Sociology of Religion: A Quarterly Review 2019, 80:2 247–262
doi: 10.1093/socrel/sry031
Advance Access Publication 24 August 2018

Does Survey Nonresponse Bias Estimates of Religious Service Attendance? Evidence from an Address-Based Sample from the Boston Area

Philip S. Brenner*
University of Massachusetts Boston

This study investigates what role, if any, nonresponse plays in inflating survey estimates of religious behavior, using a multimode survey designed to allow estimation of nonresponse bias. A sample of 3,000 Boston-area households drawn from an address-based frame was randomly divided into two subsamples, contacted by mail, and invited to participate in a survey. The first subsample was asked to complete an interactive voice response interview. The second subsample was asked to complete a survey by telephone if a number was available for the address or by personal interview if not. Finally, random samples of nonrespondents were recontacted for a personal interview. Comparison of attendance estimates from initial interviews with nonrespondent interviews within sample segments yields minor or minimal differences that are not statistically significant. Findings suggest that the mechanism generating survey nonresponse is unlikely to be a major cause of bias in religious service attendance estimates in this study.

Key words: attendance; survey research; measurement; quantitative methods.

INTRODUCTION

How many Americans attend religious services in a given week? This ostensibly simple question turns out to be difficult to answer with a reasonable amount of certainty. For religious statistics like this one to be valid and reliable, a sample of individuals who resemble the population must be randomly selected, agree to participate, understand the questions being asked, and answer honestly. A well-developed literature on survey estimates of religious service attendance has focused on the last two steps of this sequence. Starting with the last step (answering

Direct correspondence to Philip S. Brenner, University of Massachusetts Boston, 100 Morrissey Boulevard, Boston, MA 02125, USA. E-mail: philip.brenner@umb.edu.

honestly), evidence suggests that many respondents do not provide answers that reflect their "true value" of attendance. Rather, they report values that exceed their actual behavior and inflate the survey estimate by about 100% over its true value (Brenner 2011a; Chaves and Cavendish 1994; Hadaway et al. 1993, 1998; Presser and Stinson 1998). This research has led to a deeper understanding of the cause of this form of measurement bias focused on the penultimate step (understanding the question). Bias appears to be founded in respondents applying a pragmatic interpretation of the question—about one's identity as a church-going person—rather than its intended semantic meaning about actual behavior (Brenner 2017; Hadaway et al. 1998).

Yet, measurement error may not be the only form of error contributing to bias in survey estimates of religious service attendance. In the face of declining response rates, survey nonresponse has become one of the most important and researched problems in survey methodology (Brick and Williams 2013; Groves and Couper 1998). However, diagnosing nonresponse and its contribution to bias can be difficult because we typically know little or nothing about survey nonrespondents. Accordingly, few rigorous studies of the effect of survey nonresponse on estimates of church attendance have been conducted. Therefore, we know relatively little about how survey nonresponse contributes to bias in survey estimates of attendance, if it biases them at all.

Religious service attendance is an important survey measure to assess given its ubiquity in statistical models across the social sciences, used to operationalize religious behavior specifically or religiosity in general, either as a dependent or an independent variable (Brenner 2016). Thus, this study assesses survey error and one important source of it, survey nonresponse, in a measure of religious service attendance. First, the few studies presenting estimates of nonresponse bias (or a lack thereof) on religious service attendance measures are reviewed and synthesized. Then, I describe the design of the current study that marries a typical production study protocol with rigorous nonresponse follow-up to identify the potential effect of nonresponse bias on survey estimates of religious service attendance.

Previous Research

One of the most vexing problems in survey research today is survey nonresponse. However, the problem with survey nonresponse is not directly that we fail to reach some sample elements or that some decide not to participate. As long as our ability to contact sample elements and their decision to participate are unrelated to the concept(s) being measured, nonresponse is ignorable, and our main concern is the reduction in sample size that would reduce the statistical power of our estimates and coefficients. Nonresponse only causes bias if the mechanism generating it is related to the concepts being measured. Moreover, nonresponse bias, where it occurs, is a statistic-by-statistic phenomenon. Some estimates from a survey may include bias while others may be valid (Groves 2006; Groves et al. 2006).

Survey estimates of religious service attendance are not immune from the potential of nonresponse bias as even high-quality surveys with high response rates produce estimates that inflate rates of religious service attendance (Brenner 2011b). If the mechanism generating nonresponse is positively or negatively related to the sampled individual's religiosity, bias could result. A positive relationship could be generated by a set of nonexclusive causes. Religious individuals may be more likely than irreligious individuals to participate in a survey if they have a higher propensity for compliant, cooperative, or pro-social behavior (Abraham et al. 2009; Woodberry 1998), if they are more involved in or integrated into their communities (Abraham et al. 2006), or if they are easier to contact given their higher propensity to have family types or household compositions (e.g., older adults, young families with children) with more accessible at-home patterns (Groves and Couper 1998). A negative relationship could result if religious individuals, given their propensity to be politically conservative (Malka et al. 2012), share conservatives' distrust of the social institutions that collect survey data such as universities (Gauchat 2012; Johnson and Peifer 2017; Pew 2017) and federal agencies such as the Bureau of Labor Statistics and Census Bureau (Weakliem and Villemez 2004). While either a positive or negative relationship is a potential and plausible source of bias, we know relatively little about the effect of nonresponse on survey estimates of religious service attendance.

