

Accounting for Nonignorable Unit Nonresponse and Attrition in Panel Studies with Refreshment Samples

Tracy Schifeling, Christine Cheng, Jerome P. Reiter and D. Sunshine Hillygus*

Word count: 9600

Abstract

Panel surveys typically suffer from attrition, which can lead to biased inference when basing analysis only on cases that complete all waves of the panel. Unfortunately, the panel data alone cannot inform the extent of the bias due to attrition, so analysts must make strong and untestable assumptions about the missing data mechanism. Many panel studies also include refreshment samples, which are data collected from a random sample of new individuals during some later wave of the panel. Refreshment samples offer information that can be utilized to correct for biases induced by nonignorable attrition while reducing reliance on strong assumptions about the attrition process. To date, these bias correction methods have not dealt with two key practical issues in panel studies: unit nonresponse in the initial wave of the panel and in the refreshment sample itself. As we illustrate, nonignorable unit nonresponse can significantly compromise the analyst's ability to use the refreshment samples for attrition bias correction. Thus, it is crucial for analysts to assess how sensitive their inferences—corrected for panel attrition—are to different assumptions about the nature of the unit nonresponse. We present an approach that facilitates such sensitivity analyses, both for suspected nonignorable unit nonresponse in the initial wave and in the refreshment sample. We illustrate the approach using simulation studies and an analysis of data from the 2007-2008 Associated Press/Yahoo News election panel study.

Key Words: Bayesian, longitudinal, missing, nonignorable, selection.

*Department of Statistical Science, Box 90251, Duke University, Durham, NC 27708(e-mail: tracy@stat.duke.edu). Department of Statistical Science, Box 90251, Duke University, Durham, NC 27708(e-mail: christine.cheng09@gmail.com). Department of Statistical Science, Box 90251, Duke University, Durham, NC 27708 (e-mail: jerry@stat.duke.edu). Department of Political Science, Box 90204, Duke University, Durham, NC 27708 (e-mail: hillygus@duke.edu).

1. INTRODUCTION

Longitudinal or panel surveys, in which the same individuals are interviewed repeatedly at different points in time, have widespread use in political science (Hillygus, 2005), economics (Baltagi and Song, 2006), education (Buckley and Schneider, 2006), and sociology (Western, 2002), among other fields. Inevitably, panel surveys suffer from attrition; that is, units who respond in early waves fail to respond to later waves of the survey (Lynn, 2013). For example, in the most recent multi-wave panel survey of the American National Election Study, 47% of respondents attrited between the January 2008 baseline wave and the June 2010 follow-up wave. As is well known, attrition can result in biased inferences when the propensity to drop out is systematically related to the substantive outcome of interest (Olsen, 2005; Behr et al., 2005; Hogan and Daniels, 2008). For example, Bartels (1999) showed that differential attrition of respondents in the 1992-1996 American National Election Study panel resulted in an overestimation of political interest in the population.

Many panel surveys also include refreshment samples: cross-sectional, random samples of new respondents given the questionnaire at the same time as a second or subsequent wave of the panel. Refreshment samples offer information that can be leveraged to correct for nonignorable attrition via statistical modeling. Specifically, as described by Hirano et al. (1998, 2001) and Bhattacharya (2008), the analyst can estimate an additive nonignorable (AN) model, which comprises a joint model for the survey variables coupled with a selection model for the attrition process. Recently, Deng et al. (2013) show that the AN model can be extended to panels with more than two waves and one refreshment sample, including scenarios where attriters in one wave return to the sample in a later wave.

To date, applications of the AN model have largely ignored a key complication in panel surveys, namely that the initial panel and the refreshment sample may be subject to unit nonresponse (e.g., sampled individuals refuse to cooperate, cannot be contacted, or are otherwise unable to participate). For example, in their analysis of the Dutch Transportation Panel, Hirano et al. (1998) treated the 2886 households that completed the first interview as the baseline, disregarding that this represented only 47% (2886/6128) of sampled households in that wave.

Bhattacharya (2008) uses the Current Population Survey to design illustrative simulations of an AN model without any unit nonresponse in the panel and refreshment sample. Other applications of the AN model that disregard unit nonresponse include those in Nevo (2003) and Das et al. (2011).

Using only respondents in the initial wave and refreshment sample, even if their cross-sectional weights are calibrated to trusted population estimates, implicitly assumes that the unit nonresponse is missing (perhaps completely) at random (MAR). This implies strong assumptions about the unit nonresponse, e.g., the distributions of survey variables are identical for respondents and nonrespondents within groups defined by cross-tabulations of limited sets of variables. Unfortunately, as we illustrate later, when the unit nonresponse in the initial wave or refreshment sample is not missing at random (NMAR), treating the unit nonresponse as MAR compromises the analyst's ability to use the refreshment samples to correct for attrition bias. Given sharp declines in survey response rates in recent years (e.g., Peytchev, 2013) and heightened concerns about the impact on the sample representativeness (e.g., Groves et al., 2002; Singer, 2006), disregarding unit nonresponse in applications of the AN model seems difficult to justify.

In this article, we present an approach to estimating the AN model for two-wave panels with nonignorable unit nonresponse in the initial wave or refreshment sample. The approach facilitates a process for assessing the sensitivity of inferences corrected for panel attrition to various assumptions about the nature of the nonignorable unit nonresponse. Such sensitivity analyses are the best way to handle unit nonresponse, since the data do not offer information about NMAR missing values (Little and Rubin, 2002; Hogan and Daniels, 2008; Molenberghs et al., 2008). The basic idea is to introduce selection models for the unit nonresponse with interpretable sensitivity parameters that can be tuned to reflect various departures from ignorable missing data mechanisms. This approach is similar in spirit to methods used to assess the sensitivity of estimated treatment effects to unmeasured confounding in observational studies (e.g., Rosenbaum, 2010; Schwartz et al., 2012). We present the methodology for binary outcomes and categorical predictors, although similar ideas could be used for other types of variables. We estimate the models using Bayesian methods and Markov Chain Monte Carlo simulation. We also present two extensions: (i) a two-step approximation to the fully Bayesian solution that can facilitate exploration of different unit nonresponse mechanisms, and (ii) an approach for handling

nonignorable unit nonresponse when the panel includes more than one wave before the refreshment sample is collected. We apply the methodology to data from the AP-Yahoo News 2008 Election Panel Study (APYN), focusing on measuring campaign interest in the 2008 presidential election.

The remainder of the article is organized as follows. In Section 2, we offer a review of the AN model assuming no unit nonresponse in the initial wave and the refreshment sample. In Section 3 we present methods for assessing sensitivity of results to nonignorable unit nonresponse in both the initial wave and refreshment sample. We present both the fully Bayesian model and two-step approximation. In Section 4 we extend the model to a scenario with an intermediate wave between the baseline and refreshment sample where we allow for missing data in the intermediate wave, caused by either unit nonresponse or panel attrition. In Section 5, we illustrate the methodology on an analysis of APYN data.

2. REVIEW OF ADDITIVE NONIGNORABLE MODEL

Consider a two wave panel of N_p individuals with a refreshment sample of N_r new subjects in the second wave. For all $N = N_p + N_r$ subjects, the data include q time-invariant variables $\mathbf{X} = (X_1, \dots, X_q)$, such as demographic or frame variables. Let Y_1 be a scalar response variable of substantive interest collected in the initial wave of the panel. Let Y_2 represent the corresponding response variable collected in wave 2. Here, we assume that Y_1 and Y_2 are the same variable collected at different waves, although this is not necessary for the AN model. Among the N_p individuals, $N_{cp} < N_p$ provide at least some data in the second wave, and the remaining $N_{ip} = N_p - N_{cp}$ individuals drop out of the panel. Thus, the refreshment sample includes only (\mathbf{X}, Y_2) ; by design, Y_1 are missing for all the individuals in the refreshment sample. For now, we presume no nonresponse in Y_1 in the panel, in Y_2 in the refreshment sample, and in \mathbf{X} for all cases.

For each individual $i = 1, \dots, N$, let $R_i = 1$ if individual i would remain in wave 2 if included in wave 1, and let $R_i = 0$ if individual i would drop out of wave 2 if included in wave 1. We note that R_i is fully observed for all individuals in the panel but is missing for the individuals in the refreshment sample, since this latter set is not provided the chance to respond in wave 1.

The AN model requires a joint model for (Y_1, Y_2) and R given \mathbf{X} . We write this as

$$(Y_1, Y_2) | \mathbf{X} \sim f(\mathbf{X}, \Theta) \quad (1)$$

$$R | Y_1, Y_2, \mathbf{X} \sim g(\mathbf{X}, Y_1, Y_2, \boldsymbol{\tau}), \quad (2)$$

where Θ and $\boldsymbol{\tau}$ represent sets of model parameters. In the APYN application, Y_1 and Y_2 are binary and \mathbf{X} is exclusively categorical. In this case, one specification of the AN model is

$$Y_{1i} | \mathbf{X}_i \sim \text{Bern}(\pi_1), \quad \text{logit}(\pi_1) = \alpha_0 + \boldsymbol{\alpha}_X \mathbf{X}_i \quad (3)$$

$$Y_{2i} | Y_{1i}, \mathbf{X}_i \sim \text{Bern}(\pi_{2i}), \quad \text{logit}(\pi_{2i}) = \beta_0 + \beta_1 Y_{1i} + \boldsymbol{\beta}_X \mathbf{X}_i \quad (4)$$

$$R_i | Y_{1i}, Y_{2i}, \mathbf{X}_i \sim \text{Bern}(\pi_{Ri}), \quad \text{logit}(\pi_{Ri}) = \tau_0 + \tau_1 Y_{1i} + \tau_2 Y_{2i} + \boldsymbol{\tau}_X \mathbf{X}_i. \quad (5)$$

Here, we use a subscript X to denote a vector of coefficients in front of \mathbf{X}_i ; for example, $\boldsymbol{\alpha}_X$ represents the vector of coefficients of \mathbf{X} in the model for Y_1 . Throughout the article, we implicitly assume that the analyst uses dummy coding to represent the levels of each variable in \mathbf{X} and a separate regression coefficient for each dummy variable.

The key assumption in the AN model is that the probability of attrition depends on Y_1 and Y_2 through a function g that is additive in Y_1 and Y_2 ; that is, no interactions between Y_1 and Y_2 are allowed in (2). Additivity is necessary to enable point identification of the model parameters—i.e., the maximum likelihood estimate of the parameters in the model has a unique value—as there is insufficient information to estimate interactions between Y_1 and Y_2 in (2). We note that the model also can accommodate interaction terms among subsets of (\mathbf{X}, Y_1) and interaction terms among subsets of (\mathbf{X}, Y_2) . For further discussion of the additivity assumption and when it might not hold, including the consequences of incorrectly assuming additivity, see Deng et al. (2013).

As described by Hirano et al. (1998), AN models include MAR and NMAR models as special cases. When $\tau_2 = 0$ and at least one element among $\{\boldsymbol{\tau}_X, \tau_1\}$ does not equal zero, the attrition is MAR. When $\tau_2 \neq 0$, the attrition is NMAR. Hence, the AN model allows the data to decide between MAR and (certain types of) NMAR attrition mechanisms.

