respondents reported engaging in physical activity is compared between segments. Using base weights, telephone nonrespondents interviewed in-person reported an average of 39 min of physical activity, 10 min less than telephone respondents ($p \leq 0.05$) and 10 min less than personal interview respondents from addresses without a matched telephone number ($p \leq 0.05$). After post-stratification adjustment, findings are altered somewhat. Telephone nonrespondents interviewed in-person reported an average of 40 min of physical activity, 10 min less than personal interview respondents from addresses without a matched telephone number ($p \leq 0.05$) but nonrespondents' physical activity duration fails to differ from telephone respondents who report an average of 44 min of activity ($p = 0.42$). The multivariable model shows telephone nonrespondents interviewed in-person reported an average of 39 min of physical activity, 11 min less than telephone respondents ($p \leq 0.01$), and eight minutes less than nonrespondents, although the latter difference is just outside conventional significance levels ($p = 0.066$)

Frequency of physical activity, IVR subsample

Next, the number of reported physically-active days is compared between IVR respondents and nonrespondents interviewed in-person. In neither comparison using base weights (2.9 days for respondents, 3.4 days for nonrespondents; $p = 0.11$) or post-stratification weights (2.9 and 3.2 days; $p = 0.56$) do statistically significant differences emerge between IVR respondents and nonrespondents interviewed in-person. In the multivariable model, respondents report 2.8 physically-active days compared to 3.4 days for nonrespondents, a statistically significant 0.6-day difference ($p \leq 0.05$). Notably, the coefficient for the indicator denoting a matched phone number to the address was not significant in this model, suggesting that coverage would not have been a problem in this part of the design.

Duration of physical activity, IVR subsample

Finally, the number of reported minutes of physical activity between IVR respondents and nonrespondents interviewed in-person are compared. In none of the three comparisons, using base weights (38 and 42 min; $p = 0.35$), post-stratification weights (38 and 40 min; $p = 0.82$), nor the multivariable model (38 and 43 min; $p = 0.37$), do statistically significant differences emerge between IVR respondents and nonrespondents interviewed in-person. As in the previous model, the coefficient for the indicator denoting a matched phone number to the address was not significant, suggesting that coverage would not have been problematic in this part of the design.

Discussion

Self-reported physical activity was operationalized as frequency, the number of physically-active days, and duration, the time spent in physical activity on active days. If nonresponse is a cause of bias in either measure, consistent with Hypothesis 1, a significant difference should emerge between respondents in the initial survey (either telephone or IVR) and nonrespondents from the initial survey who were interviewed in the follow-up survey. If coverage is a cause of bias, consistent with Hypothesis 2, a significant difference should emerge between respondents from the telephone survey and those respondents, interviewed in-person who would have otherwise been omitted because there was no telephone number available for them.

Findings suggest that coverage, operationalized here as the difference between addresses that are able to be matched to a phone number and those not able to be matched, is a likely source of bias in the estimate of physical activity frequency. In the telephone subsample, personal interview respondents from addresses without a matched telephone number reported nearly one more physically-active day in the past week than either telephone respondents or nonrespondents interviewed in-person. Nonresponse, however, did not emerge as a likely source of bias in the frequency measure, as no significant difference emerged in the number of physically-active days between telephone respondents and telephone nonrespondents later interviewed in-person.

The coverage bias in these estimates may be at least partially related to age and other demographic differences. Addresses to which phone numbers can be matched are more likely to have older residents than addresses without matched telephone numbers. As older adults are less physically active, this difference may contribute to coverage bias. Non-covered addresses have younger, less well educated, and more racially and ethnically diverse residents. If these groups are also more likely to engage in frequent physical activity, this too may explain how coverage problems reduce the frequency measure. Importantly, however, this potential bias was not corrected by including post-stratification weights or demographic covariates in the estimation procedure. Significant differences in frequency of physical activity remained even when age and other demographic variables were used to post-stratify or to control in a multivariable model.

