

¹Departments of Political Science and Statistics, University of California, Berkeley, Berkeley, CA, USA.

E-mail: ekhartman@berkeley.edu

²Department of Statistics, University of California, Berkeley, Berkeley, CA, USA. E-mail: melodyhuang@berkeley.edu

Abstract

Survey weighting allows researchers to account for bias in survey samples, due to unit nonresponse or convenience sampling, using measured demographic covariates. Unfortunately, in practice, it is impossible to know whether the estimated survey weights are sufficient to alleviate concerns about bias due to unobserved confounders or incorrect functional forms used in weighting. In the following paper, we propose two sensitivity analyses for the exclusion of important covariates: (1) a sensitivity analysis for partially observed confounders (i.e., variables measured across the survey sample, but not the target population) and (2) a sensitivity analysis for fully unobserved confounders (i.e., variables not measured in either the survey or the target population). We provide graphical and numerical summaries of the potential bias that arises from such confounders, and introduce a benchmarking approach that allows researchers to quantitatively reason about the sensitivity of their results. We demonstrate our proposed sensitivity analyses using state-level 2020 U.S. Presidential Election polls.

Keywords: survey weighting, sensitivity analysis

1 Introduction

Given the fall of survey response rates and the increased reliance on online convenience samples, concerns about bias in surveys are rising. For example, Bradley *et al.* (2021) discuss bias in public opinion surveys used to estimate COVID-19 vaccination rates. Modern surveys nearly always rely on weighting adjustments, and researchers must carefully examine how survey weights are constructed. For example, Kennedy *et al.* (2018) found that a primary driver of bias in the 2016 U.S. Presidential Election was survey weights that failed to account for either education or the interaction of education with race. Subsequently, many surveys began including these when constructing survey weights. Despite this, the 2020 U.S. Presidential Election cycle resulted in national-level public opinion polling that exhibited some of the most bias the last 40 years. Clinton *et al.* (2021) found that the errors of 2016 do not explain problems in the 2020 election polls. Instead, they find that bias is likely driven by unobservable confounders.

While these polling misses provide an opportunity to retrospectively evaluate these issues, they also make clear that within any specific survey, the question of how to properly construct survey weights is an open one. The purpose of our paper is to provide a set of sensitivity analyses that researchers can use on weighted surveys to analyze sensitivity to two types of confounders: (1) partially observed confounders, which are measured in the survey sample but not the target population, and (2) fully unobserved confounders which are not measured in the survey sample or the target population. While these variables cannot be directly incorporated in the construction of weights, researchers are often aware of their existence through theoretical concerns.

In summary, we decompose bias in weighted estimators into two observable components—variation in the outcome and variation in the estimated weights—and two unobservable components—the correlation of the error in the estimated weights with the outcome, and how much variation in the ideal (true) weights this error explains. We then propose two sensitivity analyses. For partially observed confounders, our sensitivity analysis is based on a posited distribution of the confounder in the target population. For fully observed confounders, we

Political Analysis (2024)

vol. 32: 1–16

DOI: [10.1017/pan.2023.12](https://doi.org/10.1017/pan.2023.12)

Published

28 April 2023

Corresponding author

Erin Hartman

Edited by

Jeff Gill

© The Author(s), 2023. Published by Cambridge University Press on behalf of the Society for Political Methodology. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

propose a two-parameter sensitivity analysis based on the unobserved components of our bias decomposition, and provide graphical and numerical summaries of robustness.

We extend the sensitivity analyses developed by Hong, Yang, and Qin (2021), Huang (2022), and Shen *et al.* (2011), which assess the sensitivity in estimating causal effects with weighted estimators. Alternative approaches include sensitivity analyses which bound the worst-case bias from an unobserved confounder for matching estimators (e.g., Rosenbaum and Rubin 1983) and marginal sensitivity models that consider a multiplicative error term in weighted estimators (e.g., Soriano *et al.* 2023; Tan 2006; Zhao, Small, and Bhattacharya 2019); and approaches that rely on invoking parametric or distributional assumptions about the underlying data generating processes (e.g., Nguyen *et al.* 2017). An advantage of our framework is its generality; we do not require parametric assumptions on the data generating process for either the outcome or for sample selection. Our method is applicable to nonnegative calibration weights and inverse propensity weights (IPWs), making it relevant for many researchers.

1.1 Running Example: 2020 U.S. Presidential Election

We consider a retrospective analysis of the 2020 U.S. Presidential Election to demonstrate sensitivity analyses to partially and fully unobservable confounders. We evaluate two state-level ABC News/Washington Post (ABC/Wapo) polls conducted in October 2020. In particular, we examine polls for Michigan (October 20–25, 2020) and North Carolina (October 12–17, 2020). The “survey samples” consist of 770 respondents (Michigan) and 619 respondents (North Carolina) who reported that they were planning to vote or had already voted.

We want to demonstrate estimation of sensitivity parameters, including benchmarking against observed covariates, which requires we estimate custom weights rather than use the proprietary ones Hartman and Huang (2022). We construct weights calibrated to a target population defined using the 2020 Cooperative Election Study (CES) (Schaffner, Ansolabehere, and Luks 2021), following Caughey *et al.* (2020), which provides rich auxiliary data.¹ Our target population includes verified voters, incorporating the CES weights, `commonweight_vv_post`. We limit to units who stated they “Definitely voted” and for whom the retrospective candidate choice for the 2020 Presidential Election is not missing. This defines our “target population” and provides estimates of candidate support at the state level as a point of reference for our analysis. Using this approach, we have 1,213 units in our “target population” for Michigan, and 1,101 units for North Carolina.

We construct calibration weights by raking, which, in short, ensures the survey sample is representative by matching the weighted survey mean of observed demographic characteristics to known target population means. We include the following covariates in the construction of our survey weights: age, gender, race/ethnicity, educational attainment, party identification, and an indicator for born-again Christian.² For simplicity, we conduct raking only on marginal population means, as well as a two-way interaction between party identification and educational attainment, but alternative methods can account for higher-order interactions (e.g., Ben-Michael, Feller, and Hartman 2023; Hartman, Hazlett, and Sterbenz 2021).

Our primary outcome of interest is the Democratic margin in the two-party vote share (Democrat (D)–Republican (R)) among those who state a preference for either major party candidate. Table 1 presents the unweighted and weighted outcomes. The goal of our method is to determine how sensitive the weighted point estimate is to partially or fully unobserved confounders. As is common in U.S. presidential elections, the election is predicted to be close and the estimates are statistically indistinguishable from zero; however, the point estimate

- 1 In practice, the CES, or any large survey that employs weighting against an administrative file or census used to define the target population, such as the GSS, Afrobarometer, or Eurobarometer, could conduct a sensitivity analysis.
- 2 Whether or not to weight on party identification is an open debate. In Section E of the Supplementary Material, we demonstrate our sensitivity analysis for Michigan without weighting on party.

Table 1. Unweighted and weighted margin (D–R) in two-party vote share in percentage points. We provide the CES estimate and the true vote margin as reference.

| State | Unweighted | Weighted | CES | True election result |
|----------------|-------------|--------------|-------|----------------------|
| Michigan | 7.25 (3.59) | 4.57 (2.56) | 2.06 | 2.78 |
| North Carolina | 0.65 (4.04) | −0.37 (2.63) | −5.10 | −1.35 |

provides the best prediction for the substantive outcome. Thus, we focus on sensitivity in the point estimate. We describe how to incorporate sensitivity in measures of statistical uncertainty in the conclusion. Our sensitivity analysis tools allow researchers to transparently reason about whether their findings—in this case who leads in each state—are robust to the exclusion of confounders when estimating weights.