To illustrate how the AN assumption enables point identification, consider Y_1 and Y_2 as binary responses without any other variables, i.e., \mathbf{X} is empty. The data then comprise a

contingency table with eight cells, $\{Y_1 = y_1, Y_2 = y_2, R = r : y_1, y_2, r \in \{0, 1\}\}$. As described by Deng et al. (2013), the panel alone yields six constraints on these eight cells, namely (i) the four values of $P(Y_1 = y_1, Y_2 = y_2, R = 1)$ for each combination of $y_1, y_2 \in \{0, 1\}$, (ii) the equation $P(Y_1 = 1, Y_2 = 0, R = 0) + P(Y_1 = 1, Y_2 = 1, R = 0) = P(Y_1 = 1, R = 0)$, and (iii) the requirement that the total probability in the eight cells sums to one. The refreshment sample adds an additional constraint, which can be expressed via the equation $P(Y_2 = 1) = P(Y_2 = 1, R = 0) + P(Y_2 = 1, R = 1)$. Thus, the combined panel and refreshment data can identify models with six parameters plus the sum to one constraint. Excluding coefficients in front of \mathbf{X} , this is exactly the number of parameters in (3) – (5); hence, the AN model is point-identified.

We note that AN models can be constructed for outcomes that are not binary. For example, Hirano et al. (1998) present an AN model for normally distributed outcomes, and Bhattacharya (2008) presents an approach to estimating conditional expectations under an AN assumption. In this article, we consider AN models based on logistic regressions as in (3) – (5); hence, in addition to the no-interactions assumption, we implicitly assume that logistic regressions are reasonable statistical models for the outcomes of interest.

3. UNIT NONRESPONSE IN THE BASELINE OR REFRESHMENT SAMPLE

We now describe how to adapt the AN model when there is nonignorable unit nonresponse in the initial wave and in the refreshment sample. Our adaptations facilitate investigations of the sensitivity of inferences to the nonignorable nonresponse. Here, we assume that units that do not respond in the initial wave do not have the opportunity to respond in future waves, as is common in panel surveys. We describe the approach using the models in (3) – (5), but the ideas apply more generally.

For all N_p units in the original panel, let $W_{1i} = 1$ if unit i responds to the initial wave of the panel, and $W_{1i} = 0$ otherwise. For all N_r units in the refreshment sample, let $W_{2i} = 1$ if unit i responds in the refreshment sample, and let $W_{2i} = 0$ otherwise. Here, W_{1i} and W_{2i} describe missingness in Y_{1i} and Y_{2i} , respectively. To simplify explanations, for now we assume \mathbf{X}_i is known for all N units. In practice, this could arise when \mathbf{X}_i comprises sampling frame or administrative variables, or in surveys that have previously collected demographic information, such as internet

	\mathbf{X}	Y_1	W_1	Y_2	R	W_2
Panel	✓	✓	1	✓	1	?
				?	0	?
Refreshment sample	✓	?	0	?	?	?
				✓	?	1
	✓	?	?	?	?	0

Figure 1: Two-wave panel with attrition and unit nonresponse in the baseline and refreshment sample.

panels. We discuss relaxing this assumption about \mathbf{X}_i in Section 4. Figure 1 displays the pattern of observed and missing data across the two waves and refreshment sample.

Figure 1 reveals several key features about W_1 and W_2 . First, W_1 and W_2 are never observed jointly. Thus, when modeling, we cannot identify any parameter reflecting an association between W_1 and W_2 given the other variables. Second, since we do not observe R and W_2 jointly, we cannot identify any parameter reflecting an association between R and W_2 given the other variables. Third, since we do not observe R when $W_1 = 0$, we cannot identify any parameter reflecting an association between R and W_1 given the other variables.

Based on these constraints, we add the following selection models to (3), (4), and (5).

$$W_{1i}|Y_{1i}, \mathbf{X}_i \sim \text{Bern}(\pi_{W_{1i}}), \quad \text{logit}(\pi_{W_{1i}}) = \gamma_0 + \gamma_1 Y_{1i} + \gamma_X \mathbf{X}_i \quad (6)$$

$$W_{2i}|Y_{2i}, \mathbf{X}_i \sim \text{Bern}(\pi_{W_{2i}}), \quad \text{logit}(\pi_{W_{2i}}) = \rho_0 + \rho_1 Y_{2i} + \rho_X \mathbf{X}_i. \quad (7)$$

We let γ_1 and ρ_1 be sensitivity parameters set by the analyst, and (γ_0, γ_X) and (ρ_0, ρ_X) be estimated from the data.

For illustration, we return to the scenario with Y_1 and Y_2 binary and \mathbf{X} empty. In this case, the combined panel and refreshment sample essentially represent a contingency table with $2^5 = 32$ cells. We do not observe counts in any of the cells directly because each unit has at least one variable missing, either by design or because the unit did not respond. For example, even for the panel units who respond in both wave one and wave two, we do not know whether they would have responded had they been in the refreshment sample, i.e., their values of W_2 are unobserved. As a consequence, the ten parameters in (3) through (7) are not point-identified, as we now explain.

The complete-cases in the panel data provide four constraints, namely

$$P(Y_1 = y_1, Y_2 = y_2, W_1 = 1, R = 1) = \sum_{w_2 \in \{0,1\}} P(Y_1 = y_1, Y_2 = y_2, W_1 = 1, R = 1, W_2 = w_2) \text{ for } y_1, y_2 \in \{0, 1\}. \quad (8)$$

The attriters from the cases with $W_1 = 1$ offer two constraints, namely

$$P(Y_1 = y_1, W_1 = 1, R = 0) = \sum_{y_2, w_2} P(Y_1 = y_1, Y_2 = y_2, W_1 = 1, R = 0, W_2 = w_2) \text{ for } y_1 \in \{0, 1\}. \quad (9)$$

The respondents in the refreshment sample data provide another two constraints, namely

$$P(Y_2 = y_2, W_2 = 1) = \sum_{y_1, w_1, r} P(Y_1 = y_1, Y_2 = y_2, W_1 = w_1, R = r, W_2 = 1) \text{ for } y_2 \in \{0, 1\}. \quad (10)$$

Thus, the data offer only eight constraints to estimate ten parameters (ignoring the sum-to-one constraint, which is present in the data and the complete model).

To enable identification of the parameters in the complete model, we have to add two constraints to (3) through (7). We do so by fixing γ_1 and ρ_1 at user-specified constants; that is, we make them sensitivity parameters. These can be readily interpreted in terms of odds of response, with values far from zero indicating nonignorable nonresponse. For example, setting $\gamma_1 = 0$ implies that the nonresponse in Y_1 in the panel is missing at random (MAR) and can be ignored, whereas setting $\gamma_1 = .5$ implies that individuals with $Y_1 = 1$ have $\exp(.5) \approx 1.65$ times the odds of responding in wave 1 as individuals with $Y_1 = 0$. Similarly, setting $\rho_1 = 0$ implies that the nonresponse in the refreshment sample is MAR, whereas setting $\rho_1 = -.25$ implies that individuals with $Y_2 = 1$ have $\exp(-.25) \approx .77$ times the odds of responding in the refreshment sample as individuals with $Y_2 = 0$. Of course, it is not possible for the analyst to know γ_1 or ρ_1 . However, as we illustrate in Section 5.2, the analyst can insert a range of values of each parameter into (3) through (7) to assess the sensitivity of analyses to assumptions about the nonignorable nonresponse.

With set values of the sensitivity parameters, the model can be estimated using Gibbs sampling. Given a current draw of the completed-data, we sample the parameters using Metropolis steps. Given a current draw of the parameters, we draw values of the missing data according to the logistic regressions in (3) through (7) as follows. For ease of notation, we

suppress parameters from the conditional densities.

- Cases with $(W_1 = 1, R = 0)$: Sample the missing (Y_2, W_2) from

$$p(Y_2, W_2 \mid \mathbf{X}, Y_1, R = 0) = p(Y_2 \mid \mathbf{X}, Y_1, R = 0)p(W_2 \mid \mathbf{X}, Y_1, Y_2, R = 0) \quad (11)$$

$$p(Y_2 \mid \mathbf{X}, Y_1, R = 0) \propto p(R = 0 \mid \mathbf{X}, Y_1, Y_2)p(Y_2 \mid \mathbf{X}, Y_1) \quad (12)$$

and $p(W_2 \mid \mathbf{X}, Y_1, Y_2, R = 0)$ is given by (7).

- Cases with $W_1 = 0$: Sample the missing (Y_1, Y_2, W_2, R) from

$$p(Y_1, Y_2, W_2, R \mid \mathbf{X}, W_1 = 0) = p(Y_1 \mid \mathbf{X}, W_1 = 0)p(Y_2 \mid \mathbf{X}, Y_1)p(R \mid \mathbf{X}, Y_1, Y_2)p(W_2 \mid \mathbf{X}, Y_2) \quad (13)$$

$$p(Y_1 \mid \mathbf{X}, W_1 = 0) \propto p(W_1 = 0 \mid \mathbf{X}, Y_1)p(Y_1 \mid \mathbf{X}) \quad (14)$$

and the remaining densities from (4), (5), and (7).

- Cases with $W_2 = 1$: Sample the missing (Y_1, W_1, R) from

$$p(Y_1, W_1, R \mid \mathbf{X}, Y_2, W_2 = 1) = p(Y_1 \mid \mathbf{X}, Y_2)p(W_1 \mid \mathbf{X}, Y_1)p(R \mid \mathbf{X}, Y_1, Y_2) \quad (15)$$

$$p(Y_1 \mid \mathbf{X}, Y_2, W_2 = 1) \propto p(Y_2 \mid \mathbf{X}, Y_1)p(Y_1 \mid \mathbf{X}) \quad (16)$$

and the remaining densities from (5), and (6).

- Cases with $W_2 = 0$: Sample the missing (Y_1, Y_2, W_1, R) from

$$p(Y_1, Y_2, W_1, R \mid \mathbf{X}, W_2 = 0) = p(Y_1 \mid \mathbf{X}, W_2 = 0)p(W_1 \mid \mathbf{X}, Y_1)p(Y_2 \mid \mathbf{X}, Y_1, W_2 = 0) \times p(R \mid \mathbf{X}, Y_1, Y_2) \quad (17)$$

$$p(Y_2 \mid \mathbf{X}, Y_1, W_2 = 0) \propto p(W_2 = 0 \mid \mathbf{X}, Y_1, Y_2)p(Y_2 \mid \mathbf{X}, Y_1) \quad (18)$$

$$p(Y_1 \mid \mathbf{X}, W_2 = 0) \propto \sum_{y_2} p(W_2 = 0 \mid \mathbf{X}, Y_2 = y_2)p(Y_2 = y_2 \mid \mathbf{X}, Y_1)p(Y_1 \mid \mathbf{X}). \quad (19)$$

and all densities from (3) through (7).