A different pattern of results emerged when predicting duration of physical activity. Telephone respondents and personal interview respondents from addresses without matched telephone numbers did not differ in their reported duration of their physical activity. Therefore, coverage bias—operationalized here as the difference between addresses with and without a matched phone number—was not a significant contributor to error in the estimate of time spent in physical activity. Nonresponse, however, does appear to be a source of bias in the estimate of physical activity duration. Nonrespondents from the telephone survey later interviewed in-person reported about 10 fewer minutes of activity than telephone respondents and personal interview respondents from addresses without a matched phone number. Importantly, this potential bias was not fully corrected with post-stratification weighting or modeling with demographic and health controls. Significant differences in the duration of physical activity remained.

While both hypotheses are supported for the telephone interview component of this study, it is not clear why coverage bias emerged only in the frequency measure and nonresponse bias emerged only in the duration measure. Demographic differences between segments may explain this pattern. Residents at addresses to which a phone number could not be matched tend to be younger, more racially and ethnically diverse, and have lower levels of education than phone respondents and nonrespondents. This group is plausibly more likely to engage in high frequency and high duration physical activities, such as a job doing physically intensive work or regularly playing a sport. In comparison, phone respondents and nonrespondents report a lower frequency of physical activity but differ from each other on their reported physical activity duration. Phone respondents report higher duration physical activity than nonrespondents and differ from the other groups in some key ways. Phone respondents have the highest proportion in the oldest (65 or older) age group and have the lowest rate of employment, likely due to a higher rate of retirement, are the least racially and ethnically diverse, and have the highest level of education. Given these characteristics, they may tend to engage in less frequent but longer duration leisure-type physical activities such as walking or playing golf. Nonrespondents tend to split the difference demographically between the other two groups but are more heavily male, married, and have a higher proportion in the working-age (40–64) group. They may be more likely to engage in lower duration activities, such as minor household chores. While this post-hoc reasoning is plausible and may explain the pattern of bias found here, it is speculative and requires confirmation in future research.

The second set of models predicted physical activity in the IVR subsample. In neither of the models predicting frequency or duration of physical activity do consistently significant differences emerge between respondents and nonrespondents followed-up with personal interviews. The lack of significant differences between respondents and nonrespondents suggests nonresponse is not significantly biasing these estimates. Moreover, no difference was found between cases with and without a matched phone number in the IVR subsample, suggesting that coverage bias would not have been a problem in this part of the design.

Response rates for this survey add a caveat to this finding. Initial response rates were somewhat low: 10% in the IVR survey and 20% in the telephone survey. While these response rates are not unusually low for surveys in diverse, urban populations like the Boston area, they are one of a number of important limitations of this study. A more important limitation than the response rate from the initial round of data collection is that from the nonresponse follow-up surveys. Only 42–44% of randomly selected nonrespondents chosen for a follow-up survey were interviewed. This unfortunately means that our best estimates may still include nonresponse bias. This limitation would be more notable if we had not found bias and were defending the unbiasedness of the survey estimates. However, that our nonrespondent samples look more similar demographically to the population than the initial telephone and IVR samples (see Table 1), suggests that nonresponse in the follow-up interviews is likely not generating additional bias.

Second, bias is highly dependent on the specific details of the particular study design, such as sampling design, frame characteristics, and data collection mode. Our study finds evidence for coverage and nonresponse biases using telephone and personal interviews with an address-based sample. While these findings may generalize to other production surveys using other designs, different profiles of nonresponse and coverage errors are possible in other studies. Future research should examine these errors for mail and web surveys.