Woodberry (1998) suggests that survey nonresponse is a primary cause of inflated estimates of religious service attendance in the United States. He argues that religious respondents are more easily contacted and more cooperative than irreligious respondents. Comparing estimates from the General Social Survey (GSS) and Southern Focus Poll (SFP), Woodberry finds a 15-point difference in attendance propensities. He attributes the lower attendance propensity in the GSS to its higher response rate (which yields a more representative sample) and the higher attendance propensity in the SFP to its lower response rate (which yields a less representative sample with fewer nonattenders). However, GSS and SFP differ on multiple dimensions—including sampling frames (area probability vs. random-digit dialing [RDD]) and data collection mode (personal vs. telephone interviewing)—that could cause the observed variation between attendance estimates.

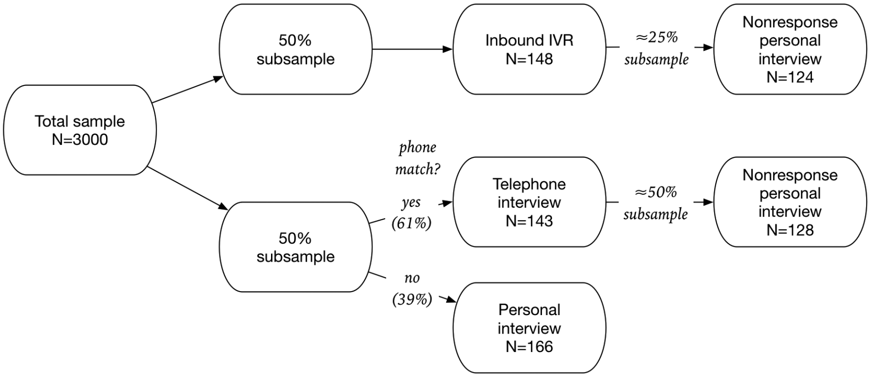
Moreover, recent work fails to support Woodberry's finding. Survey methodologists at the Pew Research Center designed a study to assess the contribution of nonresponse to total survey error. They compared a phone survey of an RDD sample that attempted at least 10 contacts with a design that increased the quantity and quality of contact attempts (Keeter et al. 2000, 2006). Although additional effort expended to complete cases doubled their response rate (50%, compared with 25% for the standard design), it did not significantly change their religious service attendance estimate: 36% and 37% of respondents in the extended and standard designs, respectively, reported weekly attendance (2006).

Given the limited extant research and the mixed nature of these findings, this article reports on the findings of a study designed to test the effect of nonresponse on survey estimates of religious service attendance that builds on this prior work, extending and updating it in two important ways. First, as much of the survey research industry moves away from RDD sampling designs (like that used by Keeter et al.) given the shift in the population away from landline telephones (Blumberg and Luke 2017), this study uses an address-based sampling (ABS) frame based on the US Postal Service Delivery File. Telephone numbers were matched to these addresses and interviewers attempted to complete phone interviews, following up nonrespondents with personal interviews. Second, a key limitation of the prior research—problem with coverage of households without telephone numbers—was addressed by sending field interviewers to addresses without matched phone numbers. In a second subsample, self-administered telephone interviewing, called interactive voice response (IVR), was used and nonrespondents were followed up with personal interviews. Initial respondents are compared with a sample of nonrespondents who were subsequently interviewed in-person (and, in the telephone interview subsample, with respondents from households without a matched telephone number) to assess the potential for nonresponse bias at each stage of data collection within each subsample.

DATA AND METHODS

Data collection started in September 2015 and lasted until April 2016. A proportionate-to-size stratified random sample of 3,000 residential addresses was drawn from an ABS frame covering five Boston neighborhoods (Dorchester, Jamaica Plain, and Mattapan) and suburbs (Milton and Quincy), selected purposively based on their demographic diversity. The sample was randomly divided into two subsamples (figure 1). Addresses in the first subsample received a letter, accompanying a 2-dollar cash incentive, inviting residents to call a toll-free number to complete an IVR survey. Invitation letters were printed in English on University of Massachusetts Boston letterhead and addressed to the household by

FIGURE 1. Study design.



name (e.g., "The Smith Household") with "or current resident" added to the addressee in case the matched name was out-of-date or otherwise incorrect. 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 minutes 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 #3 = 10%) completed the survey.

Addresses in the second subsample that were matched to a telephone number (~60% of the sample) received a nearly identical letter informing them that an interviewer would call in the next few days. Each telephone number was called a maximum of 12 times; the median and modal 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% response rate) 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 a 20-dollar incentive. Each address was visited a maximum of 13 times; the median and modal number of visits was 6. All personal interviews were completed in English. In this part of the subsample, 166 respondents (41% response rate) completed the survey.