2 Notation and Setup

We consider a finite target population with N units. The survey sample consisting of n units is drawn from the target population (where $n \ll N$); we assume the survey sample is not a simple random sample, and, as such, is not representative of the target population. Let $S_i \in \{0, 1\}$ be a survey inclusion indicator, where $S_i = 1$ when a unit is a respondent in the survey and 0 otherwise. Let $\Pr(S_i = 1)$ denote the probability of inclusion for unit i .³ Note that for a probability sample without nonresponse, $\Pr(S_i = 1)$ represents the sampling probability; for a convenience sample or a sample with nonresponse, it encodes the product of the sampling probability and response probability for unit i , both of which may be unknown. Throughout the manuscript, we refer to the collection of respondents, or units for which $S_i = 1$, as the “survey sample” (denoted S) and we assume that each unit in the target population has some positive (but possibly unknown) probability of inclusion in the survey sample (i.e., $0 < \Pr(S_i = 1) \leq 1$). We denote quantities calculated over the survey sample with a subscript S (e.g., $\text{var}_S(\cdot)$). Our outcome of interest is Y , which is observed for every unit in the survey sample. Our target estimand is the population mean $\mu = \frac{1}{N} \sum_{i=1}^N Y_i$. See Table A.1 in the Supplementary Material for a glossary of terms.

When the survey sample is a simple random sample from the target population, then the average outcome within the survey sample is an unbiased estimator for the population mean. However, in most cases, the survey suffers from unit nonresponse, or is a convenience sample, resulting in a nonrepresentative sample; as such, using the mean outcome within the survey sample may be a biased estimator of the population mean. We assume that to account for nonrandom selection into the survey sample, researchers construct survey weights that adjust the survey to be representative of the target population on observable characteristics.

One way to construct survey weights is to assume that the observable characteristics \mathbf{X} are sufficient, and conditional on these variables, the distribution of the outcome among the survey respondents and the target population is the same.

ASSUMPTION 1 (Conditional Ignorability of Response (Little and Rubin 2002)) $Y \perp\!\!\!\perp S \mid \mathbf{X}$.

Assumption 1 nonparametrically justifies post-stratification weights constructed by dividing the population proportion by the survey proportion within intersectional strata defined by \mathbf{X} . Common approaches include conditioning on variables that fully explain sampling or the outcome, but many alternatives are possible (Egami and Hartman 2021).

When weighting on continuous variables or a large number of strata, post-stratification suffers from sparsity constraints. In this case, researchers must define a feature mapping $\mathbf{X} \mapsto \phi(\mathbf{X})$ from $\mathbb{R}^P \mapsto \mathbb{R}^{P'}$ that captures important features of \mathbf{X} for use in coarsened post-stratification or

3 While the current setup assumes that the survey sample is a subset of the target population (“nested”), our results extend to a nonnested setup. See Huang (2022) for discussion.

model-based weighting methods. Researchers can construct calibration weights, subject to a set of moment constraints defined on $\phi(\mathbf{X})$. For example, researchers can use “raking” to calibrate the marginal means on all variables \mathbf{X} , letting $\phi(\mathbf{X}) = \mathbf{X}$. Alternatively, IPWs, in which researchers directly estimate probability of inclusion and weight inversely proportional to this estimate, may use $\phi(\mathbf{X}) = \mathbf{X}$, which are included link-linearly in logistic regression. We focus our conceptual discussion around IPWs, which are asymptotically equivalent to calibration weights (Ben-Michael *et al.* 2021b) but use calibration for estimation. In particular, we employ raking on the margins, a type of calibrated IPW, in our analysis. A detailed discussion on the construction of survey weights is beyond the scope of this paper. See Section B.1 of the Supplementary Material for a discussion of calibration weights; we refer readers to Haziza and Beaumont (2017) for a thorough review on how to construct survey weights and related considerations.

How to construct an appropriate feature mapping is an active field of research. For example, recent methods aim to include higher-order interactions of \mathbf{X} (e.g., Ben-Michael *et al.* 2023) or rely on kernel methods to account for important features (e.g., Hartman *et al.* 2021). While these methods are important for flexibly accounting for observable characteristics, our proposed sensitivity analysis also evaluates robustness to unobservable confounders.

When relying on model-based weighting adjustments, such as calibration, researchers may not be able to directly invoke Assumption 1. Instead, they can appeal to a linear ignorability assumption for consistent estimation of the population mean.

ASSUMPTION 2 (Linear Ignorability in $\phi(\mathbf{X})$ (Hartman *et al.* 2021)) Let $Y_i = \phi(\mathbf{X}_i)^\top \beta + \delta_i$ and $Pr(S_i = 1) = g(\phi(\mathbf{X}_i)^\top \theta + \eta_i)$ where $g(\cdot) : \mathcal{R} \mapsto [0, 1]$. Linear ignorability holds when $\delta_i \perp \eta_i$.

Assumption 2 states that the part of Y orthogonal to $\phi(\mathbf{X})$ must be independent of the part of S , the survey selection process, orthogonal to $\phi(\mathbf{X})$ via a suitable link function. See Hartman *et al.* (2021) for more details. What Assumption 2 makes clear is that researchers must carefully choose a feature mapping, $\phi(\mathbf{X})$ to account for all features of \mathbf{X} , including interactions and transformations, that affect both survey inclusion and the outcome.

Assumptions 1 and 2 are related. The advantage of Assumption 2 is that it allows researchers to address sparsity in finite data by imposing a parametric assumption, whereas Assumption 1 nonparametrically identifies the population mean. They differ most starkly in the types of violations that are problematic. Given a set of survey weights constructed using feature mapping $\phi(\cdot)$, we define the weighted estimator as follows:

DEFINITION 2.1 (Weighted Estimator for Population Mean) $\hat{\mu} = \sum_{i \in S} w_i Y_i$.

This estimator could be biased for two reasons: (1) there is an unobserved confounder U (violating Assumptions 1 and 2) or (2) researchers have failed to construct an adequately rich feature expansion $\phi(\mathbf{X})$ of the observable characteristics, such as by failing to incorporate higher-order moments or interactions (violating Assumption 2). These ignorability assumptions are strong and untestable, although some observable implications can be tested. While survey weights might make the sample representative on the features included in estimation, this does not imply that these ignorability assumptions hold. Weighting will typically mitigate bias if the weighting variables are correlated with the outcome, although it could exacerbate bias. As we can never know if Assumption 1 or 2 hold, it is essential we have tools to transparently evaluate their credibility. The focus of our paper is to help researchers evaluate if their results are sensitive to the exclusion of unobservable characteristics or more complex observable features.

3 Bias in Weighted Estimators

In the following section, we introduce the bias of weighted estimators when omitting an unobservable variable U from estimation of the survey weights. More specifically, under Assumption 2,

we assume that Y can be decomposed into $\phi(\mathbf{X})^T \beta + U + v$, where $v \perp\!\!\!\perp \{X, U, S\}$. Implicitly, this means U is included in $\delta = U + v$. Moreover, we assume β and θ are nonzero, indicating there is a correlation between Y and S , weighting is necessary, and the true weights are not all equal to 1. This would be violated, for example, if the survey was a simple random sample with no nonresponse.

