Bias-Corrected Crosswise Estimators for Sensitive Inquiries*

Yuki Atsusaka† Randolph T. Stevenson‡
atsusaka@rice.edu randystevenson@rice.edu

October 21, 2020

Abstract

The crosswise model is an increasingly popular survey technique to elicit candid answers from respondents on sensitive questions. We demonstrate, however, that the conventional crosswise estimator for the population prevalence of sensitive attributes is biased toward 0.5 in the presence of inattentive respondents who randomly choose their answers under this design. We propose a simple design-based bias correction procedure and show that our bias-corrected estimator can be easily implemented without measuring individual-level attentiveness. We also offer several useful extensions of our bias correction, including a sensitivity analysis for conventional crosswise estimates, a strategy for weighting, and a framework for multivariate regressions in which a latent sensitive trait is used as an outcome or a predictor. We illustrate our methodology by simulation studies and empirical examples and provide a practical guide for designing surveys to enable our proposed bias correction. Our method can be easily implemented through our open-source software, cWise.

Keyword: crosswise model, sensitive questions, survey design, randomized response techniques, sensitivity analysis, regression models.

Word Count: 10370

*An open-source software R package cWise:A (Cross)Wise Method to Analyze Sensitive Survey Questions, which implements our methods, is available at https://github.com/YukiAtsusaka/cWise. For helpful comments, we would like to thank Graeme Blair, Dongzhou Huang, Gary King, Shiro Kuriwaki, Jeff Lewis, John Londregan, Michelle Torres, and members of the Rice Method Research Group (Gustavo Guajardo, Colin Jones, and Yui Nishimura).

†Ph.D. Candidate, Department of Political Science, Rice University, MS-24 105 Herzstein Hall, 6100 Main St, Houston, TX 77005 (https://atsusaka.org).

‡Professor of Political Science, Rice University, MS-24 105 Herzstein Hall, 6100 Main St, Houston, TX 77005.

1 Introduction

Social scientists often use surveys to probe sensitive topics that respondents may hesitate to answer truthfully, such as racial animus (Kuklinski, Cobb and Gilens, 1997), discriminatory attitudes and behaviors about gender (De Jong, Pieters and Stremersch, 2012), support for militant organizations (Lyall, Blair and Imai, 2013), sexual behaviors (Vakilian, Mousavi and Keramat, 2014), corruption (Reinikka and Svensson, 2003), tax evasion (Korndörfer, Krumpal and Schmukle, 2014), and illicit drug use (Shamsipour et al., 2014). To mitigate social desirability bias in such inquiries, various survey techniques have been developed, including randomized response techniques (RRT), list experiments, and endorsement experiments (Blair, Imai and Zhou, 2015; Rosenfeld, Imai and Shapiro, 2016). Among these techniques, the *crosswise model* is an increasingly popular design, which shows respondents two statements — one about a sensitive attitude or behavior we want to study and one containing a piece of private information known only to each respondent — and asks them about the *joint* condition regarding the two statements (Yu, Tian and Tang, 2008; Jann, Jerke and Krumpal, 2011; Gingerich et al., 2016; Höglinger, Jann and Diekmann, 2016; Höglinger and Jann, 2018). The key property of the design is that even though researchers only observe respondents' answers to the joint condition about the two statements, they can nonetheless estimate the population prevalence of the sensitive attribute using their prior knowledge on the non-sensitive attribute.

Despite its promise, the literature has not yet delineated the statistical consequences of a common (but often overlooked) problem when implementing the crosswise model: having *inattentive respondents* who do not follow the specific instructions and randomly pick a given choice under this design. Indeed, previous research has estimated that between 2% and 28% of respondents were inattentive and randomly answering the question in past studies (Höglinger and Diekmann, 2017; Höglinger and Jann, 2018; Schnapp, 2019). While the literature has shown that measurement errors caused by such respondents substantially bias our inferences about the prevalence of sensitive attributes in other survey techniques (e.g., Blair and Imai, 2012; Blair, Chou and Imai, 2019), very little is known about the impact of potential measurement errors in the crosswise model, let alone how to address the potential problem.