We now demonstrate how to account for nonignorable nonresponse in the baseline and refreshment sample, as well as nonignorable attrition. To do so, we design simulations in which

we use the true values of γ_1 and ρ_1 . In a sensitivity analysis in a genuine setting, the analyst would not know the true values of the sensitivity parameters and would examine a range of plausible values; we illustrate this in Section 5.2. Here, our purpose is to show that the approach can result in accurate parameter estimates when true values are used, thus demonstrating that the approach offers a meaningful sensitivity analysis.

We simulate 100 datasets from our model in (3) through (7), and estimate the parameters for each dataset under two conditions: (i) with the sensitivity parameters both set to their true values, and (ii) with the sensitivity parameters both set to 0. For each simulated dataset, the panel has 10,000 units and the refreshment sample has 5,000 units. We include eight binary variables, six of which comprise \mathbf{X} . We set the values of the logistic regression coefficients to generate significant associations among the survey variables and substantial, but potentially realistic, nonresponse rates. Specifically, the parameters in the model in (3) through (7) are all set to values between -0.5 and 1. Both sensitivity parameters γ_1 and ρ_1 are set to 1. These parameter settings lead to roughly 50% of the panel completing both waves, 25% of the panel dropping out at wave two, and 25% of the panel not responding in wave one. The refreshment sample has roughly a 60% response rate. We use independent Cauchy priors with a scale parameter of 10 for the intercept terms and a scale parameter of 2.5 for all other coefficients, following the advice of Gelman et al. (2008).

As evident in Figure 2, when the sensitivity parameters are set to their true values, for all parameters the simulated coverage rates are between 0.93 and 0.99. On the other hand, as evident in Figure 3, when the sensitivity parameters are set wrongly to zero (which corresponds to MAR nonresponse), 25 parameters have simulated coverage less than 0.9. The five parameters in the top left of Figure 3 are the five intercepts, and the parameter in the bottom left is the coefficient of Y_2 in the model for R .

We also ran simulations for both cases where one sensitivity parameter is set to the truth and the other is set to zero. When we correctly account for the unit nonresponse in the baseline but incorrectly assume the refreshment sample unit nonresponse is MAR, some parameters in models for Y_2 , R and W_2 are far from their true values. When we correctly account for the unit nonresponse in the refreshment sample but incorrectly assume the baseline unit nonresponse is MAR, some parameters in the models for Y_1 and W_1 are far from their true values. Thus,

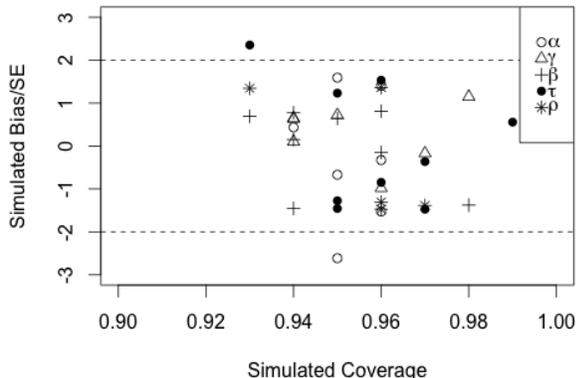


Figure 2: Simulated coverage versus simulated Bias/SE with sensitivity parameters set to truth for model in (3) - (7).

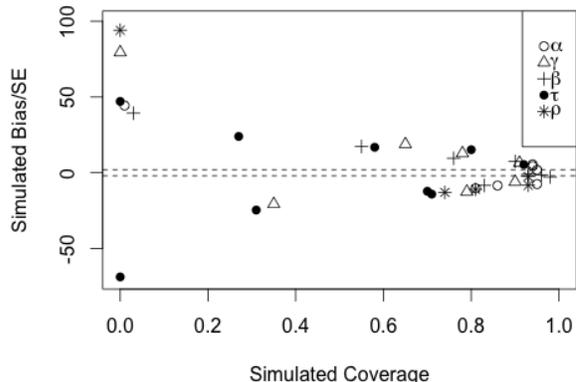


Figure 3: Simulated coverage versus simulated Bias/SE with sensitivity parameters set to 0 for model in (3) - (7).

when unit nonresponse in both the baseline and refreshment sample is nonignorable, it is important to assess sensitivity to assumptions about both sources of missing data.

We also investigated the performance of the sensitivity analysis procedure when the underlying models are misspecified. Specifically, we generated data from robit regression models (Liu, 2004; Gelman et al., 2004; Gelman and Hill, 2007) with one degree of freedom using the same parameter values as before, but fit the sequential logistic regression models to the data. When we set the sensitivity parameters to the values used previously, inferences continue to be more reliable than when setting the sensitivity parameters to zero. See the supplementary material for details and results.

Although the selection models in (6) and (7) are flexible, some analysts may prefer to characterize the missing data in the baseline or refreshment sample with alternate missing data mechanisms, for example, pattern mixture models. In such cases, the analyst can implement a two-step approach based on multiple imputation (Rubin, 1987). First, the analyst fills in the missing data caused by unit nonresponse in the baseline or refreshment sample, creating M completed datasets. After the baseline and refreshment sample are completed, we have only panel attrition, which can be handled with an AN model. Inferences can proceed via the usual multiple imputation combining rules (Rubin, 1987) or, for Bayesian analyses based on (3) through (5), by mixing the draws of parameters from the M chains (Zhou and Reiter, 2010). We ran simulation

studies of this approach basing completions of the nonresponse on the selection models in (6) and (7); results were similar to those seen in Figure 2 and Figure 3.

While we focus on binary outcomes, similar approaches could be used for sensitivity analyses with multinomial or continuous outcomes. For example, if Y_1 is continuous, the analyst could specify a sensitivity parameter for Y_1 (or some function of it) and interpret it using the usual change in log-odds ratio. If Y_1 is a categorical variable with $d > 2$ levels, we could replace γ_1 with $(\gamma_{1,1}, \dots, \gamma_{1,d-1})$, potentially allowing each value of Y_1 to have a different effect on the probability of responding. With large d , the analyst most likely would need to make simplifying assumptions about the $(\gamma_{1,1}, \dots, \gamma_{1,d-1})$ to use the approach in practice; for example, set $\gamma_{1,j} = 0$ for some set of j and $\gamma_{1,j} = \gamma_{1,j'}$ for the complementary set.

Finally, while we prefer to examine inferences at particular values of the sensitivity parameters, it is also possible to specify informative prior distributions on these parameters. These could be based on expert opinion or previous experience about the nature of the unit nonresponse. We illustrate this approach in the analysis of the APYN data in Section 5.3.

4. INTERMEDIATE WAVES

In some panel surveys, multiple waves may take place between the collection of the baseline data and refreshment sample. These intermediate waves can be subject to panel attrition. In this section, we show that the ideas of Section 3 can be extended to correct for nonignorable attrition in intermediate waves as well. We use a setting with three waves including a refreshment sample at the third wave only. We consider two variations: (i) monotone dropout, when those units that do not respond in wave 2 do not have the opportunity to respond in wave 3, and (ii) intermittent dropout, when units that do not respond in wave 2 are potentially able to respond in wave 3. This latter scenario often arises in online panel surveys, like the APYN election poll, in which all wave 1 respondents are invited to participate in subsequent surveys, even if they failed to participate in a previous wave. For simplicity, as in the original AN model, we assume there is no unit nonresponse in the baseline nor in the refreshment sample. Adding selection models like those in Section 3 does not present additional complications.

For both scenarios, we use the notation of Section 3 with slight modification. For $i = 1, \dots, N$, let $W_{2i} = 1$ if individual i provides a value of Y_2 if included in wave 1. Note that W_2 is missing for

	\mathbf{X}	Y_1	Y_2	W_2	Monotone		Intermittent	
					Y_3	R	Y_3	R
Panel	✓	✓	✓	1	✓	1	✓	1
					?	0	?	0
Refreshment sample:	✓	?	?	0	?	?	✓	1
							?	0
	✓	?	?	?	✓	?	✓	?

Figure 4: Two examples of a three wave study with a refreshment sample in the third wave and dropout in the intermediate wave. With monotone dropout, panel units that fail to respond in wave 2 ($W_2 = 0$) do not have the opportunity to continue in wave 3. With intermittent dropout, panel units that fail to respond in wave 2 can participate in wave 3. In either scenario, the patterns of observed and missing data are the same for the baseline wave and refreshment samples.

cases in the refreshment sample. For $i = 1, \dots, N$, let Y_{3i} be the response of unit i at wave 3; let $R_i = 1$ if individual i would provide a value in wave 3 if included in wave 2; and, let $R_i = 0$ if individual i would drop out of wave 3 if included in wave 2. Here, R_i is not observed for any units that are not followed up from wave 2, which includes cases in the refreshment sample and with $W_2 = 0$ for monotone dropout. The two variations are displayed in Figure 4.

4.1 Monotone Dropout

Comparing Figure 4 with Figure 1, it is apparent that the case with monotone dropout has similar structure as the case for a two-wave panel with nonresponse in the baseline and complete data in the refreshment sample. In particular, consider (Y_1, Y_2) combined as variables in a hypothetical “wave 1” and Y_3 as a hypothetical “wave 2.” In this case, only some of the variables in the “wave 1” are subject to nonignorable nonresponse. Taking advantage of this mapping, we can write a model for the monotone dropout case using the general format,

$$(Y_1, Y_2, Y_3) \mid \mathbf{X} \sim f(\mathbf{X}, \Theta) \tag{20}$$

$$R \mid Y_1, Y_2, Y_3, \mathbf{X} \sim g(\mathbf{X}, Y_1, Y_2, Y_3, \boldsymbol{\rho}) \tag{21}$$

$$W_2 \mid Y_1, Y_2, \mathbf{X} \sim h(\mathbf{X}, Y_1, Y_2, \boldsymbol{\gamma}), \tag{22}$$

where Θ , $\boldsymbol{\rho}$, and $\boldsymbol{\gamma}$ represent sets of model parameters. As in the AN model, to enable model identification we exclude interactions between Y_3 and (Y_1, Y_2) when specifying (21).