The previous paragraphs describe the first phase of the two-phase sampling design used in this study. In the second phase, nonresponding cases are sampled. 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 20-dollar incentive. Each address was visited a maximum of 12 times; the median and modal number of visits was 6. In these nonresponding segments, 124 and 128 respondents (47% and 43% response rates, respectively) completed the survey.

Measures

The dependent variable is self-reported religious service attendance, measured using the question, "How often do you attend religious services? Would you say never, about once or twice a year, several times a year, about once a month, two to three times a month, or every week or more often?" This variable is analyzed both in its original ordinal distribution and, in line with prior research (Presser and Stinson 1998; Woodberry 1998), recoded to reflect a respondent's stated propensity to attend in a given week: never (0.000); once or twice a year (1.5 ÷ 52 = 0.029); several times a year (6 ÷ 52 = 0.115); once a month (1 ÷ 4 = 0.250); two to three times a month (2.5 ÷ 4 = 0.625); and every week or more often = (1.000).

The key independent variable is a set of indicators for the three segments of the study within the telephone subsample (telephone interviews; personal interviews of phone nonrespondents; and personal interviews with residents at addresses to which a phone number could not be matched) and the two segments within the IVR subsample (IVR interviews and personal interviews of IVR nonrespondents). Indicators for the five surveyed neighborhoods and suburbs of Boston are included as covariates along with five 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 (18–39, 40–64, 65 years and older); marital status (married/not married); and sex (male/female).

The achieved sample in its five segments differs notably from the population on these five covariates. In four of the five segments of the study (excepting households to which a phone number could not be matched), whites are overrepresented compared with the population (table 1). Those with higher educational attainment are also overrepresented compared with the population, especially among IVR respondents. But the most striking difference between sample segments and the population is on respondent age. On the phone survey, 56% reported being 65 years of age or older compared with their population rate at 15%. Taken together, these differences demonstrate that some segments, especially telephone respondents, are demographically very different from the population. These differences raise the strong possibility that some survey segments may be very different from the population on the outcome of interest: religious service attendance.

Weights

Analyses take three different approaches to adjustment: sampling design weights, poststratification weights, and multivariate modeling. The proportionate-to-size stratified sampling design is self-weighting to the household level. Within household, a sampling design weight was computed to adjust the propensity of respondent selection within households of varying size and to appropriately weight nonresponding cases selected for follow-up personal interviews in the second phase of the two-phase sampling design. This sampling design weight is necessary for the computation of each estimate.

Poststratification weights, commonly used to adjust for survey nonresponse, are also computed and used to adjust estimates for comparison with results using only sampling design weights and those from multivariable models using demographic controls. Poststratification adjustment entails reweighing the sample to look like the population based on what is known about both, which is typically limited to demographic characteristics. Adjustments are made with the expectation that the distribution of sample variables whose population distribution is unknown will be adjusted toward their true population distributions along with the distribution of demographics. Poststratification weights are computed here using age, race, ethnicity, educational attainment, sex, and marital status. Estimates from the American Community Survey for the five strata are used as

TABLE 1 Comparing Demographic Groups from Sample Segments to the Population

	Subsample 1: Telephone (%)			Subsample 2: IVR (%)		Population (ACS; %)
	Segment 1	Segment 2	Segment 3	Segment 4	Segment 5	
Initial outcome	Respondent	Nonrespondent	No phone number available	Respondent	Nonrespondent	
Interview mode	Telephone interview	Personal interview	Personal interview	IVR interview	Personal interview	
Race/ethnicity						
White	69.6***	55.6**	38.4	67.4***	50.4	42.5
Black/African Am	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
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
Sex						
Female	60.8	48.4	51.8	62.1*	56.0	53.1
Marital status						
Married	42.9	50.0*	32.9	41.4	36.0	39.8
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

Note: Population parameters computed from the American Community Survey (ACS) for the five areas sampled for these analyses.
$^{***}p < .001$, $^{**}p < .01$, $^{*}p < .05$.

population parameters for weight construction. Poststratification weights are raked to the margins, combined with sampling weights, and trimmed for extreme values. Weight trimming prevents individuals in demographic categories that are extremely underrepresented in the sample (compared with their population proportions) from receiving very large weights and potentially biasing results. In the current study, weights were trimmed at 7.0, affecting fewer than 20 cases, and then iteratively recomputed to retain the study's original sample size.

Analysis

Three sets of estimates are used to make salient comparisons given the study design. The first two sets of estimates adjust proportions for each coding of the dependent variable (the raw ordinal outcome and the attendance propensity recode) using the design weights and the poststratification adjustments, respectively. The third set of estimates are predictions generated by two multivariable models. Fractional logistic regression is used to detect differences in the computed propensity of attending in a given week between the five study segments. A supplementary analysis uses ordinal logistic regression to compare the distribution of the attendance measure in its original form. The latter analysis is used to test a plausible just-so scenario—weekly and never attenders canceling each other out in the face of high nonresponse by moderate-attending individuals—that could lead the fractional logit to a nonsignificant result.