In this article, we demonstrate that the conventional crosswise estimator for the population prevalence of sensitive traits is biased toward 0.5 in the presence of inattentive respondents who do not follow the instructions and instead randomly select a given choice. It is particularly problematic in sensitive inquiries because this bias is in *exactly the direction* that would lead researchers to conclude that the crosswise estimator is showing more of the sensitive behavior or opinion than direct questioning, and so is "working" as expected. In other words, researchers may mistakenly conclude that, by using the crosswise model, they have successfully induced truthful answers (i.e., higher estimated prevalence) even when such estimates are an artifact of bias caused by inattentive respondents. To remedy this problem, we offer a simple design-based bias correction procedure that enables researchers to estimate and correct for this bias without having to measure individual-level attentiveness for each respondent. Our solution only requires researchers to include a certain kind of *anchor question* in the survey while making several assumptions whose plausibility in a given survey can be increased with simple modifications to the survey design (e.g., randomizing the order in which answer categories are presented). Moreover, we also develop several useful extensions of our proposed *bias-corrected crosswise estimator*, including a sensitivity analysis, a weighting strategy, and a framework for multivariate regressions using the latent sensitive attribute as an outcome or as a predictor.

In what follows, we first describe the crosswise model, formally define inattentive respondents, discuss the negative inferential consequences of having these respondents, illustrate these consequences in a simple simulation, and discuss partial solutions from previous studies (Section 2). In Section 3, we formally describe the bias in the conventional crosswise estimator and propose a bias-corrected crosswise estimator that uses information from an anchor question to estimate the proportion of inattentive respondents. We then demonstrate the finite sample performance of the proposed estimator via simulation in Section 4.

In Section 5, we provide three extensions of our methodology: (1) a method for conducting sensitivity analyses to simulate the amount of bias caused by inattentive respondents even in studies that have not asked the anchor question, (2) a weighting strategy that allows the estimator to be used with non-random samples, and (3) a framework for regression models in which the latent sensitive trait can be used either as an outcome or as a predictor while applying the bias correction. To illustrate these extensions, we re-analyze six previous studies using the crosswise model to show that inattentiveness can have large effects on their conclusions and provide simulations testifying to the efficacy of the weighting and regression methods. Finally, in Section 6, we provide practical advice about survey design choices that can enhance the applicability, in any specific survey, of the assumptions underlying our method. Software to implement our bias-corrected estimator and its extensions, cWise, is available online. Additional analyses, information, and demonstrations can be found in our Online Appendix.

2 The Crosswise Model with Inattentive Respondents

In this section, we describe how the crosswise model works and demonstrate potential problems caused by inattentive respondents.

2.1 Crosswise Model

Yu, Tian and Tang (2008) introduced the crosswise model to overcome several limitations with previous RRT models (Warner, 1965; Blair, Imai and Zhou, 2015).[1] In the crosswise model, the respondent is asked to read two statements whose veracity is known only to her. One statement is about a sensitive topic of interest, while the other statement is about a non-sensitive topic whose population prevalence is *ex ante* known to researchers. For example, the crosswise question below includes a sensitive statement about attitudes toward immigrants as well as a non-sensitive statement about when the respondent's mother was born (which is a piece of private information known only to her, but whose population prevalence is known).[2]

Instruction: Please read the two statements below
Statement A: I would feel uncomfortable if an immigrant family moved in next door
Statement B: My mother was born in January, February, or March

Crosswise Question: Which of the following best describes your case?
- Both statements are TRUE, or both statements are FALSE ⇝ **Crosswise Item**
- Only one of them is TRUE

Here, Statement A is the sensitive statement that researchers would have asked directly if there had been no worry about social desirability bias. The quantity of interest is the population proportion of individuals who agree with (or fit the description in) Statement A. Often, this proportion is less than 0.5 (and closer to 0 than 0.5) so that direct questioning is expected (because of social desirability) to produce an estimate biased toward 0 (i.e., most people do not want to admit to having a sensitive attitude or behavior even if they fit the description).[3] In contrast, Statement B is a non-sensitive statement whose population prevalence is

[1]Most importantly, it provides no obvious strategy for uncooperative respondents seeking to defeat the method. For this reason, the crosswise model can be considered as a special case of RRT models.