The omitted variable setup provides a flexible formulation for assessing violations of Assumption 2. Without loss of generality, we assume U is orthogonal to $\phi(\mathbf{X})$, and could be a combination of multiple underlying variables. U can always be replaced by the residual from projecting confounders onto $\phi(\mathbf{X})$. This is important when reasoning about potential confounders; it makes clear that bias is due to the part of the omitted confounder not linearly explained by the included covariates. For example, if we have omitted political interest, but we have included age and educational attainment in the construction of our weights, U is the part of political interest not linearly explained by age and educational attainment. This indicates that potential confounders well explained by $\phi(\mathbf{X})$ are less problematic. If U is associated with selection into the survey, i.e., correlated with S , then Assumption 2 is violated because $\delta_i \not\perp \eta_i$. The goal of the sensitivity analysis is to assess the robustness of point estimates to the omitted confounder U .

Alternatively, U could be interactions or nonlinear transformations of observable covariates \mathbf{X} not included in $\phi(\mathbf{X})$. For example, Kennedy *et al.* (2018) showed that simply accounting for the linear combination of race/ethnicity, region, and educational attainment is insufficient; instead, researchers also needed to account for an interaction between race/ethnicity, region, and educational attainment.

3.1 Derivation

We begin by defining the vector of estimated weights, w , as those estimated using $\phi(\mathbf{X})$, and the vector of ideal weights w^* as those that would have been estimated including $\phi(\mathbf{X})$ and U . Throughout the paper, we will assume, without loss of generality, that both the estimated and ideal weights are centered at mean 1. Finally, define the error term ε as the difference between the estimated weights and the ideal weights (i.e., $\varepsilon := w - w^*$).

The error in the weights is driven by the *residual* imbalance in an omitted confounder, after balancing on $\phi(\mathbf{X})$. This general intuition holds regardless of the weighting approach: if imbalance in U is minimal after adjusting for $\phi(\mathbf{X})$, the overall error from omitting U when constructing weights should be low. In Section C.1 of the Supplementary Material, we derive the error for IPWs, highlighting where imbalance factors into the error. In the context of our running example, an individual’s baseline level of political interest is an important predictor of both survey inclusion and many political outcomes; however, it is not typically incorporated into weight estimation because there are no population measures. If political interest is largely explained by variables included in weight estimation, such as age and educational attainment, then the residual imbalance in political interest should be small, and ε_i should be relatively small.

The bias in $\hat{\mu}$ from omitting the variable U from estimation of the survey weights can be parameterized with respect to the error term, ε . We formalize this in the following theorem.

THEOREM 3.1 (Bias of a Weighted Estimator) *Let w be the weights estimated using just $\phi(\mathbf{X})$, and let w^* be the ideal weights estimated using $\phi(\mathbf{X})$ and U . The bias in $\hat{\mu}$ from omitting U from estimation of the weights can be written as*

$$\begin{aligned} \text{Bias}(\hat{\mu}) &= \mathbb{E}(\hat{\mu}) - \mu \\ &= \begin{cases} \rho_{\varepsilon, Y} \sqrt{\text{var}_S(Y) \cdot \text{var}_S(w) \cdot \frac{R_\varepsilon^2}{1 - R_\varepsilon^2}}, & \text{if } R_\varepsilon^2 < 1, \\ \rho_{\varepsilon, Y} \sqrt{\text{var}_S(Y) \cdot \text{var}_S(w^*)}, & \text{if } R_\varepsilon^2 = 1, \end{cases} \end{aligned} \tag{1}$$

where ε is the error in the weights from omitting U , R_ε^2 is the ratio of variation in the ideal weights explained by ε (i.e., $R_\varepsilon^2 := \text{var}_S(\varepsilon)/\text{var}_S(w^*)$), and $\rho_{\varepsilon,Y}$ is the correlation, or alignment, between ε and the outcome Y (i.e., $\rho_{\varepsilon,Y} := \text{cor}_S(\varepsilon, Y)$). Quantities denoted with a subscript S are estimated over the survey sample (i.e., $S_i = 1$). Proof is given in Section C.2 of the Supplementary Materials.

Theorem 3.1 decomposes the sources of bias from omitting a confounder in weight estimation. The terms highlighted in blue are unobserved, whereas the other terms in black can be directly estimated from the survey data using sample analogs. By treating the weights as fixed, we implicitly derive the asymptotic bias. See Section C.2 of the Supplementary Material for a discussion of the finite-sample case. We provide more details about the interpretation and properties of each component of the bias formula in the following subsection.

3.2 Interpreting the Drivers of Bias

In this subsection, we discuss the different components in the bias formula from Theorem 3.1: (1) R_ε^2 , representing the ratio of the variance in the error ε and the variance in the ideal weights; (2) $\rho_{\varepsilon,Y}$, the alignment between the error in the weights, ε , and the outcome Y ; and (3) a scaling factor (i.e., $\text{var}_S(Y) \cdot \text{var}_S(w)$). The first two components are unobservable, whereas the last component is directly estimable from the observed data. We show that both of the unobserved components exist on bounded, standardized ranges, and provide intuition for how to interpret each component.

3.2.1 *Explained variation in ideal weights (R_ε^2). R_ε^2 represents the amount of variation in the ideal survey weights w^* that is explained by the error term ε . Following Huang (2022), we decompose the total variation in the ideal weights into two components: (1) the amount of variation in w^* explained by the estimated weights w and (2) the amount of variation in the ideal weights w^* explained by the estimation error ε :*

COROLLARY 3.1 (Variance Decomposition of w^* (Huang (2022))) *Let w be the estimated IPWs, and let w^* be the ideal weights. The variance of the ideal weights w^* can be decomposed linearly into two components:*

$$\text{var}_S(w^*) = \text{var}_S(w) + \text{var}_S(\varepsilon) \implies \frac{\text{var}_S(w)}{\text{var}_S(w^*)} + \underbrace{\frac{\text{var}_S(\varepsilon)}{\text{var}_S(w^*)}}_{:=R_\varepsilon^2} = 1.$$

An implication of Corollary 3.1 is that R_ε^2 is guaranteed to be bounded on the interval $[0, 1]$.

As the residual imbalance in the omitted confounder U increases, R_ε^2 increases. Intuitively, if imbalance in U across the target population and survey sample is large, the ideal weights w^* will be very different from the estimated weights w . As a result, the variance in the estimation error, $\text{var}_S(\varepsilon)$, will be large, thus increasing R_ε^2 . In contrast, if residual balance in U is small, the error will be small and thus the ideal weights w^* will be close to the estimated weights, leading to small R_ε^2 .⁴

Consider our running example, in which retrospective studies showed that voters with lower levels of educational attainment were underrepresented in the survey sample, relative to the target population. State polls, particularly in the Midwest, that omitted educational attainment, despite accounting for other demographic variables, saw significant error in point estimates. If educational attainment is not well explained by the other demographic characteristics, this will

4 In the scenario that $R_\varepsilon^2 = 1$ (i.e., the error from omitting the confounder can explain 100% of the variation in the ideal weights w^*), researchers will need to posit values for $\text{var}_S(w^*)$ to estimate the bias. This occurs, for example, with no weighting adjustment and unity weights. However, we argue that if researchers have conducted weighting and accounted for even a single covariate that can explain any variation in the selection process, $R_\varepsilon^2 < 1$.

lead to a large R_ε^2 . The bias is exacerbated by the correlation between this error and the outcome, since white voters with lower educational attainment were more likely to support Donald Trump in the 2020 U.S. Presidential Election.