Key comparisons, to be discussed, use z-tests to assess statistical significance and effect size to assess substantive significance. The latter is assessed with a variation of Cohen's d that uses an arcsine transformation of each compared propensity to adjust for differences in statistical power to detect differences of equivalent size over different values (Cohen 1988; Hojat and Xu 2004):

$$d'_{ij} = |[2(\arcsin\sqrt{p_i})] - [2(\arcsin\sqrt{p_j})]|$$

A value of Cohen's d under 0.20 is typically considered to be a negligible effect, 0.20 is considered a small effect, 0.50 a medium effect, and 0.80 a large effect. As Cohen and others (Glass et al. 1981) have warned about using this "t-shirt size" rubric too rigidly, effect sizes will be used here with some flexibility to inform the interpretation of findings.

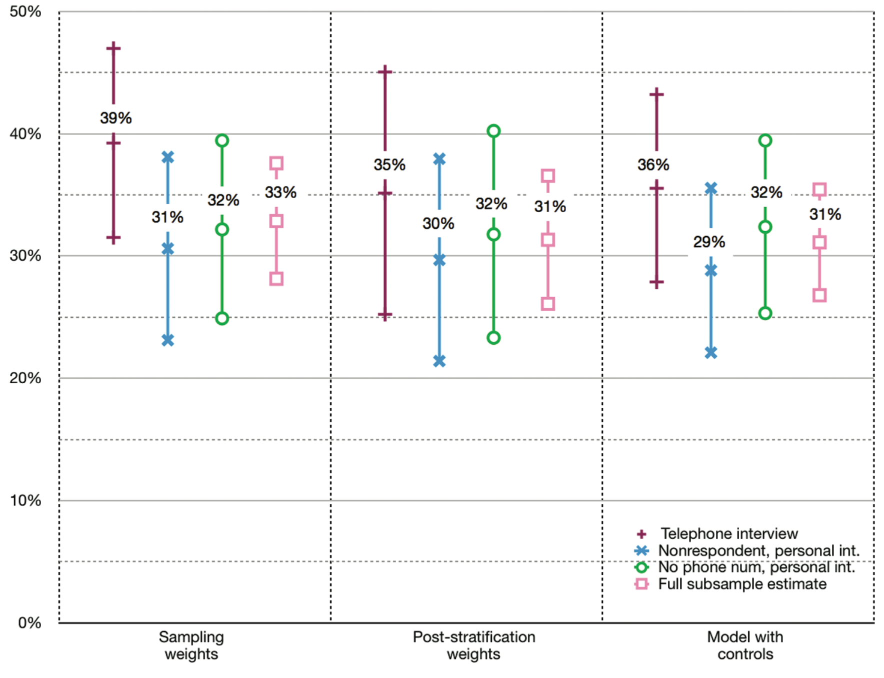
In each analysis, estimates (proportions) are presented for each set of models, accompanied by 95% confidence intervals. Estimates of attendance from the telephone interviews are compared with those from (1) personal interviews with respondents living at addresses to which phone numbers cannot be matched, and (2) follow-up personal interviews with telephone nonrespondents. Second, estimates of attendance from the IVR subsample are compared between respondents interviewed by IVR and nonrespondents who completed a follow-up personal interview. Within each subsample, full-sample estimates of attendance (including all segments in that subsample) are computed for comparison to the initial segment (telephone or IVR). This comparison allows us to produce estimates for a realistic hypothetical: how much bias would have been incurred had data collection stopped without follow-up interviews of nonrespondents in both subsamples and without pursuing households lacking a matched telephone number in the subsample assigned to telephone interviews?

RESULTS

Telephone Interview Subsample

First, I compare the attendance propensities between study segments in the subsample assigned to telephone interviewing. Figure 2 presents the propensity of religious service attendance for telephone respondents, nonrespondents subsequently interviewed in-person, and personal interview respondents from households without matched telephone numbers. The first estimates, adjusted using only sampling design weights, show an 8 percentage-point difference between telephone respondents (39%) and nonrespondents subsequently interviewed in-person (31%), and a 7 percentage-point difference between telephone respondents and respondents interviewed in-person because no phone number could be matched to their household (32%). Although these differences may look notable, they are not statistically significant (z = 1.57, p = .12; z = 1.31, p = .19) although they approach a small effect size (d' = 0.18; d' = 0.15).

FIGURE 2. Estimated propensities to attend, with 95% confidence intervals, by sample segment and estimation strategy, telephone subsample only.



Poststratification adjustment narrows these differences and increases the confidence intervals around the estimates. The difference in attendance propensity between telephone respondents (35%) and nonrespondents subsequently interviewed in-person (30%) is narrowed to 5 percentage-points, while the difference between telephone respondents and respondents interviewed in-person because no phone number could be matched to their household (32%) is narrowed to 3 percentage-points. Neither of these differences are statistically (z = 0.83, p = .41; z = 0.51, p = .61) or substantively (d' = 0.12; d' = 0.07) significant.