[2]Other kinds of private information can be used. For example, if respondents can roll a die before each question, researchers could include a statement like "The value of the dice roll for this question was 4". One practical issue is that each crosswise question needs a "fresh" piece of private information. Rolling a dice before each question serves this purpose, but in practice, respondents will usually have to roll a virtual die. To reassure respondents that such virtual dice rolls can not be recorded and used by researchers to undermine privacy, one can ask the respondent to first pick (and perhaps write down) a "secret number" between 1 and 6 and then roll a die. The appropriate statement would then be "The value of the dice roll for this question was equal to my secret number"

[3]For this reason, we assume that the quantity of interest is always less than 0.5 in the rest of the argument. Even when it is greater

ex ante known to researchers.

After reading both statements, respondents choose whether (1) both are TRUE or both are FALSE or (2) only one of them is TRUE. The information that these answers reveal (and does not reveal) to researchers is demonstrated in Figure 1. Specifically, choosing the first answer, which we call the *crosswise item*, tells researchers that the respondent fits into one of the two gray cells on the diagonal of Figure 1, while choosing the second option reveals that the individual fits into one of the two white cells on the off-diagonal. However, this information does not allow researchers to infer whether the individual agrees or disagrees with the sensitive statement *per se* (i.e., fits into the gray or white cell in the upper row).

		Statement A (sensitive item)	
		TRUE	FALSE
Statement B (non-sensitive item)	TRUE		
	FALSE		

Figure 1: **The Crosswise Model.** *Note*: The crosswise model only asks respondents to reveal whether they correspond to the gray cells (main diagonal) or white cells (off diagonal).

More generally, because the respondent's answer is tied not only to the truth of the sensitive statement but the value of the non-sensitive private information, her privacy is protected.[4] Because she does not need to reveal whether she fits the description in Statement A *individually*, in principle, she is free to answer truthfully.[5] Further, by calculating the proportion of respondents who chose the *crosswise item*, which we call the *crosswise proportion*: $\mathbb{P}(\text{TRUE-TRUE} \cup \text{FALSE-FALSE})$, and using it along with the known population prevalence of the non-sensitive private information, we can compute the proportion of respondents for which the sensitive statement is true.

To demonstrate how this calculation works, we provide a simple numerical example. Suppose that the population proportion for Statement B is known to be 0.25 while the crosswise proportion is 0.65. If the

than 0.5 (e.g., most people say Yes to the statement), we can always word the sensitive statement to flip the expected direction (i.e., including "NOT" clause) or adjust estimates post-survey (i.e., subtracting the quantity of interest from 1).

[4]The essential aspect of the design of such questions is to ensure that respondents understand that they are protected. We can achieve this goal by being clear about why the survey is offering protection and how it works. In our experience, it is also helpful to rely on types of non-sensitive private information that respondents intuitively understand can not be known by researchers (e.g., a die roll).

[5]In our experience, once a respondent understands she is protected, it is also helpful to provide a positive reason to answer candidly. For example, in an employee survey about compliance with safety regulations, one might point out (in the introduction) that obtaining accurate information about the prevalence of unsafe behaviors across the whole company may allow for *less* time and resources to be spent on safety training, leaving more for other priorities.

probabilities of the sensitive and non-sensitive statements being true are independent, it follows that:

$$
\begin{aligned}
&\mathbb{P}(\textsf{TRUE-TRUE} \cup \textsf{FALSE-FALSE}) = 0.65 \\
\Rightarrow &\mathbb{P}(\textsf{A=TRUE}) \times \mathbb{P}(\textsf{B=TRUE}) + \mathbb{P}(\textsf{A=FALSE}) \times \mathbb{P}(\textsf{B=FALSE}) = 0.65 \\
\Rightarrow &\mathbb{P}(\textsf{A=TRUE}) \times 0.25 + \mathbb{P}(\textsf{A=FALSE}) \times 0.75 = 0.65 \\
\Rightarrow &\mathbb{P}(\textsf{A=TRUE}) \times 0.25 + (1 - \mathbb{P}(\textsf{A=TRUE})) \times 0.75 = 0.65 \\
\Rightarrow &\mathbb{P}(\textsf{A=TRUE}) = \frac{0.65 - 0.75}{-0.5} = 0.2.
\end{aligned}
$$