In the case of binary outcomes, we can write this as a sequence of logistic regressions,

$$Y_{1i} | \mathbf{X}_i \sim \text{Bern}(\pi_1), \quad \text{logit}(\pi_1) = \alpha_0 + \boldsymbol{\alpha}_X \mathbf{X}_i \quad (23)$$

$$Y_{2i} | Y_{1i}, \mathbf{X}_i \sim \text{Bern}(\pi_{2i}), \quad \text{logit}(\pi_{2i}) = \beta_0 + \beta_1 Y_{1i} + \boldsymbol{\beta}_X \mathbf{X}_i \quad (24)$$

$$Y_{3i} | Y_{1i}, Y_{2i}, \mathbf{X}_i \sim \text{Bern}(\pi_{3i}), \quad \text{logit}(\pi_{3i}) = \tau_0 + \tau_1 Y_{1i} + \tau_2 Y_{2i} + \tau_3 Y_{1i} Y_{2i} + \boldsymbol{\tau}_X \mathbf{X}_i \quad (25)$$

$$R_i | Y_{1i}, Y_{2i}, Y_{3i}, \mathbf{X}_i \sim \text{Bern}(\pi_{R_i}), \quad \text{logit}(\pi_{R_i}) = \rho_0 + \rho_1 Y_{1i} + \rho_2 Y_{2i} + \rho_3 Y_{3i} + \rho_4 Y_{1i} Y_{2i} + \boldsymbol{\rho}_X \mathbf{X}_i \quad (26)$$

$$W_{2i} | Y_{1i}, Y_{2i}, \mathbf{X}_i \sim \text{Bern}(\pi_{W_{2i}}), \quad \text{logit}(\pi_{W_{2i}}) = \gamma_0 + \gamma_1 Y_{1i} + \gamma_2 Y_{2i} + \boldsymbol{\gamma}_X \mathbf{X}_i \quad (27)$$

Here, γ_2 is a sensitivity parameter that is set to reflect nonignorable response at wave 2. We can include interactions between Y_1 and Y_2 in (25) and (26) since we have cases with all these variables observed. We note that Hogan and Daniels (2008) discuss a similar model without refreshment samples, treating both γ_2 and ρ_3 as sensitivity parameters. As in the AN model in Section 2, the refreshment sample provides enough information to identify ρ_3 . Deng et al. (2013) show that with an additional refreshment sample at wave two, γ_2 would be identifiable from the data as well.

We now illustrate via simulations that the model can adjust for nonignorable monotone dropout in an intermediate wave and, hence, can offer valid sensitivity analyses. We simulate 100 datasets from the model in (23) through (27) and use a Metropolis-within-Gibbs sampler to estimate the parameters for each dataset under two conditions: (i) with the sensitivity parameter set to its true value, and (ii) with the sensitivity parameter set to 0, meaning we assume Y_2 is missing at random. For each simulated dataset, the panel has 10,000 units and the refreshment sample has 5,000 units. We include six binary covariates. We again set the values of the coefficients to produce reasonable nonresponse rates and associations among the variables. The parameters in the the model in (23) through (27) are all set to values between -0.6 and 0.6, except for the sensitivity parameter γ_2 which is set to 3. These parameter settings lead to roughly 50% of the panel completing all three waves, 25% of the panel dropping out at wave 2, and 25% of the panel dropping out in wave 3. Again we use independent Cauchy priors with a scale parameter of 10 for the intercept terms and a scale parameter of 2.5 for all other coefficients (Gelman et al., 2008). We should note that one of the 100 simulated trials produced obviously invalid results, we suspect due to lack of convergence of the MCMC. We replaced this rogue trial with another run.

As evident in Figure 5, when we set $\gamma_2 = 3$, the simulated coverage rate for all parameters is

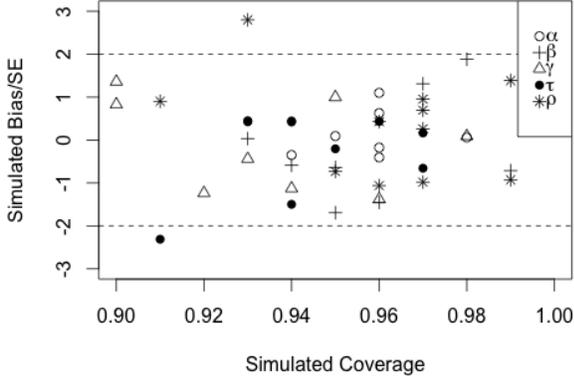


Figure 5: Simulated coverage versus simulated Bias/SE with sensitivity parameter set to truth for monotone dropout model in (23) - (27).

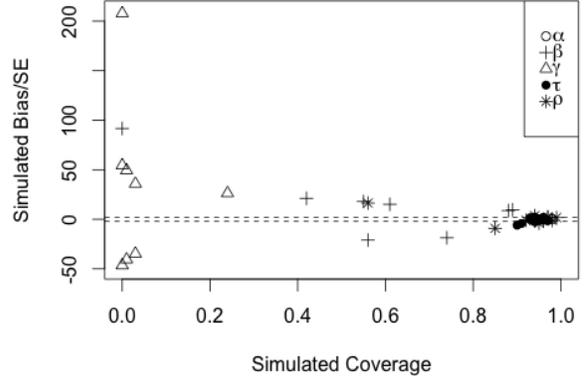


Figure 6: Simulated coverage versus simulated Bias/SE with sensitivity parameter set to 0 for monotone dropout model in (23) - (27).

at least 90%, When we instead set $\gamma_2 = 0$, incorrectly assuming Y_2 is missing at random, many parameters have low simulated coverage rates, especially the parameters in the models associated with wave two. The parameter in the top left of Figure 6 with the largest bias is the γ intercept.

4.2 Intermittent Dropout

Under intermittent dropout, units that do not respond in wave two still are given the opportunity to respond in wave three. As a result, we now observe $W_2 = 0$ jointly with both Y_3 and R . This information offers four additional constraints, so that we can add four parameters to (23) through (27). In particular, we add terms for W_2 and Y_1W_2 in the logit equations in (25) and (26). We have

$$\text{logit}(\pi_{3i}) = \tau_0 + \tau_1 Y_{1i} + \tau_2 Y_{2i} + \tau_3 Y_{1i} Y_{2i} + \tau_4 W_{2i} + \tau_5 W_{2i} Y_{1i} + \tau_X \mathbf{X}_i \quad (28)$$

$$\text{logit}(\pi_{R_i}) = \rho_0 + \rho_1 Y_{1i} + \rho_2 Y_{2i} + \rho_3 Y_{3i} + \rho_4 Y_{1i} Y_{2i} + \rho_5 W_{2i} + \rho_6 W_{2i} Y_{1i} + \rho_X \mathbf{X}_i. \quad (29)$$

Here, γ_2 remains a sensitivity parameter that is set by the analyst.

To illustrate, we run simulations with the same dimension and sample size as in the monotone dropout simulation study. The parameters in the model in (23), (24), (28), (29), and (27) are all between -0.4 and 0.6, except for the sensitivity parameter γ_2 which is set to 1.5. These parameter settings lead to about 45% of the panel completing all three waves, about 27% not responding in

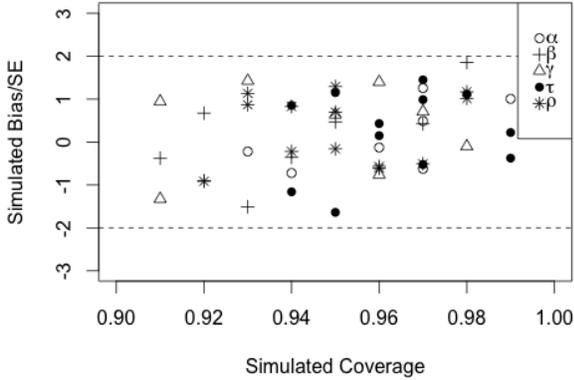


Figure 7: Simulated coverage versus simulated Bias/SE with sensitivity parameter set to truth in intermittent dropout model in (23), (24), (28), (29), and (27).

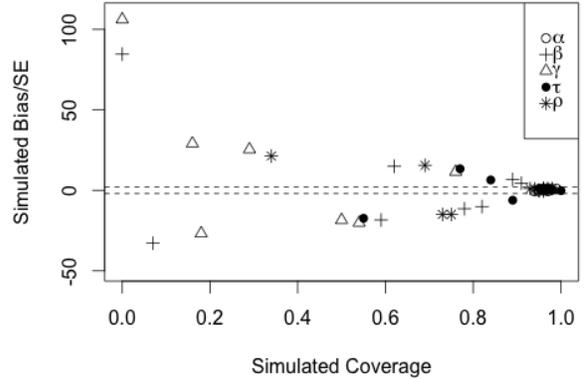


Figure 8: Simulated coverage versus simulated Bias/SE with sensitivity parameter set to 0 in intermittent dropout model in (23), (24), (28), (29), and (27).

wave 3 only, 17% not responding in wave 2 only, and 11% not responding in wave 2 or 3.

As evident in Figure 7 and Figure 8, when $\gamma_2 = 1.5$, all the parameters have near nominal simulated coverage rates and unremarkable simulated bias. When γ_2 is wrongly set to zero, many parameters have coverage rate far less than 90%. The two parameters with largest bias are the intercepts β_0 and γ_0 .

5. DEALING WITH ATTRITION AND NONRESPONSE IN AN ANALYSIS OF THE APYN

The APYN is an eleven-wave panel survey with three refreshment samples intended to measure attitudes about the 2008 presidential election. The panel was sampled from the probability-based KnowledgePanel Internet panel, which recruits panel members via a probability-based sampling method using known published sampling frames that cover 96% of the U.S. population. Sampled non-internet households are provided a laptop computer or MSN TV unit and free internet service. The study was a collaboration between the AP and Yahoo Inc., with support from Knowledge Networks (KN).

We analyze two of the eleven waves, specifically wave 1 and wave 9. Wave 1 was fielded on November 2, 2007 to $N_p = 3548$ panelists at least 18 years old, out of whom only 2730 completed

the interview (i.e., $\sum W_{1i} = 2730$). Wave 9, the last wave to include a refreshment sample, was fielded in October 2008. Individuals who failed to participate in the first wave of the survey (i.e., with $W_{1i} = 0$) were not subsequently included in the follow up waves. However, individuals who completed the first wave (i.e., $W_{1i} = 1$) were invited to participate in all subsequent waves, even if they skipped one or more of the follow-up waves. Of the 2730 wave 1 respondents, only $N_{cp} = 1715$ remain in the panel by wave 9. The refreshment sample at wave 9 was fielded to $N_r = 1085$ individuals, of whom 461 responded. For the remainder of the analysis, we refer to the November 2007 wave as “wave one” and the October 2008 wave as “wave two.”

Following Deng et al. (2013), we analyze campaign interest, which is a strong predictor of democratic attitudes and behaviors (Prior, 2010), is used in identifying likely voters in pre-election polls (Traugott and Tucker, 1984), and has been shown to be susceptible to panel attrition bias (Bartels, 1999). Our outcome of interest is based on the survey question, “How much thought, if any, have you given to candidates who may be running for president in 2008?” Following common usage of this measure (e.g., Traugott and Tucker, 1984), we dichotomize the response into those who respond “A lot” and all other responses, so that $Y_{ti} = 1$ if unit i at wave $t \in \{1, 2\}$ responds “a lot,” and $Y_{ti} = 0$ otherwise. In wave 1, 29.8% of the 2730 respondents answer $Y_{1i} = 1$. The percentage with $Y_{2i} = 1$ increases dramatically to 65.0% among the 1715 complete cases at wave two. In the refreshment sample, 72.2% of the 461 respondents answer $Y_{2i} = 1$. We predict campaign interest from age (four categories), education (college degree or not), gender (male or not), and race (black or not); see Table 1 for summaries of these variables. We note that Deng et al. (2013) ignored unit nonresponse in the refreshment samples and original panel entirely, effectively acting as if this nonresponse was MCAR.