A multivariable model (table 2, model 1) controlling for demographic variables produced similar predicted proportions. These estimates show a 7 percentage-point difference between telephone respondents (36%) and nonrespondents subsequently interviewed in-person (29%), and a 4 percentage-point difference between telephone respondents and respondents interviewed in-person because no phone number could be matched to their household (32%). While the first of these differences may look substantial, neither is statistically significant (z = 1.30, p = .19; z = 0.58, p = .56) although the first approaches a small effect, but is still modest in size (d' = 0.14; d' = 0.07).

A second set of analyses use an ordinal logistic regression (table 2, model 2) to generate predicted proportions for each of the original ordinal attendance categories. Figure 3 illustrates these estimates for the lowest ("never") and highest ("once a week or more") attendance categories. The middle categories are omitted from the figure as estimates are identical or nearly identical between study segments. A 6 percentage-point difference emerges between the predicted proportion of telephone respondents (21%) reporting "never" and the estimates for both nonrespondents subsequently interviewed in-person (27%) and respondents interviewed in-person because no phone number could be matched to their household (27%). Neither difference is statistically significant (z = 1.56, p = .12; z = 1.48, p = .14) and both approach a small effect but are modest in size (d' = 0.15; d' = 0.14). Similarly, the predicted proportion of respondents reporting "once a week" show a 7 percentage-point difference between telephone respondents (30%) and both nonrespondents subsequently interviewed in-person (23%) and respondents interviewed in-person because no phone number could be matched to their household (23%). Neither difference is statistically significant (z = 1.54, p = .12; z = 1.48, p = .14) although they both approach a small effect size (d' = 0.15 for both).

IVR subsample

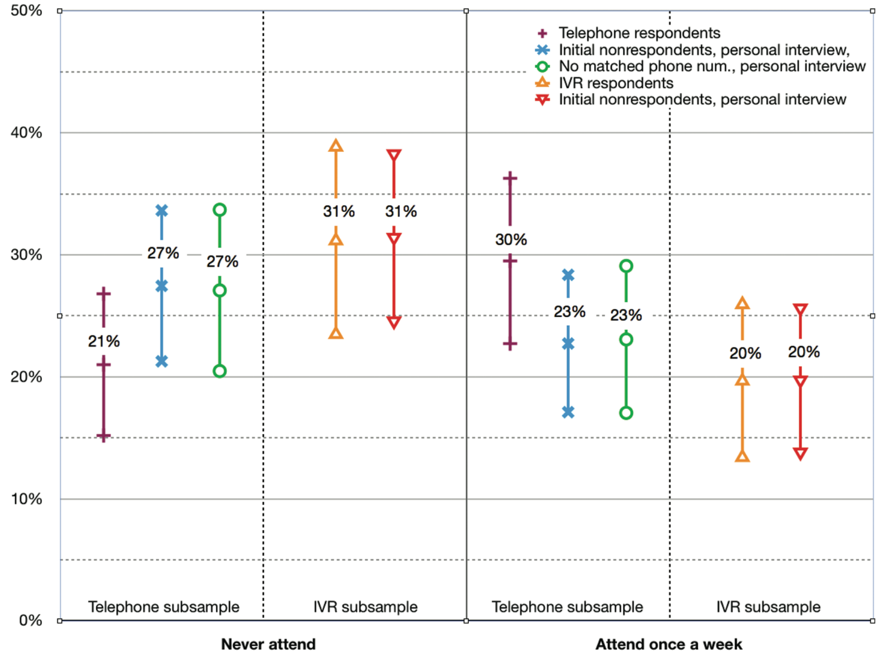
The second set of estimates compare the attendance propensities between study segments in the subsample assigned to IVR interviews, comparing IVR respondents to nonrespondents subsequently interviewed in-person (figure 4). The marginal propensities for IVR respondents and nonrespondents subsequently interviewed in-person in each set of comparisons—using sampling design weights (25% and 26%, respectively), post-stratification adjustments (22% and 28%, respectively), and multivariable model with controls (28% for both)—fail to differ

TABLE 2 Predicting Attendance Propensity Using Fractional and Ordinal Logistic Regressions