With this understanding of the crosswise model, we now describe how inattention can be understood in this model and trace its effects on the conventional estimator of the population prevalence of the sensitive trait.

2.2 Inattentive Respondents (and Why We Should Care)

The problem of inattentive respondents is well-known to survey researchers, who have used a variety of strategies for detecting them (e.g., attention checks or evidence of "straight-lining"). Estimates of the prevalence of inattentiveness in online-panels based on such measures are often as high as 8% to 12% (Crump, McDonnell and Gureckis, 2013; Meade and Craig, 2012; Maniaci and Rogge, 2014). Furthermore, researchers using the crosswise model have also found inattentive responses to be common, as we might expect given its relatively complex instructions and unfamiliar format.[6] Specifically, while Schnapp (2019) estimates that only 2% of respondents were inattentive, others have estimated the proportion of inattentive responses to be 12%, 13%, and 28% (in Höglinger and Diekmann (2017), Enzmann, 2017, cited in Schnapp (2019), and Höglinger and Jann (2018) respectively).[7]

In this article, we explicitly define inattentive respondents as *respondents who randomly pick the crosswise item or guess their answers in the crosswise question*.[8] That is, we rule out (by assumption) respondents who willfully choose the opposite answers that they would not have chosen had they been truthful. Such behavior is not due to "inattention" since respondents must necessarily understand the question and know

[6] Of course, we are not suggesting that *all* surveys are prone to the presence of inattentive respondents. Depending on survey mode, training of survey administrators, research context and topics, and the length and structure of surveys, it is possible (and ideal) to design and implement a survey so that it does not have any inattentive respondent.

[7] Inattentive respondents are also likely to be common (and may lead to bias) in surveys using other types of questions for sensitive inquiries (e.g., Blair, Chou and Imai, 2019).

[8] We assume that these are forced-choice questions that require respondents to provide some answer.

their "correct" choices to select the opposite answers.[9] Thus, inattentiveness in our case is equivalent to random guessing in the crosswise question. This definition of inattentiveness is particularly well-suited to the crosswise question format since one of its main strengths over other RRT questions is the lack of effective answer strategy for individuals who want to try to answer in a way that will "make themselves look good." Given this, most uncooperative respondents will find it easier just to guess randomly, which makes this behavior equivalent to inattentiveness.

Our first contribution to the literature is to clarify that point estimates of the prevalence of the sensitive trait from the crosswise model (hereafter *crosswise estimates*) are highly sensitive to the presence of inattentive respondents (as we have defined them above).[10] Specifically, the presence of inattentive respondents creates a bias in crosswise estimates toward 0.5.[11] This bias is particularly problematic in the context of sensitive inquiries as researchers wish to use the crosswise model exactly because they think that people might under-report a sensitive attitude or behavior in direct questioning. Thus, they often take crosswise estimates that yield higher prevalence rates than direct questioning as *prima facie* evidence that the crosswise model "worked" by successfully incentivizing respondents to reveal previously hidden attitudes or behaviors.

To illustrate this problem, Figure 2 plots the estimated proportion of sensitive attributes (with its 95% confidence interval) against the percentage of inattentive respondents based on hypothetical, but typical, data from the crosswise model.[12] It portrays that there is always a positive bias in the crosswise estimate and the bias increases as the percentage of inattentive respondents increases. While this "naïve" crosswise estimator is still unbiased under the (rare) condition that every respondent properly follows the instructions (the left arrow), this condition does not meet in almost all survey data. Our next contribution is to present a simple design-based approach to estimate and correct for the bias caused by inattentive respondents so that analysts can make more credible inferences about sensitive topics.