- 3.2.2 *Alignment between the Error and Outcome ($\rho_{\varepsilon,Y}$)*. The alignment⁵ between the error in the weights and the outcome, captured by the correlation $\rho_{\varepsilon,Y}$, also affects the potential for bias. For example, if positive ε values correspond to large Y values, then this implies that units that are being overrepresented also have a larger Y value, resulting in positive bias. Conversely, if negative ε values correspond to large Y values, this implies that units with a large Y value are underrepresented, resulting in negative bias.

In our running example, assume that the survey sample has higher average baseline political interest than the target population and also that, conditioning on our weighting variables, individuals with greater political interest are more likely to vote for a Democratic candidate. Omitting political interest from weight estimation would lead to larger weights placed on Democratic voters in the survey, resulting in a bias that overstates Democratic vote share.

- 3.2.3 *Bias Scaling Factor*. The final terms in Equation (1), $\text{var}_S(w)$ and $\text{var}_S(Y)$, are not dependent on the unmeasured confounder U and are directly estimable from the survey data. They act as a scaling factor that increases the sensitivity of the point estimate to violations of the assumptions, even if the alignment $\rho_{\varepsilon,Y}$ and the variance explained by the error R_ε^2 are close to zero.

It is well known that if $\text{var}_S(w)$ is large, such as from extreme weights, this increases the variance in the estimator. This is related to the design effect (Kish 1965), and underscores the importance of choosing prognostic and substantively meaningful variables to weight on as well as design-stage considerations that minimize the variance in the weights. This also emphasizes the bias-variance trade-off when constructing survey weights.

The term $\text{var}_S(Y)$ is the variance of the outcome variable Y , related to the “problem difficulty” in Meng (2018), which is not something that researchers can control. It formalizes the intuition that if the outcome variable Y has a high degree of heterogeneity, it is potentially more sensitive to sources of bias.

The bias decomposition highlights an important point: an omitted variable must be related to both the outcome *and* the response process in order for there to be bias from omitting it. Even if the omitted variable is imbalanced between the survey sample and the target population (i.e., $R_\varepsilon^2 > 0$), if it is not related to the outcome (i.e., $\rho_{\varepsilon,Y} = 0$), there will be no bias. Similarly, if an omitted variable is related to the outcome (i.e., $|\rho_{\varepsilon,Y}| > 0$), but is balanced between the survey sample and the target population (i.e., $R_\varepsilon^2 = 0$), no bias will occur. This framework helps formalize the types of variables that researchers should consider when assessing the sensitivity in their estimates.

3.3 Performing Sensitivity Analyses

The bias decomposition in Theorem 3.1 provides a natural basis for performing a sensitivity analysis. By positing values for the unobserved parameters, R_ε^2 and $\rho_{\varepsilon,Y}$, researchers can estimate bias and evaluate the robustness to residual confounding. In the following sections, we propose two sensitivity analyses. The first sensitivity analysis (Section 4) shows that when a confounder is observed across the survey sample but not the target population, Theorem 3.1 can be rewritten as a function of a single unobserved parameter. The second sensitivity analysis (Section 5) allows researchers to assess the sensitivity to fully unobserved confounders using a two-parameter sensitivity analysis. We propose a set of tools that allow researchers to (1) summarize the amount of sensitivity in their point estimate and (2) benchmark the analysis using observed covariate data.

5 We use the term “alignment” based on a similar concept in Kern *et al.* (2016).

4 Sensitivity Analysis for Partially Observed Confounders

Researchers typically have greater control over the variables they can measure among survey respondents, and they may have strong theoretical reasons to believe that a variable is related to both sampling and the outcome; however, if the covariate is not measured across the target population, it cannot be incorporated into the weights, and thus it should be assessed in sensitivity analyses. We formalize this issue by defining a variable measured in the survey sample, but not in the target population, as a *partially observed confounder*, denoted by V . This extends to excluded functions of observables, for example, higher-order moments or interactions of the observable covariates.

In this section, we propose a sensitivity analysis that evaluates robustness to partially observed confounders against a hypothetical distribution of this confounder in the target population. These confounders can be identified using theory, but in Section D.2 of the Supplementary Material, we suggest a data-driven approach for detecting such confounders. This is useful in settings when researchers are unsure from a purely substantive standpoint whether or not a partially observed variable must be included in the weights. A summary of our suggested approach for detecting and evaluating sensitivity to partially unobserved confounders is provided in Figure 1.

Recall that the error in the weights is driven by the residual imbalance in an omitted confounder (see Section C.1 of the Supplementary Material for details). With information about V across the survey sample, the sensitivity analysis can be reduced to one sensitivity parameter: the distribution of V in the target population. We focus here on how to incorporate such a parameter in calibration weighting, and discuss additional details for IPWs in Section D.1 of the Supplementary Material.

With calibration weighting, raking on the margins, a sensitivity analysis for partial confounding reduces to including an additional moment constraint for $\mathbb{E}(V)$, or the posited population mean, across a range of plausible values for the population mean. Calibration will solve for weights that will simultaneously meet the original and the additional moment constraints, even without knowing the joint distribution of the original covariates with the partially observed covariate. The resulting error in the weights can be directly calculated and used to estimate bias. When V is binary, $\mathbb{E}(V)$ is bounded on $[0, 1]$; this is easy to extend to categorical variables with three levels, but difficult to visually evaluate beyond that. A limitation to this approach is that when $\mathbb{E}(V)$ is unbounded, researchers must specify a theoretically relevant range for $\mathbb{E}(V)$, informed through substantive knowledge or existing data on $\mathbb{E}(V)$ from similar target populations. If researchers do not have a strong substantive prior for a reasonable range of $\mathbb{E}(V)$, it may be helpful to redefine $\mathbb{E}(V)$ in terms of standard deviations from the sample average (see Nguyen *et al.* (2017) for more discussion). This naturally extends when raking across additional moments or interactions, although it requires the researcher to specify more parameters.

4.1 Running Example: 2020 U.S. Presidential Election

Existing literature indicates that political participation is also correlated with propensity to respond to surveys (Peress 2010); however, it is not commonly incorporated into survey weights because it is not available in target populations data. To illustrate our proposed sensitivity analysis

- Step 1.** Use theory or data-driven method to detect V (Egami and Hartman 2021), see Supplementary Materials D.2.
- Step 2.** Posit values for the population average of V (i.e., $\mathbb{E}(V_i | \mathbf{X}_i)$ for propensity score weights, or $\mathbb{E}(V)$ for calibration weights) and re-estimate the survey weights.
- Step 3.** Use the re-estimated weights to obtain an updated $\hat{\mu}$.

Figure 1. Sensitivity analysis for partially observed confounders.