	Model 1. Fractional logit			Model 2. Ordinal logit		
	Coefficient	*p*	SE	Coefficient	*p*	SE
Study segments						
Telephone						
Respondent	0.16		0.27	0.37		0.25
Nonrespondent, personal interview	−0.18		0.26	−0.01		0.24
No phone number, personal interview						
IVR						
Respondent	−0.24		0.29	−0.23		0.26
Nonrespondent, personal interview	−0.24		0.28	−0.24		0.25
Sex						
Female	0.09		0.23	0.13		0.18
Male						
Marital status						
Married	0.27		0.25	0.26		0.19
Not married						
Age						
18–39 years						
40–64 years	0.64**		0.24	0.71***		0.21
65 and older	0.74**		0.28	0.52*		0.27
Race/ethnicity						
White	−0.80**		0.28	−0.76**		0.21
Black/African American						
Asian, other race	−0.64		0.35	−0.73*		0.33
Latino/a, any race	−0.23		0.41	−0.20		0.34
Education						
Less than HS	0.07		0.40	0.23		0.34
High school	0.13		0.27	0.05		0.24
Some college	0.03		0.26	0.14		0.23
College degree+						
Neighborhood/suburb						
Dorchester	0.05		0.33	−0.01		0.31
Jamaica Plain	−0.68		0.38	−0.87*		0.35
Mattapan						
Milton	0.41		0.40	0.41		0.37
Quincy	−0.55		0.38	−0.63		0.35
Constant/Cutpoint 1	−0.75		0.42	−1.22		0.38
Cutpoint 2				0.02		0.38
Cutpoint 3				0.61		0.37
Cutpoint 4				0.87		0.37
Cutpoint 5				1.24		0.38
N	661			661		
χ^2 (df)	57.4 (18)			68.7 (18)		

Note: Omitted categories in italics.
$^{***}p < .001$, $^{**}p < .01$, $^{*}p < .05$.

FIGURE 3. Estimated propensities to attend never or weekly, with 95% confidence intervals, by sample segment and estimation strategy.

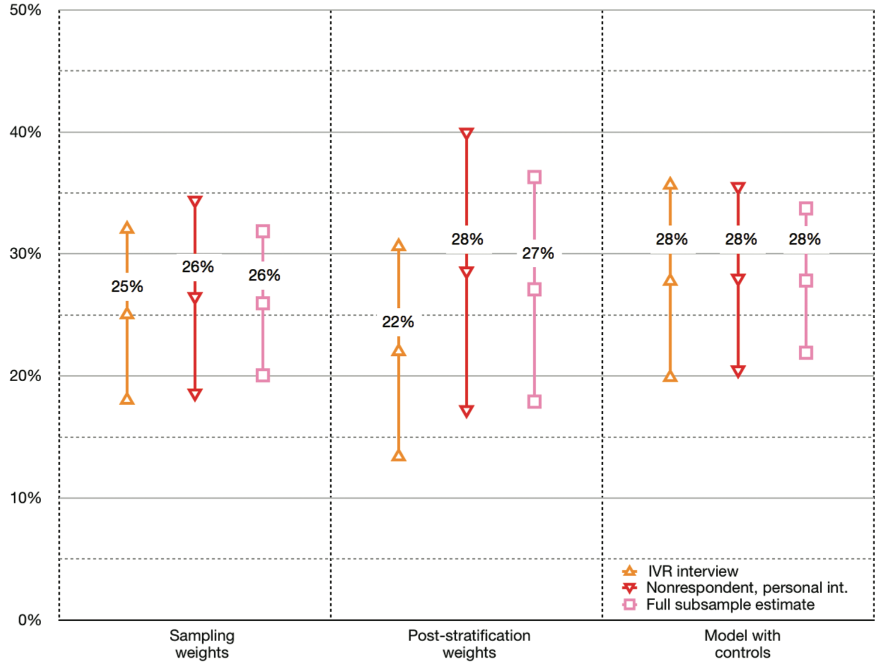


from each other, statistically ($z = 0.24$, $p = .81$; $z = 0.89$, $p = .38$; $z = 0.02$, $p = .99$) and substantively ($d' = 0.03$, 0.15, and 0.00, respectively). Predictions from the ordinal logistic regression (table 2, model 2), illustrated in figure 3, demonstrate no differences between the proportions of IVR respondents and nonrespondents interviewed in-person reporting "never" (31% for both) or "once a week" (20% for both). Unsurprisingly, neither of these differences are statistically ($z = 0.00$, $p = .98$; $z = 0.00$, $p = .98$) nor substantively ($d' = 0.00$ for both) significant.

DISCUSSION AND CONCLUSION

In the context of other larger and well-established sources of bias, namely measurement (Brenner 2011a; Hadaway et al. 1993, 1998; Presser and Stinson 1998), the contribution of nonresponse to bias in survey estimates of religious service attendance appears relatively minor in these data. The largest relative bias in the study inflates the estimate for telephone respondents (using sampling weights) by less than 20% over the full telephone subsample estimate—6 percentage-points from 33% to 39%. Although this difference does seem substantial on its face, the effect size is below Cohen's cutoff for a small effect ($d' = 0.13$) and is modest in size, especially when it is compared with the effect of measurement bias (which typically registers a medium effect size). Note that the computation of effect size is independent of the standard error of the estimate. Thus, while

FIGURE 4. Estimated propensities to attend, with 95% confidence intervals, by sample segment and estimation strategy, IVR subsample only.



increasing the sample size would reduce the standard error and potentially make this difference statistically significant, the effect size would remain unchanged, *ceteris paribus*. Moreover, poststratification adjustment or a model-based approach using demographic controls further reduce relative bias. In comparison, measurement bias inflates survey estimates of religious service attendance by up to 100% (Brenner 2011a; Hadaway et al. 1993, 1998; Presser and Stinson 1998) and is not well explained by demographic covariates (Brenner 2012).