Although previous research has not shown what the bias looks like as in Figure 2, several solutions to the presence of inattentive respondents have been proposed. The first approach is to remove inattentive survey takers from data and perform estimation and inference on the "cleaned" data (Höglinger and Diekmann,

[9]To be clear, such "strategic uncooperative behaviors" or "strategic noncompliance" is NOT an effective way for respondents to "make themselves look good" and so are not incentivized by in the crosswise model.

[10]Of course, this will not be a problem if researchers define the prevalence of sensitive attributes *only among* attentive respondents as their quantity of interest.

[11]One can get some intuition for this result by considering the case in which every respondent randomly chooses the crosswise item 50% of the time. In this extreme case, the crosswise proportion will (by definition) be 0.5, which yields $\mathbb{P}(\mathsf{A}{=}\mathsf{TRUE}) = \frac{0.5-1+\mathbb{P}(\mathsf{B}{=}\mathsf{TRUE})}{2\mathbb{P}(\mathsf{B}{=}\mathsf{TRUE})-1} = \frac{\mathbb{P}(\mathsf{B}{=}\mathsf{TRUE})-0.5}{2(\mathbb{P}(\mathsf{B}{=}\mathsf{TRUE})-0.5)} = 0.5$.

[12]In this simulation, the ground truth for the proportion with the sensitive trait is set to 0.1, whereas the known probability for the non-sensitive statement is set to 0.15. The sample size is set to 2000.

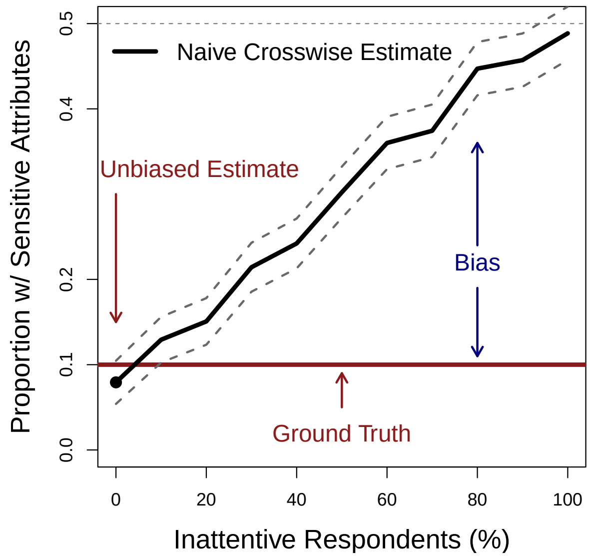

Figure 2: **Consequences of Inattentive Respondents.** *Note*: This plot illustrates that the bias increases as the percentage of inattentive responses grows. Gray dashed lines show 95% confidence intervals.

2017; Höglinger and Jann, 2018). However, this approach leads to a biased estimate of the quantity of interest unless researchers make the *ignorability assumption* that having a sensitive attribute is ignorable with respect to one's attentiveness in a survey.[13] This is an unreasonably strong assumption in most situations and needs to be empirically tested. Even if the ignorability assumption holds, this approach necessarily affects inferences as it drops some observations and decreases relative efficiency.

The second solution is to detect whether respondents answered the crosswise question randomly via direct questioning (i.e., "Did you lie about your answer to the last question?") and then to adjust the prevalence estimates accordingly (Schnapp, 2019). This approach is valid if researchers assume that direct questioning is itself not susceptible to inattentiveness or social desirability bias as well as that the crosswise question does not affect respondents' answers to the direct question. Such an assumption, however, is highly questionable. Below, we present an alternative solution to the problem, which yields an unbiased estimate of the quantity of interest with a remarkably weaker set of assumptions than in existing solutions.

[13]For a more general treatment of the consequences of dropping inattentive respondents in randomized experiments, see Aronow, Baron and Pinson (2019).