Moreover, while this study was not designed to assess measurement error, the comparison between telephone and IVR respondents may yet give some insight into social desirability bias, a particular form of measurement error discussed in the introduction. Social desirability bias is commonly linked to interviewer-administered surveys. Respondents are believed to exaggerate normative behaviors, such as engaging in physical activity, when reporting on these behaviors to survey interviewers but not when reporting on self-administered surveys. We can partially assess the role of measurement error attributable to social desirability by comparing phone respondents to IVR respondents but limiting analysis to cases from addresses with a matched phone number to eliminate the potential for differential coverage between phone and IVR estimates. These comparisons (analysis not shown) yield no differences between phone interviews and IVR in the frequency measure when weighted with sampling weights or post-stratification weights, or controlling for demographics and health. A significant difference does emerge for activity duration: IVR respondents report about 12 min less activity than phone respondents when weighted using sampling weights and when controlling for demographic and health variables. This difference, however, is reduced in size (three minutes) and to non-significance when post-stratification weights are used.

Thus, while there is some evidence for measurement error here, there is also evidence that post-stratification adjustment helps to correct this bias. Note also that the difference between phone interviews and IVR may be due to some differences in these modes other than those associated with the social presence of the interviewer or lack thereof. For instance, the duration question may be easier to answer with an interviewer who can help respondents compute an answer in minutes (e.g, 3 1/2 h = 210 min) in contrast with IVR which makes higher cognitive demands and requires more arithmetic competence.

Third, while this study rigorously estimates potential nonresponse bias, the use of personal interviewing limited the research geographically to Boston. Given that the city has been repeatedly recognized as one of the country's "fittest" (American College of Sports Medicine 2017), the mechanism generating nonresponse may differ between Boston and other cities and regions in the US. Notably, the physical activity question was asked in the context of a survey that was explicitly focused on health and correlates of health and was described to the respondent as such. Thus, respondents may have reacted differentially in a way that was related to the physical activity questions.

Conclusion

Researchers have primarily been focused on the potential for social desirability bias and other measurement errors in measures of health behaviors, such as physical activity. However, nonresponse and coverage may also potentially bias estimates of these behaviors. The purpose of this study was to examine these forms of survey error and assess their potential to alter estimates of physical activity. Using a two-phase sampling design with telephone surveys of an address-based sample, evidence of both nonresponse and coverage biases was found.

Moreover, procedures typically used to adjust for these sources, post-stratification and multivariable modeling, failed to correct estimates from the telephone survey for these sources of bias. Thus, an key lesson from these findings is one that reiterates an important message from the survey methods literature. Survey errors, including nonresponse and coverage biases, are statistic-by-statistic concerns and may affect different measures differently, even for measures of the same underlying concept. Moreover, when and where they emerge, nonresponse and coverage biases cannot always be "controlled for" in a multivariable model or "weighted away" in survey measures of physical activity.

While the coverage problem found here is highly dependent on the particular study design used, an address-based sampling design with matched telephone numbers, the problem of nonresponse is likely translatable to other telephone surveys using other sampling designs, such as RDD, and perhaps other data collection methods as well. Indeed, the problem of nonresponse bias in measures of physical activity should be a concern for any telephone survey that asks questions about physical activity as well as other measures of health (Fowler et al. 2019, 2020).

Limitations notwithstanding, these findings suggest that survey estimates of physical activity may indeed include nonresponse or coverage biases, even if researchers use post-stratification weights or estimate multivariable models with demographic control variables.

Acknowledgements

The author would like to acknowledge Lee Hargraves, Jack Fowler, Carol Cosenza, Tony Roman, and the staff at the Center for Survey Research at the University of Massachusetts Boston.

Disclosure statement

The author has no financial or other conflicts of interest to report.

Funding

This research was supported by a grant from the National Science Foundation, SES-1424433.

Author note

Philip S. Brenner is Associate Professor of Sociology and Senior Research Fellow in the Center for Survey Research at the University of Massachusetts Boston, where he also serves as Director of the Graduate Certificate Program in Survey Research. His research examines social desirability bias and other errors in interviewer- and self-administered surveys.

ORCID

Philip S. Brenner http://orcid.org/0000-0002-8848-1114

References

Adams, Swann Arp, Charles E. Matthews, Cara B. Ebbeling, Charity G. Moore, Joan E. Cunningham, Jeanette Fulton, and James R. Hebert. 2005. "The Effect of Social Desirability and Social Approval on Self-Reports of Physical Activity." *American Journal of Epidemiology* 161:389–98. doi:10.1093/aje/kwi054.