The design of the study limits its ability to evaluate the potential for other error sources, such as mode effects. Direct comparisons between respondents in IVR and telephone modes are limited as being in the latter is conflated with the ability to match a phone number to the case. However, one potential measurement error can be assessed. Some of the extant research has suggested that moving from an interviewer-administered mode to a self-administered mode increases the validity of the self-report by reducing the need for the respondent to impress the interviewer; here, with a report of frequent attendance (Kreuter et al. 2008). However, the lack of difference between IVR (a self-administered mode) respondents and nonrespondents who subsequently completed an in-person interview suggests that no such bias is occurring here. Moreover, results in the subsample assigned to telephone interviewing do not appear to map cleanly on to an explanation based in social desirability bias. Telephone interviews arguably increase the privacy of the interview (as others present cannot hear the questions being asked) and decrease the effect of the social presence of the interviewer, but in this study respondents completing personal interviews report lower rates of attendance than those on the phone, possibly due to skepticism over the legitimacy of telephone interview requests (Holbrook et al. 2003).

Two important limitations of this study should be noted. First, that the follow-up interviews of nonrespondents themselves encounter nonresponse makes nonresponse bias a possibility in these estimates. This is a typical problem with studies of nonresponse as they rely on follow-up interviews, subsequent interviews in longitudinal design, and similar techniques that also encounter nonresponse. To assess the potential of residual nonresponse to alter these findings, I make some assumptions about the counterfactual case in which no residual nonrespondents remain (i.e., we are able to interview every nonrespondent following the initial wave). I estimated the attendance rate that would be needed to generate at least a small effect size difference ($d \geq 0.20$) between the achieved sample and a full sample that includes information from all the residual nonrespondents. In the subsample assigned to a telephone interview, residual nonrespondents would need a reported attendance rate of 18% *or lower* to achieve an overall attendance rate of 22%. The difference between this hypothetical rate and that from the achieved sample (31%) is the minimum needed to reach a small effect size. In the subsample assigned to a IVR interview, residual nonrespondents would need a reported attendance rate of 17% *or lower* to achieve an overall attendance rate of 19%. The difference between this hypothetical rate and that from the achieved sample (28%) is the minimum needed to reach a small effect size. Although these rates are possible, they are nearly half those of the achieved sample segments, and as such, seem implausible. Thus, it seems unlikely that residual nonresponse would alter these findings dramatically.

Second, although this study rigorously estimates potential nonresponse bias, the use of personal interviewing limited the research geographically to Boston. Given the city's and the region's relative irreligiosity (Gallup 2018; Public Religion Research Institute 2015), the mechanism generating nonresponse may differ between Boston and elsewhere in the United States. Notably, however, the attendance question was the only religion question on this survey that focused on health and correlates of health and was described to respondents as such during recruitment. Thus, it is unlikely that Boston area respondents would have reacted differentially in a way that was related to the religious service attendance question. Moreover, the patterns of nonresponse in the initial wave of telephone surveys are similar to those from general population surveys as whites, older adults, and those with higher education tend to be more likely to agree to the request for survey participation (Curtin et al. 2000; Groves et al. 2000; Voigt et al. 2003).

With this list of limitations as an important caveat, these findings suggest that the contribution of nonresponse to bias in estimates of attendance is modest in size, and may be negligible, especially in comparison with the well-established threat attributed to measurement biases. Moreover, nonresponse bias appears to be well attenuated in this study by the demographic covariates commonly associated with religious service attendance and survey response—race, ethnicity, education, sex, marital status, and, especially age. Thus, the mechanism generating nonresponse in these data appears to be ignorable with cases either missing completely at random (MCAR) for the IVR subsample or MCAR or missing at random (MAR), once demographic controls are applied, for the telephone subsample. In sum, the findings of the current research support those from prior work (Keeter et al. 2000, 2006). Future work should extend this research to other parts of the United States.

FUNDING

This work was supported by a grant to the author from the National Science Foundation (SES-1424433).

REFERENCES

Abraham, Katharine G., Sara Helms, and Stanley Presser. 2009. "How Social Processes Distort Measurement: The Impact of Survey Nonresponse on Estimates of Volunteer Work in the United States." *American Journal of Sociology* 114: 1129–65.

Abraham, Katharine G., Aaron Maitland, and Suzanne M. Bianchi. 2006. "Nonresponse in the American Time Use Survey: Who Is Missing From the Data and How Much Does It Matter?" *Public Opinion Quarterly* 70: 676–703.

Blumberg Stephen J., and Julian V. Luke. 2017. *Wireless Substitution: Early Release of Estimates from the National Health Interview Survey, January–June 2017*. National Center for Health Statistics. https://www.cdc.gov/nchs/data/nhis/earlyrelease/wireless201712.pdf. Accessed April 14, 2018.

Brenner, Philip S. 2011a. "Exceptional Behavior or Exceptional Identity? Overreporting of Church Attendance in the US." *Public Opinion Quarterly* 75: 19–41.