5.1 Missing Demographic Characteristics

In the model and simulation study in Section 3, we assume that all \mathbf{X} values are known, even for units in the panel and refreshment sample that do not participate. In the APYN data we analyze, however, we do not have the demographic variables for the 818 panel units with $W_{1i} = 0$ nor for the 624 refreshment sample units with $W_{2i} = 0$. We use the following strategy to create a dataset with complete values of all \mathbf{X} .

Let $(\mathbf{x}_1, \dots, \mathbf{x}_{32})$ represent each of the $4 \times 2 \times 2 \times 2 = 32$ combinations of \mathbf{X} defined by

cross-tabulating the four demographic characteristics. We compute the weighted sample proportions from the 2012 Current Population Survey (CPS), which we call \bar{x}_k^{CPS} , where $k = 1, \dots, 32$ indexes the combinations. Let $\bar{x}_k^{PAN} = \sum_{i=1}^{N_p} I(\mathbf{X}_i = \mathbf{x}_k)/N_p$, where $I(\cdot) = 1$ when the condition inside the parentheses is true and $I(\cdot) = 0$ otherwise. These are the completed-data proportions in the panel, computed after imputation of missing \mathbf{X} . Similarly, let the completed-data proportions in the refreshment sample be $\bar{x}_k^{REF} = \sum_{i=1}^{N_r} I(\mathbf{X}_i = \mathbf{x}_k)/N_r$, where $k = 1, \dots, 32$. We impute missing \mathbf{X} values in the panel and refreshment sample so that \bar{x}_k^{PAN} and \bar{x}_k^{REF} closely match \bar{x}_k^{CPS} for all k . We do so because the CPS and APYN both target the population of U.S. adults. Let $N_{pm} = \sum_{i=1}^{N_p} (1 - W_{1i})$ be the number of cases in the panel with missing \mathbf{X} values. Similarly, let $N_{rm} = \sum_{i=1}^{N_r} (1 - W_{2i})$. Using the N_p panel cases, for $k = 1, \dots, 32$ let $\tilde{x}_k^{PAN} = \sum_{i=1}^{N_p} W_{1i} I(\mathbf{X}_i = \mathbf{x}_k)/N_p$. Define the corresponding quantities for the refreshment sample as $\tilde{x}_k^{REF} = \sum_{i=1}^{N_r} W_{2i} I(\mathbf{X}_i = \mathbf{x}_k)/N_r$. We use the following imputation algorithm.

1. Set a counter $t = 0$.
2. Find k such that $\bar{x}_k^{CPS} - \tilde{x}_k^{PAN}$ is maximized.
3. Among cases with $W_{1i} = 0$ yet to have \mathbf{X} imputed, determine how many additional cases to set the missing $\mathbf{X}_i = \mathbf{x}_k$ so that $|\bar{x}_k^{CPS} - \bar{x}_k^{PAN}|$ is minimized. Call this n_k . Let $t = t + n_k$.
4. Among cases with $W_{1i} = 0$ yet to have \mathbf{X} imputed, set n_k cases values of \mathbf{X} equal to \mathbf{x}_k .
When $t > N_{pm}$, only impute \mathbf{X} values for the remaining $(N_{pm} - t + n_k)$ cases.
5. Go back to step 2 until all cases with missing \mathbf{X} in the panel have been imputed.
6. Repeat from step 1 for the refreshment sample, replacing (N_p, N_{pm}, W_{1i}) with (N_r, N_{rm}, W_{2i}) and $(\bar{x}_k^{PAN}, \tilde{x}_k^{PAN})$ with $(\bar{x}_k^{REF}, \tilde{x}_k^{REF})$.

Table 1 displays the marginal probabilities for each demographic variable before and after the imputation. The imputation scheme generates a completed-data population that mimics the CPS marginal frequencies. When no data source is representative of the same population as the panel, the analyst should impute missing values in \mathbf{X} , for example using some MAR model such as a Bayesian bootstrap (Rubin, 1981).

We repeated the simulation study of Section 3 but allowing \mathbf{X} to be missing for nonrespondents; see the supplementary materials for details. Repeated sampling properties of the

Variable	Definition	Panel		Refreshment		CPS
		Before	After	Before	After	
AGE0	= 1 for age 18-29, 0 otherwise	.15	.21	.11	.22	.22
AGE1	= 1 for age 30-44, 0 otherwise	.28	.26	.21	.26	.26
AGE2	= 1 for age 45-59, 0 otherwise	.32	.28	.34	.27	.27
AGE3	= 1 for age above 60, 0 otherwise	.25	.25	.34	.25	.25
COLLEGE	= 1 for having college degree, 0 otherwise	.3	.28	.31	.28	.28
MALE	= 1 for male, 0 otherwise	.45	.48	.43	.48	.48
BLACK	= 1 for African American, 0 otherwise	.08	.12	.07	.13	.13

Table 1: Marginal probabilities of the demographic variables in the panel and refreshment sample before and after imputing the missing covariates. When imputing the missing demographic variables, we did so to match joint probabilities available in the 2012 Current Population Survey.

inferences for the coefficients in the regression models for Y_1 , Y_2 , and R continue to be reasonable. The inferences for some coefficients in the models for W_1 and W_2 exhibit low coverage rates. We do not consider this to be a significant concern, as the models for W_1 and W_2 are not of substantive interest—they are only used for sensitivity analysis.

The APYN data file includes survey weights at each wave. For individuals who responded to wave 1, their weights are the product of design-based weights and post-stratification adjustments for unit nonresponse. In general, analysts should ignore such post-stratification adjustments when assessing sensitivity of design-based analyses to nonignorable unit nonresponse. Instead, after imputing the missing values for the nonrespondents in the initial wave, analysts can use the design weights for all N_p records, assuming they are available for the unit nonrespondents. Unfortunately, in the APYN data we do not have the design weights for nonrespondents to the initial wave (or the refreshment sample), so that we are not able to perform traditional design-based analyses after imputing the missing values. Since our goal is to illustrate the sensitivity analysis rather than perform a specific finite population inference, absent the design-weights for nonrespondents we use Bayesian approaches for inference for the panel. We note that, in cases where design weights are available for nonrespondents, any design-based analysis of wave 2 data also should ignore post-stratification adjustments for attrition, since the AN model is used to account for nonignorable attrition.

5.2 Results of Sensitivity Analysis

We begin by fitting the AN model using main effects for the demographic variables and assuming MAR nonresponse in both wave one and the refreshment sample. This is equivalent to fitting the model in (3) through (7) with $\gamma_1 = \rho_1 = 0$. We use Cauchy prior distributions with a scale parameter of 10 for the intercept terms and scale parameter of 2.5 for all other terms (Gelman et al., 2008); we obtained similar results using Cauchy(0, 10) prior distributions on all of the coefficients. To fit the model, we run three MCMC chains from different starting points for 200,000 iterations. We discard the first 20,000 iterations as burn-in and keep every tenth draw. Diagnostics suggest the chains converge with adequate effective sample sizes (at least 500 for each parameter). For each of set of $(\alpha, \beta, \gamma, \tau, \rho)$, the multivariate potential scale reduction factor is very close to one. We used 200,000 iterations to ensure satisfactory effective sample sizes, particularly for the intercepts.

Table 2 summarizes the parameter estimates for the models in (3) through (5). The coefficient of Y_2 in the model for R is significant, which suggests the attrition is nonignorable. Based on the model with MAR nonresponse, holding all else constant, a panel participant with “a lot” of interest at wave two is less likely to respond in wave two than a disinterested participant. This tracks the disparities in the marginal probabilities of $Y_{2i} = 1$: 65% in the complete-cases in the panel and 72.2% in the refreshment sample. The strongest predictor of interest in wave two is interest in wave one. Additionally, older and college-educated participants are more likely to be interested in the campaign.

The negative coefficient of Y_2 in the attrition model contradicts conventional wisdom that politically interested respondents are *less* likely to attrite (Bartels, 1999). This counterintuitive finding potentially could be an artifact of nonignorable nonresponse in the panel or refreshment sample. For example, suppose politically disinterested individuals refused to respond in the refreshment sample at higher rates than politically interested individuals, given covariates. The resulting over-representation of $Y_{2i} = 1$ in the complete cases in the refreshment sample would show up as nonignorable attrition like that seen in Table 2.

We therefore analyze the sensitivity of conclusions to various mechanisms for nonignorable nonresponse in wave one and the refreshment sample. For nonresponse in wave one, we consider

Dependent Variable	Y_1	Y_2	R
Intercept	-1.58 (-1.85, -1.32)	-.13 (-.53, .26)	.52 (.07, 1.13)
AGE1	.21 (-.09, .51)	.10 (-.21, .40)	.22 (-.03, .48)
AGE2	.75 (.47, 1.05)	.33 (.03, .64)	.32 (.06, .57)
AGE3	1.21 (.92, 1.50)	1.08 (.73, 1.43)	.47 (.19, .76)
COLLEGE	.11 (-.07, .30)	.67 (.44, .91)	.59 (.39, .79)
MALE	-.04 (-.20, .13)	-.03 (-.23, .17)	.11 (-.05, .27)
BLACK	.73 (.45, 1.02)	.13 (-.25, .53)	-.25 (-.54, .03)
Y_1	–	2.12 (1.78, 2.47)	.41 (.16, .66)
Y_2	–	–	-.82 (-1.65, -.13)

Table 2: Coefficient estimates and 95% posterior intervals for AN model assuming MAR nonresponse in wave one and in refreshment sample ($\gamma_1 = \rho_1 = 0$). Results based on all $N = 4633$ sampled individuals in wave 1 and the refreshment sample. The model for R suggests nonignorable attrition.

$\gamma_1 \in \{\log 0.5, 0, \log(2)\}$. These three settings imply that sampled individuals with “a lot” of interest in wave one have, respectively, half, the same, or double the odds of responding in wave one as other individuals. For nonresponse in the refreshment sample, we consider $\rho_1 \in \{0, \log(2), \log(3)\}$. These values reflect an assumption that politically interested individuals have greater odds of responding to a cross-sectional survey than politically disinterested individuals, as it is well established that people with more interest in the survey topic tend to respond at higher levels than those less interested in the topic (Goyder, 1987; Groves et al., 2000, 2004). This direction could explain the 7.2% disparity in the panel and refreshment sample respondents’ marginal percentages of Y_2 , whereas assuming politically-interested respondents are less likely to participate in wave 2 only would magnify the apparent bias due to attrition. We consider it unlikely that $\rho_1 > \log(3)$, which would imply that politically interested individuals have more than 3 times higher odds of responding.

To illustrate the approach, we perform a sensitivity analysis for all coefficients in the models in (3) through (7). Of course, analysts need not perform such extensive analyses and instead can focus on results most relevant to their analysis.

Table 3, Table 4, and Table 5 display results under different settings of the sensitivity parameters, using the Gibbs sampler outlined in Section 3. In the model for Y_1 (Table 3), the results are insensitive to values of ρ_1 , which reflects assumptions about nonresponse in the

refreshment sample. This is not the case for γ_1 , which reflects assumptions about nonresponse in the first wave. The intercept varies from about -1.3 when interested wave one participants are less likely to respond to -1.8 when interested wave one participants are more likely to respond. Other coefficients, notably the age coefficients, change magnitudes as well, although the general conclusions remain the same across all values of ρ_1 .