Ainsworth, Barbara E., David R. Jacobs, and Arthur S. Leon. 1992. "Validity and Reliability of Self-Reported Physical Activity Status: The Lipid Research Clinics Questionnaire." *Medicine and Science in Sports and Exercise* 25:92–8.

Albanes, Demetrius, Joan Conway, Philip Taylor, Paul Moe, and Joseph Judd. 1990. "Validation and Comparison of Eight Physical Activity Questionnaires." *Epidemiology* 1:65–71.

American College of Sports Medicine. 2017. "*ACSM American Fitness Index: 2017 Health and Community Fitness Status of the 50 Largest Metropolitan Areas*." Retrieved November 7, 2019 (https://www.americanfitnessindex.org/wp-content/uploads/2014/02/ACSM-AFI-Report-2017.pdf).

Batty, G. David, and Catherine Gale. 2009. "Impact of Resurvey Non-Response on the Associations between Baseline Risk Factors and Cardiovascular Disease Mortality." *Journal of Epidemiology & Community Health* 63(11):952–5. doi:10.1136/jech.2008.086892.

Bethlehem, Jelke, Fannie Cobben, and Barry Schouten. 2011. *Handbook of Nonresponse in Household Surveys*. New York: Wiley.

Blumberg, Stephen J., and Julian V. Luke. 2007. "Coverage Bias in Traditional Surveys of Low-Income and Young Adults." *Public Opinion Quarterly* 71(5):734–49. doi:10.1093/poq/nfm047.

Boshuizen, H. C., A. L. Viet, H. S. J. Picavet, A. Botterweck, and A. J. M. van Loon. 2006. "Non-Response in a Survey of Cardiovascular Risk Factors in the Dutch Population: Determinants and Resulting Biases." *Public Health* 120:297–308. doi:10.1016/j.puhe.2005.09.008.

Brenner, Philip S. 2017. "Differential Effects of Time Constraints on Athletic Behavior and Survey Reports of Athletic Behavior." *Sociological Spectrum* 37 (2):97–110. doi:10.1080/02732173.2016.1274249.

Brenner, Philip S., and John DeLamater. 2014. "Social Desirability Bias in Self-Reports of Physical Activity: Is an Exercise Identity the Culprit?" *Social Indicators Research* 117 (2):489–504. doi:10.1007/s11205-013-0359-y.

Brenner, Philip S., and John DeLamater. 2016. "Lies, Damned Lies, and Survey Self-Reports? Identity as a Cause of Measurement Bias." *Social Psychology Quarterly* 79:333–54. doi:10.1177/0190272516628298.

Brownstein, Naomi, William D. Kalsbeek, Joyce Tabor, Pamela Entzel, Eric Daza, and Kathleen Mullan Harris. n.d. *Non-Response in Wave IV of the National Longitudinal of Adolescent Health*. Chapel Hill, NC: University of North Carolina.

Caspersen, Carl J., Kenneth E. Powell, and Gregory M. Christenson. 1985. "Physical Activity, Exercise, and Physical Fitness: Definitions and Distinctions for Health-Related Research." *Public Health Reports* 100:126–31.

Chantala, Kim, William D. Kalsbeek, and Eugenio Andraca. 2005. *Non-Response in Wave III of the Add Health Study*. Chapel Hill, NC: University of North Carolina.

de Rezende, Leandro Fórnias Machado, Thiago Hérick de Sá, Grégore Iven Mielke, Juliana Yukari Kodaira Viscondi, Juan Pablo Rey-López, and Leandro Martin Totaro Garcia. 2016. "All-Cause Mortality Attributable to Sitting Time: Analysis of 54 Countries Worldwide." *American Journal of Preventive Medicine* 51:253–63. doi:10.1016/j.amepre.2016.01.022.