———. 2011b. "Identity Importance and the Overreporting of Religious Service Attendance: Multiple Imputation of Religious Attendance Using American Time Use Study and the General Social Survey." *Journal for the Scientific Study of Religion* 50: 103–15.

———. 2012. "Investigating the Effect of Bias in Survey Measures of Church Attendance." *Sociology of Religion* 73: 361–83.

———. 2016. "Cross-National Trends in Religious Service Attendance." *Public Opinion Quarterly* 80: 563–83.

———. 2017. "Narratives of Response Error from Cognitive Interviews of Survey Questions About Normative Behavior." *Sociological Methods & Research* 46: 540–64.

Brick, J. Michael, and Douglas Williams. 2013. "Explaining Rising Nonresponse Rates in Cross-Sectional Surveys." *Annals of the American Academy of Political and Social Science* 645: 36–59.

Chaves, Mark, and James C. Cavendish. 1994. "More Evidence on US Catholic Church Attendance." *Sociology of Religion* 33: 376–81.

Cohen, Jacob. 1988. *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ: Lawrence Erlbaum.

Curtin, Richard, Stanley Presser, and Eleanor Singer. 2000. "The Effects of Response Rate Changes on the Index of Consumer Sentiment." *Public Opinion Quarterly* 64: 413–28.

Gallup Organization. 2018. "The Religious Regions of the U.S." http://news.gallup.com/poll/232223/religious-regions.aspx. Accessed April 18, 2018.

Gauchat, Gordon. 2012. "Politicization of Science in the Public Sphere: A Study of Public Trust in the United States, 1974 to 2010." *American Sociological Review* 77: 167–87.

Glass, Gene V., Barry McGaw, and Mary Lee Smith. 1981. *Meta-analysis in Social Research*. Beverly Hills, CA: Sage.

Groves, Robert M. 2006. "Nonresponse Rates and Nonresponse Bias in Household Surveys." *Public Opinion Quarterly* 70: 646–75.

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

Groves, Robert M., Eleanor Singer, and Amy Corning. 2000. "Leverage-Saliency Theory of Survey Participation: Description and an Illustration." *Public Opinion Quarterly* 64: 299–308.

Groves, Robert M., Mick P. Couper, Stanley Presser, Eleanor Singer, Roger Tourangeau, Giorgina P. Acosta, and Lindsay Nelson. 2006. "Experiments in Producing Nonresponse Bias." *Public Opinion Quarterly* 70: 720–36.

Hadaway, C. Kirk, Penny Long Marler, and Mark Chaves. 1993. "What the Polls Don't Show: A Closer Look as US Church Attendance." *American Sociological Review* 58: 741–52.

———. 1998. "Overreporting Church Attendance in America: Evidence That Demands the Same Verdict." *American Sociological Review* 63: 122–30.

Hojat, Mohammadreza, and Gang Xu. 2004. "A Visitor's Guide to Effect Sizes." *Advances in Health Sciences Education* 9: 241–9.

Holbrook, Allyson, Melanie C. Green, and Jon A. Krosnick. 2003. "Telephone Versus Face-to-Face Interviewing of National Probability Samples with Long Questionnaires." *Public Opinion Quarterly* 67: 79–125.

Johnson, David R., and Jared L. Peifer. 2017. "How Public Confidence in Higher Education Varies by Social Context." *Journal of Higher Education* 88: 619–644.

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: 759–79.

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

Kreuter, Frauke, Stanley Presser, and Roger Tourangeau. 2008. "Social Desirability Bias in CATI, IVR, and Web Surveys: The Effects of Mode and Question Sensitivity." *Public Opinion Quarterly* 72: 847–65.

Malka, Ariel, Yphtach Lelkes, Sanjay Srivastava, Adam B. Cohen, and Dale T. Miller. 2012. "The Association of Religiosity and Political Conservatism: The Role of Political Engagement." *Political Psychology* 33: 275–99.

Pew Research Center. 2017. Sharp Partisan Divisions in Views of National Institutions. http://www.people-press.org/2017/07/10/sharp-partisan-divisions-in-views-of-national-institutions/. Accessed July 11, 2017.

Presser, Stanley and Linda Stinson. 1998. "Data Collection Mode and Social Desirability Bias in Self-Reported Religious Attendance." *American Sociological Review* 63: 137–45.

Public Religion Research Institute. 2015. The Top Two Religious Groups That Dominate American Cities. https://www.prri.org/spotlight/the-top-two-religious-traditions-that-dominate-american-cities. Accessed July 13, 2016.

Voigt, Lynda F., Thomas D. Koepsell, and Janet R. Daling. 2003. "Characteristics of Telephone Survey Respondents According to Willingness to Participate." *American Journal of Epidemiology* 157: 66–73.

Weakliem, David, and Wayne J. Villemez. 2004. "Public Attitudes Toward the Census: Influences and Trends." *Social Science Quarterly* 85: 857–71.

Woodberry, Robert D. 1998. "When Surveys Lie and People Tell the Truth: How Surveys Oversample Church Attenders." *American Sociological Review* 63: 119–22.