In the model for Y_2 (Table 4), the results are insensitive to the values of γ_1 and, for the most part, to the values of ρ_1 as well. Only the intercept varies substantially with the values of ρ_1 . We discuss implications of this for estimating marginal probabilities of Y_2 below.

In the model for R (Table 5), the estimates for the coefficient of Y_2 vary widely depending on the sensitivity parameter settings. In particular, when we assume that the nonresponse in the refreshment sample is MAR ($\rho_1 = 0$), the attrition is NMAR with 95% credible intervals for the coefficient of Y_2 including only negative values. When we assume that nonresponse in the refreshment sample is NMAR with $\rho_1 = \log(3)$, we again estimate nonignorable attrition but now these interval estimates include all positive values; that is, a politically interested panel member has greater odds of responding at wave two than a disinterested one. When we assume that $\rho_1 = \log(2)$, the 95% credible intervals for the coefficient of Y_2 include zero, so that we would not have evidence to reject MAR attrition. Clearly, the nature of the unit nonresponse affects whether or not we call the attrition nonignorable.

We also estimate $P(Y_2 = 1)$ in the population for each setting of the sensitivity parameters. To do so, we use a Bayesian version of a post-stratified estimator. We write the population probability as

$$\begin{aligned}
 P(Y_2 = 1) &= \sum_{k=1}^{32} P(Y_2 = 1|Y_1 = 1, \mathbf{X} = \mathbf{x}_k)P(Y_1 = 1|\mathbf{X} = \mathbf{x}_k)P(\mathbf{X} = \mathbf{x}_k) \\
 &\quad + \sum_{k=1}^{32} P(Y_2 = 1|Y_1 = 0, \mathbf{X} = \mathbf{x}_k)P(Y_1 = 0|\mathbf{X} = \mathbf{x}_k)P(\mathbf{X} = \mathbf{x}_k)
 \end{aligned} \tag{30}$$

where \mathbf{x}_k ranges over the thirty-two possible demographic combinations. We set each $P(\mathbf{X} = \mathbf{x}_k) = \bar{x}_k^{CPS}$. For each scenario, we compute 50,000 draws of $P(Y_2 = 1)$ based on (30) and draws from the posterior distributions of the parameters from each model. These 50,000 draws represent the posterior distribution of $P(Y_2 = 1)$.

(γ_1, ρ_1)	Intercept	AGE1	AGE2	AGE3	COLLEGE	MALE	BLACK
(0, 0)	-1.58 (-1.85, -1.32)	.21 (-.09, .51)	.75 (.47, 1.05)	1.21 (.92, 1.50)	.11 (-.07, .30)	-.04 (-.20, .13)	.73 (.45, 1.02)
(0, log 2)	-1.59 (-1.85, -1.32)	.21 (-.09, .52)	.76 (.47, 1.05)	1.21 (.92, 1.51)	.11 (-.07, .29)	-.04 (-.20, .13)	.73 (.45, 1.01)
(0, log 3)	-1.59 (-1.86, -1.33)	.22 (-.08, .52)	.77 (.49, 1.06)	1.22 (.94, 1.51)	.11 (-.07, .29)	-.04 (-.21, .13)	.72 (.44, 1.01)
(log 2, 0)	-1.81 (-2.07, -1.55)	.38 (.08, .68)	.95 (.67, 1.24)	1.32 (1.03, 1.61)	.12 (-.06, .31)	-.07 (-.24, .10)	.57 (.29, .84)
(log 2, log 2)	-1.81 (-2.08, -1.56)	.38 (.09, .68)	.95 (.67, 1.24)	1.32 (1.04, 1.62)	.12 (-.06, .30)	-.07 (-.24, .10)	.56 (.28, .84)
(log 2, log 3)	-1.82 (-2.09, -1.56)	.39 (.09, .69)	.96 (.68, 1.25)	1.33 (1.05, 1.62)	.12 (-.06, .30)	-.07 (-.24, .09)	.56 (.28, .83)
(log $\frac{1}{2}$, 0)	-1.29 (-1.55, -1.03)	.00 (-.29, .30)	.51 (.23, .80)	1.04 (.76, 1.33)	.10 (-.08, .27)	.02 (-.15, .19)	.90 (.62, 1.19)
(log $\frac{1}{2}$, log 2)	-1.30 (-1.56, -1.04)	.01 (-.29, .31)	.52 (.24, .81)	1.05 (.77, 1.34)	.10 (-.08, .28)	.02 (-.15, .18)	.89 (.61, 1.17)
(log $\frac{1}{2}$, log 3)	-1.30 (-1.57, -1.04)	.01 (-.29, .32)	.53 (.24, .82)	1.06 (.77, 1.35)	.10 (-.08, .28)	.01 (-.15, .18)	.89 (.61, 1.17)

Table 3: Posterior means and 95% posterior intervals for coefficients in regression of Y_1 on \mathbf{X} for different values of sensitivity parameters. Here, γ_1 is an additional odds of response in the first wave, and ρ_1 is an additional odds of response in the refreshment sample.

Figure 9 summarizes the posterior distributions when $\gamma_1 = 0$ and $\rho_1 \in \{\log .75, 0, \log 1.5, \log 2, \log 3\}$. When the nonresponse in the refreshment sample is MAR ($\rho_1 = 0$), the posterior mean for $P(Y_2 = 1)$ is 0.69. When we set $\rho_1 = \log 1.5$, the credible interval for $P(Y_2 = 1)$ is (0.60, 0.69), which barely includes 0.69. In contrast, setting $\rho_1 = \log 2$ results in an interval of (0.57, 0.66), which no longer includes 0.69. Thus, if it is plausible that politically interested participants in the refreshment sample have at least 1.5 times higher odds of responding to the refreshment sample in wave two than politically disinterested participants, arguably we should not feel comfortable assuming the refreshment sample unit nonresponse is MAR. Going in the other direction, if the sensitivity parameter is set to $\log .75$, the interval is (0.68, 0.76), which includes 0.69. These results suggest that the estimate of political interest in the population at wave 2 is sensitive to nonresponse in the refreshment sample, but the estimate would be significantly different only at rather substantial levels of bias (i.e., when political interested participants have 1.5 higher odds of responding).

We also estimate $P(Y_1 = 1)$ to estimate differences in interest in the presidential candidates

(γ_1, ρ_1)	Intercept	AGE1	AGE2	AGE3	COLLEGE	MALE	BLACK	Y_1
(0, 0)	-.13 (-.53, .26)	.10 (-.21, .40)	.33 (.03, .64)	1.08 (.73, 1.43)	.67 (.44, .91)	-.03 (-.23, .17)	.13 (-.25, .53)	2.12 (1.78, 2.47)
(0, log 2)	-.63 (-.99, -.26)	.17 (-.14, .48)	.44 (.14, .75)	1.21 (.87, 1.55)	.76 (.54, .99)	-.02 (-.22, .19)	.09 (-.31, .50)	2.18 (1.86, 2.53)
(0, log 3)	-.90 (-1.24, -.55)	.19 (-.11, .50)	.49 (.18, .78)	1.22 (.89, 1.56)	.79 (.57, 1.01)	.00 (-.2, .2)	.05 (-.35, .46)	2.13 (1.80, 2.47)
(log 2, 0)	-.10 (-.49, .31)	.09 (-.21, .39)	.32 (.02, .63)	1.07 (.72, 1.42)	.66 (.43, .89)	-.03 (-.23, .16)	.15 (-.24, .54)	2.11 (1.77, 2.45)
(log 2, log 2)	-.60 (-.96, -.23)	.16 (-.15, .47)	.44 (.13, .74)	1.21 (.87, 1.56)	.76 (.53, .98)	-.02 (-.22, .19)	.10 (-.30, .51)	2.19 (1.86, 2.52)
(log 2, log 3)	-.87 (-1.21, -.51)	.19 (-.11, .49)	.48 (.18, .78)	1.23 (.90, 1.56)	.79 (.57, 1.01)	.00 (-.2, .2)	.06 (-.34, .46)	2.15 (1.81, 2.49)
(log $\frac{1}{2}$, 0)	-.20 (-.59, .20)	.11 (-.20, .42)	.35 (.05, .67)	1.10 (.75, 1.45)	.68 (.45, .91)	-.03 (-.23, .16)	.11 (-.28, .51)	2.13 (1.80, 2.48)
(log $\frac{1}{2}$, log 2)	-.69 (-1.05, -.31)	.17 (-.14, .48)	.46 (.14, .76)	1.21 (.86, 1.55)	.77 (.55, 1.00)	-.01 (-.21, .19)	.07 (-.34, .49)	2.18 (1.86, 2.52)
(log $\frac{1}{2}$, log 3)	-.97 (-1.30, -.62)	.20 (-.10, .51)	.50 (.20, .81)	1.22 (.89, 1.56)	.80 (.58, 1.02)	.00 (-.19, .20)	.03 (-.37, .43)	2.10 (1.77, 2.44)

Table 4: Posterior means and 95% posterior intervals for coefficients in regression of Y_2 on (\mathbf{X}, Y_1) for different values of sensitivity parameters. Here, γ_1 is an additional odds of response in the first wave, and ρ_1 is an additional odds of response in the refreshment sample.

from wave one to wave two. We write the population probability as

$$P(Y_1 = 1) = \sum_{k=1}^{32} P(Y_1 = 1 | \mathbf{X} = \mathbf{x}_k) P(\mathbf{X} = \mathbf{x}_k). \quad (31)$$

When $\gamma_1 = 0$ and $\rho_1 \in \{\log .75, 0, \log 1.5, \log 2, \log 3\}$, the posterior mean is .30 with a 95% credible interval of (.28, .32). This interval does not overlap with the credible intervals for $P(Y_2 = 1)$. If $\rho_1 = 0$ and $\gamma_1 = \log .5$, the estimate of $P(Y_1 = 1)$ is .33 with an interval of (.31, .35). Keeping $\rho_1 = 0$ and setting $\gamma_1 = \log 2$, the estimate of $P(Y_1 = 1)$ is .27 with an interval of (.26, .29). Thus, regardless of the assumptions about unit nonresponse, there clearly was a large uptick in political interest from November 2007 to October 2008.

5.3 Using a Prior Distribution on Sensitivity Parameters

As an alternative to fixing the sensitivity parameters at various plausible values, analysts may want a single set of inferences that averages over their prior beliefs about the values of the sensitivity parameters (Molenberghs et al., 2001, 1999). In this section, we illustrate this process for the APYN analysis by constructing and using prior distributions for γ_1 and ρ_1 . To do so, we first specify prior distributions on the proportions of nonrespondents who gave “a lot” of thought to the candidates, and convert these beliefs into distributions for γ_1 and ρ_1 .