Durante, Richard, and Barbara E. Ainsworth. 1996. "The Recall of Physical Activity: Using a Cognitive Model of the Question-Answering Process." *Medicine & Science in Sports and Exercise* 28:1282–91.

Fowler, Floyd J., Philip S. Brenner, Anthony M. Roman, and J. Lee Hargraves. 2020. "The Effects of Nonresponse and Sampling Omissions on Estimates on Various Topics in Federal Surveys: Telephone and IVR Surveys of Address-Based Samples." *Journal of Official Statistics* 36(3):631–45. doi:10.2478/jos-2020-0032.

Fowler, Floyd J., Philip S. Brenner, Trent D. Buskirk, and Anthony Roman. 2019. "Telephone Health Survey Estimates: Effects of Nonresponse and Sample Limitations." *Health Services Research* 54:700–6. doi:10.1111/1475-6773.13110.

Gray, Rebecca, Pamela Campanelli, Kavita Deepchand, and Patricia Prescott-Clarke. 1996. “Exploring Survey Non-Response: The Effect of Attrition on a Follow-Up of the 1984–85 Health and Life Style Survey.” *The Statistician* 45 (2):163–83. doi:10.2307/2988406.

Groves, Robert M. 2006. “Nonresponse Rates and Nonresponse Bias in Household Surveys.” *Public Opinion Quarterly* 70(5):646–75. doi:10.1093/poq/nfl033.

Groves, Robert M., and Mick P. Couper. 1998. *Nonresponse in Household Interview Surveys*. New York: Wiley.

Groves, Robert M., Mick P. Couper, Stanley Presser, Eleanor Singer, Roger Tourangeau, Giorgina Piani Acosta, and Lindsay Nelson. 2006. “Experiments in Producing Nonresponse Bias.” *Public Opinion Quarterly* 70(5): 720–36. doi:10.1093/poq/nfl036.

Harvey, J. T., M. J. Charity, N. A. Sawyer, and R. M. Eime. 2018. “Non-Response Bias in Estimates of Prevalence of Club-Based Sport Participation from an Australian National Physical Activity, Recreation, and Sport Survey.” *BMC Public Health* 18:895. doi:10.1186/s12889-018-5793-y.

Hill, Anthony, Julian Roberts, Paul Ewings, and David Gunnell. 1997. “Non-Response Bias in a Lifestyle Survey.” *Journal of Public Health Medicine* 19 (2):203–7. doi:10.1093/oxfordjournals.pubmed.a024610.

Jordan, Jeremy S., Matthew Walker, Aubrey Kent, and Yuhei Inoue. 2011. “The Frequency of Nonresponse Analysis in the *Journal of Sport Management*.” *Journal of Sport Management* 25 (3):229–39. doi:10.1123/jsm.25.3.229.

Keeter, Scott, Carolyn Miller, Andrew Kohut, Robert Groves, and Stanley Presser. 2000. “Consequences of Reducing Nonresponse in a National Telephone Survey.” *Public Opinion Quarterly* 64:125–48. doi:10.1086/317759.

Keeter, Scott, Courtney Kennedy, Michael Dimock, Jonathan Best, and Peyton Craighill. 2006. “Gauging the Impact of Growing Nonresponse on Estimates from a National RDD Telephone Survey.” *Public Opinion Quarterly* 70 (5):759–79. doi:10.1093/poq/nfl035.

Klesges, Robert C., Linda, H. Eck, Michael W. Mellon, William Fulliton, Grant W. Somes, and Cindy L. Hanson. 1990. “The Accuracy of Self-Reports of Physical Activity.” *Medicine and Science in Sports and Exercise* 22:690–7.

Lee, I-Min, Eric J. Shiroma, Felipe Lobelo, Pekka Puska, Steven N. Blair, and Peter T. Katzmarzyk. 2012. “Effect of Physical Inactivity on Major Non-Communicable Diseases Worldwide: An Analysis of Burden of Disease and Life Expectancy.” *The Lancet* 380(9838):219–29. doi:10.1016/S0140-6736(12)61031-9.