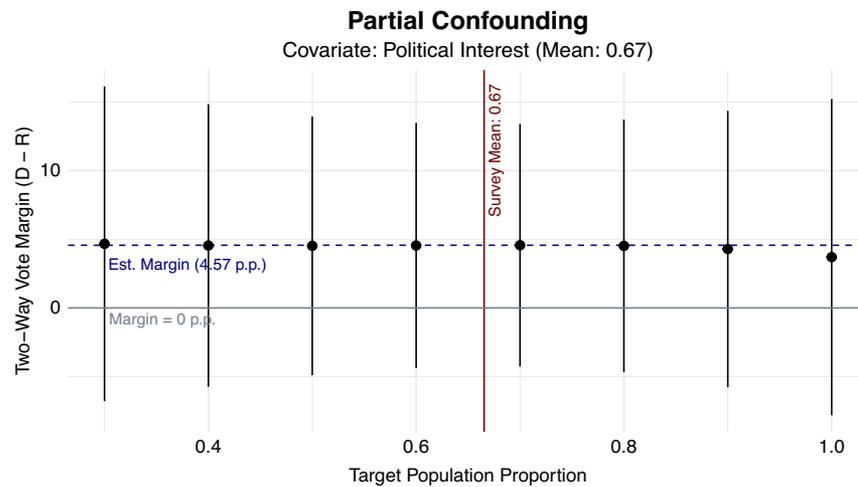


Figure 2. Plot of estimates for two-way vote margin (D–R) as the proportion of politically interested individuals in the target population changes.

for partially observed confounders, we focus on the Michigan poll. We use interest in the upcoming election as a proxy for political participation, and posit it is a partially observed confounder. Below, to visually demonstrate sensitivity analysis for a partially observed confounder, we recode this as a binary variable.⁶

In the survey, 67% of the respondents are encoded as “very closely” following the 2020 election, using the estimated weights. The sensitivity analysis varies the proportion of individuals in the target population who are politically interested, from 0% to 100%, the natural range of the variable, and re-estimates the weights and evaluates how the point estimate changes. (See Figure 2 for visualization.) The range could be reduced with a strong substantive argument or external data. We see that the point estimate is very insensitive to excluding political interest, moving very little across the entire range of the sensitivity parameter, and that the substantive result (i.e., Democratic two-way vote share is greater than 0) is robust to omitting this partially observed confounder. Confidence intervals include 0 across the full range of the sensitivity parameter. We argue that the analysis is very insensitive to this partially observed confounder, and researchers need not include it in the weights in this survey. In Section D.2.1 of the Supplementary Material, we confirm that this political interest variable is, in fact, not a partially observed confounder using our algorithm for detecting such confounders; political interest is rendered irrelevant to the outcome using only fully observed covariates within the survey sample.

5 Sensitivity Analysis for Fully Unobserved Confounders

In this section, we introduce a sensitivity analysis for fully unobserved confounders. The method uses Theorem 3.1 as the foundation for a two-parameter sensitivity analysis. Researchers posit values for both the alignment, $\rho_{\epsilon, Y}$, and the variation in the ideal weights explained by the error, R_{ϵ}^2 , to evaluate potential bias in $\hat{\mu}$. We summarize our approach in Figure 3.

To help researchers conduct the sensitivity analysis, we propose three methods that allow researchers to (1) summarize the degree of robustness in their point estimate using a single “robustness value,” (2) graphically evaluate robustness to fully unobserved confounders with bias contour plots, and (3) benchmark potential bias using observed covariates.

⁶ We define political interest “1” for individuals who responded that they were following the 2020 election “very closely,” and “0” otherwise.

Step 1. Using the observed survey sample data, estimate the variance of the outcome Y and the estimated weights w (i.e., estimate $\text{var}_S(Y)$ and $\text{var}_S(w)$).

Step 2. Vary $\rho_{\varepsilon,Y}$ across the range $[-1, 1]$.

Step 3. Vary R_ε^2 across the range $[0, 1]$.

Step 4. Evaluate the bias of $\hat{\mu}$ for each $\rho_{\varepsilon,Y}$ and R_ε^2 value using Equation 1.

Step 5. Interpret bias using the robustness value RV_{b^*} , contour plots, and benchmarking against observables.

Figure 3. Summary: sensitivity analysis for unobserved confounders.

5.1 Numerical Summary: Robustness Value

We propose a standardized numerical summary of sensitivity, in the form of a “robustness value,” RV_{b^*} , to allow researchers to succinctly assess the plausible existence of fully unobserved confounders. Researchers must first specify a substantively meaningful threshold b^* at which bias would change the substantive conclusion. The robustness value then represents the minimum amount of variation that the error term ε must explain in both the ideal weights w^* and the outcome Y in order for the bias to be large enough to change the substantive result.⁷ Following Cinelli and Hazlett (2020) and Huang (2022), the robustness value is estimated as follows:

$$RV_{b^*} = \frac{1}{2} \left(\sqrt{a_{b^*}^2 + 4a_{b^*}} - a_{b^*} \right), \text{ where } a_{b^*} = \frac{(\hat{\mu} - b^*)^2}{\text{var}_S(Y) \cdot \text{var}_S(w)}. \tag{2}$$

Researchers can pick different target values, b^* , that are substantively meaningful; for example, in the context of U.S. election polling, if researchers are interested in estimating a candidate’s vote share, a logical value is $b^* = 0.5$, which represents the threshold for where candidate’s vote share changes past the 50% threshold, thus changing the prediction that the candidate will win or lose. Similarly, if the outcome is vote margin, a logical value is $b^* = 0$, which indicates where the predicted winning candidate would change. Other natural b^* also include substantively meaningful deviations from the point estimate, such as a 20% difference. In the conclusion, we discuss how to additionally incorporate uncertainty in the estimation of weights using a percentile bootstrap to determine the bias necessary to change the statistical significance of the results (Huang and Pimentel 2022).

RV_{b^*} is bounded on an interval from 0 to 1. When RV_{b^*} is close to 1, this implies that the error in the weights must explain close to 100% of the variation in both w^* and Y in order for the bias to substantively change the point estimate. On the other hand, when RV_{b^*} is close to 0, then if the error in the weights can account for even a small amount of variation in w^* and Y , the resulting bias will be large enough to alter the substantive result of the point estimate.

5.1.1 Running Example: 2020 U.S. Presidential Election. Recall from Section 1.1 that the point estimate for our weighted ABC/Wapo poll projects Biden to win the popular vote by a 4.57 p.p. margin (± 2.56 p.p.) in Michigan, and Biden to lose the popular vote by a -0.37 p.p. margin (± 2.63 p.p.) in North Carolina. We let $b^* = 0$, indicating that we are interested in bias that would change the predicted winning candidate. For the Michigan poll, $RV_{b^*=0} = 0.11$; the error in the weights needs to explain 11% of the variation in both the outcome and the ideal weights to reduce the estimated vote margin to zero. Researchers need to substantively defend if this is plausible. For the North Carolina poll, $RV_{b^*=0} = 0.01$; the error only needs to explain 1% of the variation in both the outcome and the ideal weights to reduce the estimated vote margin to zero. As such, we conclude that there is a greater

⁷ When $R_\varepsilon^2 = \rho_{\varepsilon,Y}^2 = RV_{b^*}$, then the bias will equal $\hat{\mu} - b^*$.

Table 2. Point estimate and robustness value for ABC/Wapo 2020 U.S. Presidential Election poll.

| | Point estimate | Standard error | $R_{b^*=0}$ |
|----------------|----------------|----------------|-------------|
| Michigan | 4.57 p.p. | 2.56 p.p. | 0.11 |
| North Carolina | -0.37 p.p. | 2.63 p.p. | 0.01 |

degree of sensitivity in the North Carolina poll to an omitted variable substantively altering the predicted winner.

5.2 Graphical Summary: Bias Contour Plots

While the robustness value is a useful summary measure, R_{b^*} only represents a single point of the combination of $\{\rho_{\epsilon, \gamma}, R_{\epsilon}^2\}$ that could lead to substantively meaningful bias. However, the estimation error may not equally explain the alignment and the variation in the ideal weights.

To provide a fuller understanding of how bias may vary across different $\{\rho_{\epsilon, \gamma}, R_{\epsilon}^2\}$ values, we propose the use of bias contour plots. To construct the bias contour plots, researchers evaluate the bias at values of $\rho_{\epsilon, \gamma}$ in the range $[-1, 1]$ on the x -axis and R_{ϵ}^2 in the range $[0, 1)$ on the y -axis. The bias is calculated using Equation (1). This approach fully captures potential bias across the range of both of the sensitivity parameters.