(γ_1, ρ_1)	Intercept	AGE1	AGE2	AGE3	COLLEGE	MALE	BLACK	Y_1	Y_2
(0, 0)	.52 (.07, 1.13)	.22 (-.03, .48)	.32 (.06, .57)	.47 (.19, .76)	.59 (.39, .79)	.11 (-.05, .27)	-.25 (-.54, .03)	.41 (.16, .66)	-.82 (-1.65, -.13)
(0, log 2)	.02 (-.29, .36)	.21 (-.04, .46)	.24 (-.01, .50)	.29 (-.01, .58)	.46 (.26, .65)	.12 (-.04, .28)	-.26 (-.55, .03)	.10 (-.22, .39)	.20 (-.45, .85)
(0, log 3)	-.17 (-.45, .11)	.20 (-.06, .45)	.19 (-.07, .45)	.16 (-.15, .46)	.36 (.15, .57)	.12 (-.04, .29)	-.27 (-.56, .02)	-.18 (-.57, .17)	.85 (.17, 1.62)
(log 2, 0)	.57 (.11, 1.23)	.22 (-.03, .48)	.32 (.07, .57)	.48 (.20, .76)	.60 (.4, .8)	.11 (-.05, .27)	-.25 (-.54, .03)	.43 (.18, .67)	-.90 (-1.75, -.21)
(log 2, log 2)	.04 (-.27, .38)	.21 (-.04, .46)	.25 (.0, .5)	.31 (.02, .59)	.47 (.26, .66)	.12 (-.04, .28)	-.26 (-.54, .03)	.12 (-.19, .41)	.12 (-.51, .75)
(log 2, log 3)	-.15 (-.42, .14)	.20 (-.06, .45)	.20 (-.05, .46)	.18 (-.13, .48)	.38 (.17, .58)	.12 (-.04, .29)	-.27 (-.56, .02)	-.13 (-.50, .21)	.74 (.08, 1.47)
(log $\frac{1}{2}$, 0)	.44 (.03, 1.00)	.22 (-.03, .47)	.31 (.06, .56)	.45 (.17, .73)	.58 (.38, .78)	.11 (-.05, .27)	-.25 (-.54, .03)	.38 (.13, .63)	-.69 (-1.48, -.04)
(log $\frac{1}{2}$, log 2)	-.02 (-.32, .31)	.20 (-.05, .45)	.23 (-.02, .48)	.26 (-.03, .55)	.44 (.23, .64)	.12 (-.04, .28)	-.26 (-.54, .03)	.05 (-.29, .35)	.32 (-.34, 1.00)
(log $\frac{1}{2}$, log 3)	-.20 (-.47, .07)	.19 (-.06, .45)	.18 (-.08, .44)	.12 (-.18, .42)	.34 (.13, .54)	.13 (-.03, .29)	-.28 (-.57, .02)	-.25 (-.64, .10)	1.01 (.34, 1.76)

Table 5: Posterior means and 95% posterior intervals for coefficients in regression of R on (\mathbf{X}, Y_1, Y_2) for different values of sensitivity parameters. Here, γ_1 is an additional odds of response in the first wave, and ρ_1 is an additional odds of response in the refreshment sample.

For γ_1 , we construct the prior distribution to reflect our belief that nonrespondents are not as politically interested as respondents. In wave 1 of the panel, 29.8% of the respondents indicated that they gave “a lot” of thought about the candidates. Thus, we make the 97.5 percentile of the prior distribution for $Pr(Y_1 = 1 | W_1 = 0)$ equal to .298. We set the 2.5 percentile of this prior distribution equal to .149, based on consultations with three public opinion experts. Finally, we set the prior distribution for $Pr(Y_1 = 1 | W_1 = 0)$ equal to a normal distribution with the matching 95% central interval. One could match to other distributions as well, such as a beta distribution; the normal distribution sufficed to represent our prior beliefs.

We next convert the prior distribution for $Pr(Y_1 = 1 | W_1 = 0)$ into a prior distribution for γ_1 . To do so, we use the facts that

$$Pr(W_1 = 1 | Y_1 = 0) = Pr(Y_1 = 0, W_1 = 1) / Pr(Y_1 = 0) \quad (32)$$

$$Pr(W_1 = 1 | Y_1 = 1) = Pr(Y_1 = 1, W_1 = 1) / Pr(Y_1 = 1). \quad (33)$$

In these equations, we can estimate the two joint probabilities using the empirical percentages from the N_p cases. For any estimate of the marginal probability of Y_1 , we can determine the corresponding value of (γ_0, γ_1) by unwinding the selection model for $(W_1 | Y_1)$. Ignoring the

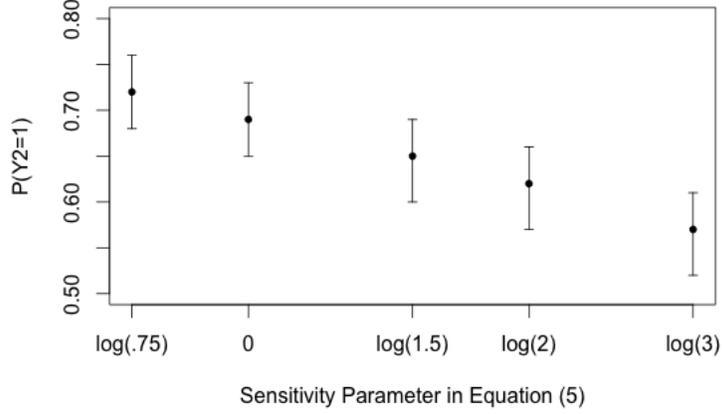


Figure 9: Posterior means and 95% credible intervals for $P(Y_2 = 1)$ for $\rho_1 \in \{\log .75, 0, \log 1.5, \log 2, \log 3\}$.

effects of \mathbf{X} in (6) for simplification, given an estimate of $Pr(Y_1 = 1)$ we have

$$\text{logit}(Pr(Y_1 = 0, W_1 = 1)/Pr(Y_1 = 0)) = \gamma_0 \quad (34)$$

$$\text{logit}(Pr(Y_1 = 1, W_1 = 1)/Pr(Y_1 = 1)) = \gamma_0 + \gamma_1. \quad (35)$$

Of course, because of the unit nonresponse, we do not know the marginal probability for Y_1 . We instead find the sample space for (γ_0, γ_1) by solving (34) and (35) for all possible empirical percentages of Y_1 . For $i = 1, \dots, 818$ (the total number of nonrespondents in wave 1 of the panel), we compute $p_1^{(i)} = 815 + i$, where 815 is the number of respondents who indicated that they gave “a lot of thought” about the candidates. Then, for each i , we solve

$$\text{logit} \left(\frac{Pr(Y_1 = 0, W_1 = 1)}{(N_p - p_1^{(i)})/N_p} \right) = \gamma_0^{(i)} \quad (36)$$

$$\text{logit} \left(\frac{Pr(Y_1 = 1, W_1 = 1)}{p_1^{(i)}/N_p} \right) = \gamma_0^{(i)} + \gamma_1^{(i)}. \quad (37)$$

The collection of $(\gamma_0^{(i)}, \gamma_1^{(i)})$ represents the sample space. We plot the implied CDF of the set of $\gamma_1^{(i)}$ using the probabilities from the normal prior distribution for $Pr(Y_1 = 1|W_1 = 0)$ —since γ_1 is effectively a transformed version of $Pr(Y_1 = 1|W_1 = 0)$ —and visually match it to a normal CDF function. For γ_1 , the resulting normal distribution has mean .39 and standard deviation .22.

For ρ_1 , we follow a similar process. For the refreshment sample nonrespondents, we use a normal distribution with central 95% interval from .361 to .722. Here, .722 is the proportion of refreshment sample respondents who had given “a lot” of thought to the candidates. For $i = 1, \dots, 624$ (the total number of nonrespondents in the refreshment sample), we set $p_2^{(i)} = 333 + i$, and solve the following two equations:

$$\text{logit} \left(\frac{\text{Pr}(Y_2 = 0, W_2 = 1)}{(N_r - p_2^{(i)})/N_r} \right) = \rho_0^{(i)} \quad (38)$$

$$\text{logit} \left(\frac{\text{Pr}(Y_2 = 1, W_2 = 1)}{p_2^{(i)}/N_r} \right) = \rho_0^{(i)} + \rho_1^{(i)} \quad (39)$$

As a result, we use a normal distribution with mean .78 and standard deviation .38.

We modify the MCMC algorithm to incorporate these prior distributions. We then run three chains each for 200,000 iterations, discarding the first 20,000 iterations as burn-in and saving every 10th draw. For each of set of $(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\tau}, \text{ and } \boldsymbol{\rho})$, the multivariate potential scale reduction factor is less than 1.01. Table 6 summarizes the posterior distributions of the regression coefficients. With these prior beliefs on the nonresponse, we find no strong evidence for NMAR attrition, even though the model that assumes MAR nonresponse in wave one and the refreshment sample (see Table 2) does suggest NMAR attrition. We also computed posterior inferences for $P(Y_2 = 1)$, obtaining a posterior mean of .63 (95% interval: .54 to .71). Comparing this to Figure 9, the posterior mean is in the same range as other settings of sensitivity parameters, and the 95% interval is wider than under fixed settings of the sensitivity parameters.

5.4 Model Diagnostics

To check the fit of the models, we follow the advice in Deng et al. (2013) and use posterior predictive diagnostics (Meng, 1994; Gelman et al., 2005; He et al., 2010; Burgette and Reiter, 2010). Using every 100th draw from the posterior distributions of the parameters in (3) through (7), we generate $T = 500$ datasets with new values of (Y_1, Y_2, R, W_1, W_2) , holding \mathbf{X} constant. Let $\{D^{(1)}, \dots, D^{(T)}\}$ be the collection of the T replicated datasets. We then compare statistics of interest in $\{D^{(1)}, \dots, D^{(T)}\}$ to those in the observed data D . Specifically, suppose that S is some statistic of interest, such as a marginal or conditional probability in our context. For $t = 1, \dots, T$, let $S_{D^{(t)}}$ be the values of S computed from $D^{(t)}$, and let S_D be the value of S computed from D .

Dependent Variable	Y_1	Y_2	R
Intercept	-1.72 (-2.02, -1.42)	-.53 (-1.08, .07)	.13 (-.33, .79)
AGE1	.31 (.00, .63)	.15 (-.17, .46)	.21 (-.03, .46)
AGE2	.87 (.57, 1.18)	.42 (.10, .74)	.26 (.00, .52)
AGE3	1.28 (.98, 1.58)	1.18 (.82, 1.53)	.33 (-.02, .67)
COLLEGE	.12 (-.06, .30)	.74 (.50, .97)	.48 (.23, .72)
MALE	-.06 (-.23, .11)	-.02 (-.22, .18)	.11 (-.04, .28)
BLACK	.63 (.33, .93)	.10 (-.30, .52)	-.25 (-.54, .03)
Y_1	–	2.16 (1.82, 2.51)	.16 (-.35, .56)
Y_2	–	–	.00 (-1.20, 1.18)

Table 6: Coefficient estimates and 95% posterior intervals for AN model with informative prior distributions on sensitivity parameters.