Leenders, Nicole Y. J. M., W. Michael Sherman, H. N. Nagaraja, and C. Lawrence Kien. 2001. “Evaluation of Methods to Access Physical Activity in Free-Living Conditions.” *Medicine and Science in Sports and Exercise* 33: 1233–40.

Lim, Sungwoo, Stephen Immerwahr, Sunghee Lee, and Tiffany Harris. 2013. “Estimating Nonresponse Bias in a Telephone-Based Health Surveillance Survey in New York City.” *American Journal of Epidemiology* 178(8): 1337–41. doi:10.1093/aje/kwt121.

Macera, Caroline A., Kirby L. Jackson, Dorothy R. Davis, Jennie J. Kronenfeld, and Steven N. Blair. 1990. “Patterns of Non-Response to a Mail Survey.” *Journal of Clinical Epidemiology* 43(12):1427–30. doi:10.1016/0895-4356(90)90112-3.

Matthews, Charles E., and Patty S. Freedson. 1995. “Field Trial of a Three-Dimensional Activity Monitor: Comparison with Self-Report.” *Medicine and Science in Sports and Exercise* 27:1071–8.

Matthews, Charles E., Patty S. Freedson, James R. Hebert, Edward J. Stanek, III, Philip A. Merriam, and Ira S. Ockene. 2000. “Comparing Physical Activity Assessment Methods in the Seasonal Variation of Blood Cholesterol Study.” *Medicine and Science in Sports and Exercise* 32:976–84.

Rzewnicki, Randy, Yves Vanden Auweele, and Ilse De Bourdeaudhuij. 2003. “Addressing Overreporting on the International Physical Activity Questionnaire (IPTQ) Telephone Survey with a Population Sample.” *Public Health Nutrition* 6:299–305. doi:10.1079/PHN2002427.

Sallis, James, and Brian Saelens. 2000. “Assessment of Physical Activity by Self-Reports: Status, Limitations, and Future Directions.” *Research Quarterly for Exercise and Sport* 71(sup2):1–14. doi:10.1080/02701367.2000.11082780.

Shahar, Eyal, Aaron R. Folsom, and Rodney Jackson. 1996. “The Effect of Nonresponse on Prevalence Estimates for a Referent Population: Insights from a Population-Based Cohort Study.” *Annals of Epidemiology* 6(6): 498–506. doi:10.1016/S1047-2797(96)00104-4.

Shephard, R. J. 2003. “Limits to the Measurement of Habitual Physical Activity by Questionnaires.” *British Journal of Sports Medicine* 37:197–206. doi:10.1136/bjsm.37.3.197.

van Loon, A. Jeanne M., Marja Tijhuis, H. Susan J. Picavet, Paul G. Surtees, and Johan Ormel. 2003. “Survey Non-Response in The Netherlands: Effects on Prevalence Estimates and Associations.” *Annals of Epidemiology* 13(2): 105–10. doi:10.1016/s1047-2797(02)00257-0.

Wadden, Thomas A., Jena S. Tronieri, and Meghan L. Butryn. 2020. “Lifestyle Modification Approaches for the Treatment of Obesity in Adults.” *American Psychologist* 75 (2):235–51. doi:10.1037/amp0000517.

Wang, Y. Claire, Klim McPherson, Tim Marsh, Steven L. Gortmaker, and Martin Brown. 2011. "Health and Economic Burden of the Projected Obesity Trends in the USA and UK." *Lancet* 378:815–25. doi:10.1016/S0140-6736(11)60814-3.

Washburn, Richard, Gregory W. Heath, and Allen W. Jackson. 2000. "Reliability and Validity Issues concerning Large-Scale Surveillance of Physical Activity." *Research Quarterly for Exercise and Sport* 71 (sup2):104–13. doi:10.1080/02701367.2000.11082793.