The contour plots allow researchers to visualize the “killer confounder” region, which represents the values of $\{\rho_{\epsilon, \gamma}, R_{\epsilon}^2\}$ for which the bias is large enough to substantively change the meaning of the point estimate. The boundary of the killer confounder region is defined by b^* (i.e., the same threshold value chosen for the robustness value); R_{b^*} is one point on the boundary of the killer confounder region, where $\rho_{\epsilon, \gamma} = R_{\epsilon}^2$. As such, it is important to report bias contour plots in order to assess the full set of possible parameter values that may result in a killer confounder.

If the area of the killer confounder region dominates much of the plot, then the point estimate is very sensitive to fully unobserved confounders; in contrast, if the area is relatively small and contained to regions that are defensible as unlikely, then the result is robust.

5.2.1 Running Example: 2020 U.S. Presidential Election. Figure 4 presents the bias contour plot for the ABC/Wapo polls for Michigan and North Carolina. The killer confounder region, shaded in blue, represents the sensitivity values where the estimated margin (D–R) would be opposite of the substantive finding from our estimated vote margin; the blue line represents values of $\{\rho_{\epsilon, \gamma}, R_{\epsilon}^2\}$ where $b^* = 0$. There are two key takeaways to highlight from the bias contour plots. First, for Michigan, we see that even if the omitted variable is relatively well balanced (i.e., $R_{\epsilon}^2 = 0.05$ on the y -axis of the Michigan plot), if the error is highly aligned with the outcome, i.e., correlated with the outcome by more than 0.5 on the x -axis, the omitted variable would be a killer confounder. Similarly, if the error in the weights is not highly aligned with the outcome (i.e., $\rho_{\epsilon, \gamma} = 0.05$), but is highly imbalanced (i.e., $R_{\epsilon}^2 = 0.75$), the confounder would also be a killer confounder. This showcases the importance of considering *both* sensitivity parameters when assessing the plausibility of a killer confounder.

Second, we see that the killer confounder region for North Carolina is larger than that for Michigan. As such, consistent with the robustness value, there is a greater degree of sensitivity to an omitted confounder altering the estimated result in North Carolina than in Michigan.

5.3 Formal Benchmarking

Both bias contour plots and the robustness value are useful methods for summarizing overall sensitivity in a point estimate. However, in practice, it is challenging to assess whether or not they are substantively meaningful. In the context of our running example in Michigan, it can be difficult

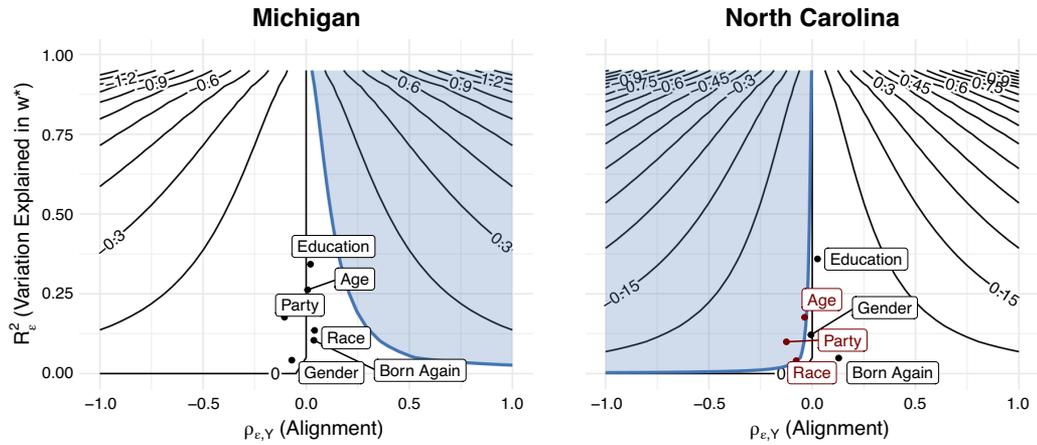


Figure 4. Bias contour plots. The shaded blue region represents the killer confounder region, in which the bias is large enough to reduce the margin to or below zero, changing the predicted winner. We also plot the results from formal benchmarking, where each point represents the parameter value of an omitted confounder with equivalent confounding strength of an observed covariate.

to answer whether or not it is plausible for an omitted confounder to be strong enough to explain 11% of the variation in both the ideal weights and the outcome process. Similarly, it is challenging to know from visual inspection the plausibility of the killer confounder region.

To address these challenges, we propose a procedure that allows researchers to use observed covariates to benchmark potential parameter values. Furthermore, we introduce several measures of relative confounding strength to help researchers assess sensitivity.

DEFINITION 5.1 (Benchmarked Error) Define a benchmark error term for observed covariate j as

$$\hat{\varepsilon}^{-j} := w^{-j} - w,$$

where w^{-j} represents the weights estimated using all covariates in $\phi(\mathbf{X})$, except for the j th covariate (i.e., $\phi(\mathbf{X}^{-j})$).

For example, for *education* $\hat{\varepsilon}^{-j}$ is the error, defined as the difference between our estimated weights and those estimated omitting *education*. In this setting, $\hat{\varepsilon}^{-j}$ captures the residual imbalance in educational attainment, after accounting for all other demographic variables.

We consider the confounding strength of an omitted confounder in terms of both the variation explained in the true weights and the alignment of the error and outcome. Formal benchmarking allows researchers to estimate the parameter values for an omitted confounder with equivalent confounding strength to an observed covariate, defined below. We can extend this to estimate the parameter values for an omitted confounder with confounding strength proportional, but not equivalent, to the benchmarked confounding strength. See Section B.2 of the Supplementary Material.

DEFINITION 5.2 (Equivalent Confounding Strength) An omitted confounder U has equivalent confounding strength to an observed covariate $\mathbf{X}^{(j)}$ if

$$\text{var}_S(\varepsilon) / \text{var}_S(w^*) = \text{var}_S(\hat{\varepsilon}^{-j}) / \text{var}_S(w^*), \tag{3}$$

$$\text{cor}_S(\varepsilon, Y) = \text{cor}_S(\hat{\varepsilon}^{-j}, Y), \tag{4}$$

where $\hat{\varepsilon}^{-j}$ is defined in Definition 5.1.

An omitted confounder with equivalent confounding strength explains the same amount of variation in the true weights as an observed covariate $\mathbf{X}^{(-j)}$, given $\phi(\mathbf{X}^{(-j)})$ (Equation (3)) and has the same level of alignment to the outcome as the benchmarked covariate (Equation (4)).

Using formal benchmarking, we can estimate the sensitivity parameters as

$$\hat{R}_\epsilon^2 = \frac{\hat{R}_\epsilon^{2-(j)}}{1 + \hat{R}_\epsilon^{2-(j)}}, \quad \hat{\rho}_{\epsilon,Y} = \hat{\rho}_{\epsilon,Y}^{-(j)}, \tag{5}$$

where $\hat{R}_\epsilon^{2-(j)} := \text{var}_S(\hat{\epsilon}^{-(j)})/\text{var}_S(w)$, and $\hat{\rho}_{\epsilon,Y}^{-(j)} := \text{cor}_S(\hat{\epsilon}^{-(j)}, Y)$ can be directly estimated using the benchmarked error term $\hat{\epsilon}^{-(j)}$. The benchmarked bias is estimated by plugging the values of Equation (5) to Equation (1). This procedure can be extended to account for subsets of covariates, thereby allowing researchers to posit parameter values, given the aggregate confounding strength of different combinations of covariates. Similarly, in two-step weighting approaches with a second-stage nonresponse adjustment to the weights, researchers could benchmark against the design weights.