We compute the two-sided posterior predictive probability,

$$ppp = (2/T) * \min\left(\sum_{t=1}^T I(S_{D^{(t)}} - S_D > 0), \sum_{t=1}^T I(S_{D^{(t)}} - S_D < 0)\right). \quad (40)$$

Small values of ppp , for example, less than 5%, suggest that the replicated datasets are systematically different from the observed dataset, with respect to that statistic. When the value of ppp is not small, the imputation model generates data that look like the completed data. As statistics S , we use the following quantities: (i) the percentage of cases with $W_1 = 1$, (ii) among all cases with $W_1 = 1$, the percentage of cases with $Y_1 = 1$, (iii) the percentage of cases with $W_2 = 1$, (iv) among all cases with $W_2 = 1$, the percentage of cases with $Y_2 = 1$, and (v) the MLEs of the coefficients in the logistic regression of Y_2 on (Y_1, \mathbf{X}) , computed with the cases with $W_1 = R = 1$. Each of these quantities is a function of replicated or observed data. We calculate ppp values for each S for all settings of the sensitivity parameters in Table 3 plus $(0, \log 1.5)$ and $(0, \log .75)$.

The ppp values do not suggest obvious problems with model fit. Across all eleven settings, the smallest value of ppp is 0.43. About 40% of the ppp values are at least 0.90. In other words, the proposed AN model generates replicated data that look like the original data, so that the model does not appear to be implausible. We note that one could examine other diagnostics such as the partial posterior predictive p-value (Bayarri and Berger, 2000, 1998), as additional checks on the fit of the model.

6. CONCLUDING REMARKS

Most applications of the AN model have ignored nonresponse in the panel and refreshment sample entirely, effectively treating it as missing completely at random. As we have demonstrated, this assumption when unreasonable can lead to substantial bias in estimates. The sensitivity analyses presented here can help analysts investigate how much inferences change with different assumptions about nonignorable nonresponse in either or both of these samples. Similar sensitivity analyses can be used to handle nonignorable dropout when there are multiple waves between the original panel and refreshment sample. To select values of the sensitivity parameters, analysts should consider the extent of selection bias that seems possible for their particular application. For example, analysts can use similar values of ρ_1 and γ_1 when they expect similar reasons explain unit nonresponse in both the initial wave and refreshment sample. Analysts might consider quite different values when this is not expected, for example if the two surveys have different incentive structures. After specifying ranges of interest, analysts can investigate various combinations within the ranges to identify regions for which inferences of interest do not (and do) change meaningfully.

Conceptually, this approach to sensitivity analysis could extend to panels with more waves than considered here. The number of parameters to estimate and the number of sensitivity parameters could become cumbersome, particularly with large dimensions. It may be necessary to use computationally expedient models for the underlying survey data, such as latent class models (Dunson and Xing, 2009; Kuniyama and Dunson, 2013), as proposed and implemented by Si et al. (2014) for two wave surveys with many categorical variables. It also may be necessary to use simplifying assumptions about the sensitivity parameters, for example, set sensitivity parameters equal across waves. Another possibility is to characterize the survey variables with some low rank representation, and perform sensitivity analysis using that representation. Developing sensitivity analysis for surveys with many waves and refreshment samples is an area for future research.

7. ACKNOWLEDGMENTS

This research was supported by grants from the National Science Foundation [NSF-SES-10-61241, NSF-SES-11-31897]. We thank Chase Harrison and Michael Henderson for providing guidance on response rate disparities among politically interested respondents.

8. SUPPLEMENTARY MATERIALS

Supplementary materials for this article may be found online at http://www.oxfordjournals.org/our_journals/jssam. These include (i) description and results of the simulation with missing \mathbf{X} for unit nonrespondents, and (ii) description and results of the simulation with mismatch between the data generation model and the estimation model.

REFERENCES

- Baltagi, B. H. and Song, S. H. (2006). “Unbalanced panel data: A survey.” *Statistical Papers*, 47(4): 493–523.
- Bartels, L. M. (1999). “Panel effects in the American national election studies.” *Political Analysis*, 8(1): 1–20.
- Bayarri, M. J. and Berger, J. O. (1998). “Quantifying Surprise in the Data and Model Verification.” *Bayesian Statistics*, 6: 53 – 82.
- (2000). “P Values for Composite Null Models.” *Journal of the American Statistical Association*, 95(452): 1127–1142.
- Behr, A., Bellgardt, E., and Rendtel, U. (2005). “Extent and determinants of panel attrition in the European Community Household Panel.” *European Sociological Review*, 21(5): 489–512.
- Bhattacharya, D. (2008). “Inference in panel data models under attrition caused by unobservables.” *Journal of Econometrics*, 144: 430–446.
- Buckley, J. and Schneider, M. (2006). “Are Charter school parents more satisfied with schools? Evidence from Washington, DC.” *Peabody Journal of Education*, 81: 57–78.
- Burgette, L. F. and Reiter, J. P. (2010). “Multiple imputation via sequential regression trees.” *American Journal of Epidemiology*, 172: 1070–1076.
- Das, M., Toepoel, V., and van Soest, A. (2011). “Nonparametric Tests of Panel Conditioning and Attrition Bias in Panel Surveys.” *Sociological Methods and Research*, 40(1): 32–56.
- Deng, Y., Hillygus, D. S., Reiter, J. P., Si, Y., and Zheng, S. (2013). “Handling attrition in longitudinal studies: The case for refreshment samples.” *Statistical Science*, 28: 238 – 256.

- Dunson, D. B. and Xing, C. (2009). “Nonparametric Bayes modeling of multivariate categorical data.” *Journal of the American Statistical Association*, 104(487): 1042–1051.
- Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2004). *Bayesian Data Analysis*. Chapman and Hall/CRC, second edition.
- Gelman, A. and Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press.
- Gelman, A., Jakulin, A., Pittau, M. G., and Su, Y.-S. (2008). “A weakly informative default prior distribution for logistic and other regression models.” *The Annals of Applied Statistics*, 2(4): 1360–1383.
- Gelman, A., Van Mechelen, I., Verbeke, G., Heitjan, D. F., and Meulders, M. (2005). “Multiple imputation for model checking: Completed-Data plots with missing and latent data.” *Biometrics*, 61: 74–85.
- Goyder, J. (1987). *The silent minority: Nonrespondents on sample surveys*. Cambridge: Polity Press.
- Groves, R. M., Dillman, D., Eltinge, J. L., and Little, R. J. (2002). *Survey nonresponse*. New York: Wiley.
- Groves, R. M., Presser, S., and Dipko, S. (2004). “The role of topic interest in survey participation decisions.” *Public Opinion Quarterly*, 68(1): 2–31.
- Groves, R. M., Singer, E., and Corning, A. (2000). “Leverage-saliency theory of survey participation: description and an illustration.” *The Public Opinion Quarterly*, 64(3): 299–308.
- He, Y., Zaslavsky, A. M., and Landrum, M. B. (2010). “Multiple imputation in a large-scale complex survey: A guide.” *Statistical Methods in Medical Research*, 19: 653–670.
- Hillygus, D. S. (2005). “Campaign effects and the dynamics of turnout intention in election 2000.” *Journal of Politics*, 67(1): 50–68.
- Hirano, K., Imbens, G., Ridder, G., and Rubin, D. (2001). “Combining panel data sets with attrition and refreshment samples.” *Econometrica*, 69: 1645–1659.

- Hirano, K., Imbens, G. W., Ridder, G., and Rubin, D. B. (1998). “Combining panel data sets with attrition and refreshment samples.” Technical Report 230, National Bureau of Economic Research.
- Hogan, J. W. and Daniels, M. J. (2008). *Missing Data in Longitudinal Studies*. Boca Raton: Chapman and Hall.
- Kunihama, T. and Dunson, D. B. (2013). “Bayesian modeling of temporal dependence in large sparse contingency tables.” *Journal of the American Statistical Association*, 108(504): 1324–1338.
- Little, R. and Rubin, D. (2002). *Statistical Analysis with Missing Data*. Hoboken, N.J.: Wiley-Interscience.
- Liu, C. (2004). “Robit Regression: A Simple Robust Alternative to Logistic and Probit Regression.” In Gelman, A. and Meng, X.-L. (eds.), *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives*. John Wiley and Sons, Ltd.
- Lynn, P. (2013). “Alternative sequential mixed-mode designs: effects on attrition rates, attrition bias, and costs.” *Journal of Survey Statistics and Methodology*, 1(2): 183–205.
- Meng, X. (1994). “Posterior predictive p-values.” *Annals of Statistics*, 22: 1142–1160.
- Molenberghs, G., Beunckens, C., Sotito, C., and Kenward, M. G. (2008). “Every missingness not at random model has a missingness at random counterpart with equal fit.” *Journal of the Royal Statistical Society, Series B*, 70: 371–388.
- Molenberghs, G., Goetghebeur, E. J. T., Lipsitz, S. R., and Kenward, M. G. (1999). “Nonrandom Missingness in Categorical Data: Strengths and Limitations.” *The American Statistician*, 53(2): 110–118.
- Molenberghs, G., Kenward, M. G., and Goetghebeur, E. (2001). “Sensitivity Analysis for Incomplete Contingency Tables: The Slovenian Plebiscite Case.” *Journal of the Royal Statistical Society, Series C*, 50(1): 15 – 29.

- Nevo, A. (2003). “Using weights to adjust for sample selection when auxiliary information is available.” *Journal of Business and Economic Statistics*, 21: 43–52.
- Olsen, R. J. (2005). “The problem of respondent attrition: Survey methodology is key.” *Monthly Labor Review*, 128: 63–71.
- Peytchev, A. (2013). “Consequences of survey nonresponse.” *The ANNALS of the American Academy of Political and Social Science*, 645(1): 88–111.
- Prior, M. (2010). “Youve either got it or you dont? The stability of political interest over the life cycle.” *The Journal of Politics*, 72(03): 747–766.
- Rosenbaum, P. R. (2010). *Observational Studies*. New York: Springer.
- Rubin, D. B. (1981). “The Bayesian Bootstrap.” *The Annals of Statistics*, 9(1): 130–134.
- (1987). *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley & Sons.
- Schwartz, S. L., Li, F., and Reiter, J. P. (2012). “Sensitivity analysis for unmeasured confounding in principal stratification.” *Statistics in Medicine*, 31: 949–962.
- Si, Y., Reiter, J. P., and Hillygus, D. S. (2014). “Semi-parametric Selection Models for Potentially Non-ignorable Attrition in Panel Studies with Refreshment Samples.” *Political Analysis*.
- Singer, E. (2006). “Introduction Nonresponse Bias in Household Surveys.” *Public Opinion Quarterly*, 70(5): 637–645.
- Traugott, M. W. and Tucker, C. (1984). “Strategies for predicting whether a citizen will vote and estimation of electoral outcomes.” *Public Opinion Quarterly*, 48(1B): 330–343.
- Western, B. (2002). “The impact of incarceration on wage mobility and inequality.” *American Sociological Review*, 67(4): 526–546.
- Zhou, X. and Reiter, J. P. (2010). “A note on Bayesian inference after multiple imputation.” *The American Statistician*, 64: 159–163.