An alternative is to estimate the *minimum relative confounding strength* (MRCS), or how many times stronger (or weaker) the confounding strength must be, relative to the benchmarked covariate, in order to change the substantive direction of a point estimate.

$$\text{MRCS}(j) = \frac{\hat{\mu} - b^*}{\text{Bias}(\phi(\mathbf{X}^{(-j)}))}.$$

An MRCS larger than 1 implies that the omitted confounder must be stronger than the observed covariate in order to be a killer confounder. Similarly, if the MRCS is smaller than 1, then a confounder weaker than the observed covariate would be a killer confounder.

The MRCS is especially useful in cases when researchers have a strong understanding of how observed covariates relate to survey sample selection and the outcome. For example, Kennedy *et al.* (2018) found that white voters in Midwestern states with lower levels of educational attainment underrepresented in many 2016 surveys. As such, researchers conducting surveys in subsequent elections could investigate the relative confounding strength of these observed covariates. An MRCS greater than 1 implies that an unobserved confounder has to be *stronger* than the confounding strength of these important observed covariates to be a killer confounder, which may be unlikely given the strength of these covariates. In such a case, we would conclude that our results are insensitive.

Benchmarking allows researchers to incorporate their substantive knowledge in understanding sensitivity. However, it is important to emphasize the limits of the benchmarked and summary measures. Our sensitivity framework allows researchers to transparently discuss the plausibility of unobserved confounders and the impact of their omission on the point estimates. However, none of our summary measures can fully eliminate the possibility of killer confounders. In particular, while a large MRCS value may indicate a large degree of robustness, it could also mean that none of the observed covariates used to benchmark are sufficiently explanatory of the outcome or the survey inclusion probability. This emphasizes the need for researchers to evaluate observable covariates that are highly prognostic in the diagnostics. Without strong substantive understanding of such covariates, the sensitivity analyses here have limited diagnostic value. Therefore, consistent with Cinelli and Hazlett (2020), we do not propose any thresholds or cutoff values for the robustness value or MRCS; similarly, we caution researchers from blindly using benchmarking across observed covariates, as the plausibility of fully unobserved confounders still depends on researchers' substantive judgment and context.

5.3.1 *Running Example: 2020 U.S. Presidential Election.* Figure 4 contains benchmarking against observed covariates for our running example. Each point, representing the bias of a confounder

Table 3. Formal benchmarking results for the ABC/Wapo polls for Michigan and North Carolina. The estimated bias is reported in percentage points.

| Variable | Michigan | | | | North Carolina | | | |
|------------|----------------------|---------------------------|-------|-----------|----------------------|---------------------------|-------|-----------|
| | \hat{R}_ϵ^2 | $\hat{\rho}_{\epsilon,Y}$ | MRCS | Est. bias | \hat{R}_ϵ^2 | $\hat{\rho}_{\epsilon,Y}$ | MRCS | Est. bias |
| Party | 0.18 | -0.11 | -2.44 | -1.87 | 0.10 | -0.13 | 0.31 | -1.17 |
| Age | 0.26 | 0.01 | 37.63 | 0.12 | 0.18 | -0.04 | 0.74 | -0.49 |
| Education | 0.34 | 0.02 | 8.75 | 0.52 | 0.36 | 0.02 | -0.71 | 0.52 |
| Gender | 0.04 | -0.07 | -8.14 | -0.56 | 0.12 | -0.01 | 5.07 | -0.07 |
| Race | 0.13 | 0.04 | 7.86 | 0.58 | 0.04 | -0.08 | 0.81 | -0.45 |
| Born again | 0.10 | 0.04 | 10.02 | 0.46 | 0.05 | 0.13 | -0.46 | 0.80 |

with equivalent confounding strength as the estimated benchmarked R_ϵ^2 and $\rho_{\epsilon,Y}$, is labeled with the observed covariate. Numeric results are presented in Table 3.

There are several key takeaways to highlight from the formal benchmarking results. First, we see that omitting a confounder with similar confounding strength as party identification, one of the strongest predictors of vote choice in U.S. politics, would result in the largest bias, given the other covariates included in weighting, understating Democratic support by 1.87 p.p. in Michigan, and 1.17 p.p. in North Carolina.

Second, the benchmarking results allow us to consider the full range of the sensitivity parameters. For example, in the North Carolina poll, omitting a variable similar to born-again Christian results in a large alignment $\rho_{\epsilon,Y}$ (0.13), but a relatively low R_ϵ^2 (0.05). This implies that while this variable is well balanced across the survey sample and the target population, the error is, as expected, highly explanatory of outcome. Variables with similar confounding strength could result in significant bias. In contrast, benchmarking against age results in a fairly large R_ϵ^2 (0.18), indicating high imbalance, but the error has a low correlation with the outcome. These covariates represent the types of confounders that, if omitted, would result in reasonably large magnitudes of bias.

Finally, the MRCS estimates have large magnitudes in Michigan, meaning that an omitted confounder would have to be substantially (more than two times) stronger than any of the observed covariates to result in enough bias to alter the estimated outcome that Biden wins the popular vote, our chosen b^* . Different choices of b^* would change the MRCS. In contrast, we see that in North Carolina, the MRCS estimates for several covariates are less than 1. In particular, an omitted variable need only be 30% as strong as party identification, or 70%–80% as strong as age or race/ethnicity to be a killer confounder.

5.3.2 *Takeaway.* From the sensitivity analysis, the Michigan results are quite robust; the error from a confounder would have to explain a large degree of variation in either the outcome or the ideal weights to overturn the prediction that Biden wins the popular vote in Michigan. For example, a confounder would have to be over twice as strong as party identification, one of the strongest predictors of vote choice in American politics. In contrast, a relatively weak confounder would change the estimated results in North Carolina.

6 Concluding Remarks

The social sciences rely heavily on surveys to answer a broad range of important questions. In the face of rising nonresponse and growing reliance on convenience samples, survey weights are a powerful tool that allow researchers to address nonrepresentative survey samples. Our proposed suite of sensitivity analysis tools allow researchers to reason about potential bias due to partially

and fully unobserved confounders. This includes tools for estimating bias, summary statistics, and graphical analyses with formal benchmarking against observed covariates.

This paper addresses the sensitivity of a point estimate to omitting a confounder, and, as such, does not explicitly account for changes in the uncertainty estimates. Our bias decomposition holds within finite samples; thus, researchers can apply a percentile bootstrap, calculating an adjusted weighted estimate conditional on the sensitivity parameters over repeated bootstrap samples, to construct valid intervals (Huang and Pimentel 2022; Soriano *et al.* 2023; Zhao *et al.* 2019). These confidence intervals account both for changes in the estimate from omitting a confounder and changes in estimated standard errors.

We focus on sensitivity to the decisions a researcher makes in the construction of survey weights. However, there is an increased emphasis on the use of outcome modeling in survey analysis, such as through model-assisted estimation (Breidt and Opsomer 2017) and doubly robust estimation (Chen, Li, and Wu 2020). Our framework readily extends to such settings. Researchers can consider the sensitivity in survey weights, given an outcome model, by replacing Y in our bias decomposition with the residual of the Y from the prediction from the outcome model. See Huang (2022) for a more detailed discussion.

Finally, we note that the application in this paper is situated in a very well-studied and highly theorized substantive area. Multiple papers have evaluated the bias in public opinion polling during the 2016 and 2020 U.S. Presidential Elections (i.e. Kennedy *et al.* 2018; Hartman *et al.* 2021, to name a few). As a result, the example is particularly useful for discussing how to perform the sensitivity analysis. However, we note that if researchers are operating in less theorized substantive areas, reasoning about the plausibility of omitted variables may be more challenging. We emphasize that the utility of a sensitivity analysis will always be dependent on a researcher's understanding of the survey context, and cannot be used to *replace* substantive knowledge. We propose a suite of tools to allow researchers to more transparently reason about sensitivity as well as better incorporate their contextual knowledge into the analysis. The sensitivity analysis and all of the corresponding sensitivity tools can be implemented using our R package `senseweight`.

Acknowledgments

We wish to thank Graeme Blair, Naoki Egami, Kirill Kalinin, Xiao Lu, and Erin Rossiter for thoughtful, valuable feedback. We also wish to thank the participants of the UNC-Chapel Hill FAQ Seminar and the C-Dem Workshop: Survey Design Under Constraints—Best Practices.

Funding

Melody Huang is supported by the National Science Foundation Graduate Research Fellowship under Grant No. 2146752.

Data Availability Statement

Replication code for this article is available at <https://doi.org/10.7910/DVN/YJSJEX> (Hartman and Huang, 2022).

Supplementary Material

For supplementary material accompanying this paper, please visit <https://doi.org/10.1017/pan.2023.12>.

References

- Ben-Michael, E., A. Feller, and E. Hartman. 2023. "Multilevel Calibration Weighting for Survey Data." *Political Analysis*, 1-19. doi:<https://doi.org/10.1017/pan.2023.9>.
- Ben-Michael, E., A. Feller, D. A. Hirshberg, and J. R. Zubizarreta. 2021b. "The Balancing Act in Causal Inference." Preprint, [arXiv:2110.14831](https://arxiv.org/abs/2110.14831).

- Bradley, V. C., S. Kuriwaki, M. Isakov, D. Sejdinovic, X.-L. Meng, and S. Flaxman. 2021. “Unrepresentative Big Surveys Significantly Overestimated US Vaccine Uptake.” *Nature* 600 (7890): 695–700.
- Breidt, F. J., and J. D. Opsomer. 2017. “Model-Assisted Survey Estimation with Modern Prediction Techniques.” *Statistical Science* 32 (2): 190–205.
- Caughey, D., A. J. Berinsky, S. Chatfield, E. Hartman, E. Schickler, and J. S. Sekhon. 2020. *Target Estimation and Adjustment Weighting for Survey Nonresponse and Sampling Bias*. Elements in Quantitative and Computational Methods for the Social Sciences. Cambridge: Cambridge University Press.
- Chen, Y., P. Li, and C. Wu. 2020. “Doubly Robust Inference with Nonprobability Survey Samples.” *Journal of the American Statistical Association* 115 (532): 2011–2021.
- Cinelli, C., and C. Hazlett. 2020. “Making Sense of Sensitivity: Extending Omitted Variable Bias.” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 82 (1): 39–67.
- Clinton, J., et al. 2021. *Task Force on 2020 Pre-Election Polling: An Evaluation of the 2020 General Election Polls*. New York: American Association for Public Opinion Research.
- Egami, N., and E. Hartman. 2021. “Covariate Selection for Generalizing Experimental Results: Application to a Large-Scale Development Program in Uganda.” *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 184 (4): 1524–1548.
- Hartman, E., C. Hazlett, and C. Sterbenz. 2021. “Kpop: A Kernel Balancing Approach for Reducing Specification Assumptions in Survey Weighting.” Preprint, [arXiv:2107.08075](https://arxiv.org/abs/2107.08075).
- Hartman, E., and M. Huang. 2022. “Replication Data for: ‘Sensitivity Analysis for Survey Weights.’” <https://doi.org/10.7910/DVN/YJSJEX>
- Haziza, D., and J.-F. Beaumont. 2017. “Construction of Weights in Surveys: A Review.” *Statistical Science* 32 (2): 206–226.
- Hong, G., F. Yang, and X. Qin. 2021. “Did You Conduct a Sensitivity Analysis? A New Weighting-Based Approach for Evaluations of the Average Treatment Effect for the Treated.” *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 184 (1): 227–254.
- Huang, M. 2022. “Sensitivity Analysis in the Generalization of Experimental Results.” Preprint, [arXiv:2202.03408](https://arxiv.org/abs/2202.03408).
- Huang, M., and S. D. Pimentel. 2022. “Variance-Based Sensitivity Analysis for Weighting Estimators Result in More Informative Bounds.” Preprint, [arXiv:2208.01691](https://arxiv.org/abs/2208.01691).
- Kennedy, C., et al. 2018. “An Evaluation of the 2016 Election Polls in the United States.” *Public Opinion Quarterly* 82 (1): 1–33.
- Kern, H. L., E. A. Stuart, J. Hill, and D. P. Green. 2016. “Assessing Methods for Generalizing Experimental Impact Estimates to Target Populations.” *Journal of Research on Educational Effectiveness* 9 (1): 103–127.
- Kish, L. 1965. *Survey Sampling*. New York: Wiley.
- Little, R. J., and D. B. Rubin. 2002. *Statistical Analysis with Missing Data: Little/Statistical Analysis with Missing Data*. Hoboken: Wiley.
- Meng, X.-L. 2018. “Statistical Paradises and Paradoxes in Big Data (I) Law of Large Populations, Big Data Paradox, and the 2016 US Presidential Election.” *Annals of Applied Statistics* 12 (2): 685–726.
- Nguyen, T. Q., C. Ebnesajjad, S. R. Cole, and E. A. Stuart. 2017. “Sensitivity Analysis for an Unobserved Moderator in RCT-to-Target-Population Generalization of Treatment Effects.” *Annals of Applied Statistics* 11 (1): 225–247.
- Peress, M. 2010. “Correcting for Survey Nonresponse using Variable Response Propensity.” *Journal of the American Statistical Association* 105 (492): 1418–1430.
- Rosenbaum, P. R., and D. B. Rubin. 1983. “Assessing Sensitivity to an Unobserved Binary Covariate in an Observational Study with Binary Outcome.” *Journal of the Royal Statistical Society: Series B* 45 (2): 212–218.
- Schaffner, B., S. Ansolabehere, and S. Luks. 2021. “Cooperative Election Study Common Content, 2020.” Version V4. <https://doi.org/10.7910/DVN/E9N6PH>
- Shen, C., X. Li, L. Li, and M. C. Were. 2011. “Sensitivity Analysis for Causal Inference using Inverse Probability Weighting.” *Biometrical Journal* 53 (5): 822–837.
- Soriano, D., E. Ben-Michael, P. J. Bickel, A. Feller, and S. D. Pimentel. 2023. “Interpretable Sensitivity Analysis for Balancing Weights.” *Journal of the Royal Statistical Society Series A: (Statistics in Society)*: 0964–1998.
- Tan, Z. 2006. “A Distributional Approach for Causal Inference using Propensity Scores.” *Journal of the American Statistical Association* 101 (476): 1619–1637.
- Zhao, Q., D. S. Small, and B. B. Bhattacharya. 2019. “Sensitivity Analysis for Inverse Probability Weighting Estimators via the Percentile Bootstrap.” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 81 (4): 735–761.