3 The Proposed Methodology

In this section, we formally derive the bias in the conventional crosswise estimator, offer a simple design-based method to estimate the bias, and introduce a resulting bias-corrected estimator of the population proportion of individuals who have sensitive attributes. The key idea in this section is that the bias is a function of the proportion of (in)attentive respondents; while this quantity is unknown in the conventional crosswise model, researchers can still estimate it by adding another question that resembles the original crosswise question and establishing a set of assumptions.

3.1 The Setup

Suppose that we consider a single sensitive question in a survey with n respondents, who were drawn from a finite population via simple random sampling. Suppose also that we apply the crosswise model to estimate the prevalence of the sensitive trait and there are no missing data. While we leave future research to tackle potential missing values in the crosswise model, we relax the random sampling assumption later to incorporate sample weights into the proposed estimator when using non-representative samples.

Let π be the population proportion of individuals who have the sensitive trait in question (our quantity of interest) and thus who fit the description in Statement A. Let p be the population proportion of people who have the non-sensitive attribute and thus who fit the characterization in Statement B. In the crosswise model, p is *ex ante* known to researchers. Finally, let λ be the population proportion of individuals who would choose the crosswise item.

Given these quantities (and assuming that the probabilities of A and B being true are independent from each other), Yu, Tian and Tang (2008) introduced the following identity as a foundation of the crosswise model:

$$\mathbb{P}(\text{TRUE-TRUE} \cup \text{FALSE-FALSE}) = \lambda = \pi p + (1 - \pi)(1 - p). \tag{1a}$$

Solving the identity with respect to the quantity of interest yields $\pi = \frac{\lambda+p-1}{2p-1}$. Yu, Tian and Tang (2008) then proposed the naïve crosswise estimator as follows:

$$\widehat{\pi}_{CM} = \frac{\widehat{\lambda} + p - 1}{2p - 1}, \tag{1b}$$

where $\widehat{\lambda}$ is the observed crosswise proportion from the crosswise model, and $p \neq 0.5$.

We call Equation (1b) the naïve estimator because it does not take into account the presence of inattentive respondents who randomly choose their answers in this design. The novelty of our approach is to explicitly incorporate these respondents in the above identity. Namely, when one or more respondents do not follow the instruction and randomly pick their answers, the crosswise proportion becomes:

$$\lambda = \Big\{\pi p + (1-\pi)(1-p)\Big\}\gamma + \kappa(1-\gamma), \tag{1c}$$

where γ is the proportion of *attentive* respondents and κ is the probability with which inattentive respondents pick the crosswise item.

Note that Equation (1c) is a strict generalization of Equation (1a), when γ is assumed to be 1. While the naïve crosswise estimator is unbiased as long as $\gamma = 1$ (see Online Appendix A.5), it is no longer an unbiased estimator of our estimand (π) whenever $\gamma < 1$. More specifically, we can define and quantify the bias in the conventional crosswise estimator as follows (see Online Appendix A.1):

$$\begin{aligned} B_{CM} &\equiv \mathbb{E}[\widehat{\pi}_{CM}] - \pi \\ &= \left(\frac{1}{2} - \frac{1}{2\gamma}\right)\left(\frac{\lambda - \kappa}{p - \frac{1}{2}}\right). \end{aligned}$$

Here, B_{CM} is a bias with respect to our quantity of interest caused by inattentive respondents. Importantly, under several regularity conditions, which are met in typical crosswise models, the bias term is always positive (see Online Appendix A.2). This means that the conventional crosswise estimator always *overestimates* the population prevalence of sensitive attributes in the presence of inattentive respondents. As discussed above, this property of the bias yields a highly problematic consequence in the context of sensitive inquiries. Figure A.1 in the Online Appendix visualizes the property of the bias in the conventional crosswise estimator.

3.2 Bias-Corrected Crosswise Estimator

To address this pervasive issue, we propose the following bias-corrected crosswise estimator:

$$\widehat{\pi}_{BC} = \widehat{\pi}_{CM} - \widehat{B}_{CM} \tag{2a}$$

where $\widehat{B}_{CM}$ is an unbiased estimator of the bias:

$$\widehat{B}_{CM} = \left(\frac{1}{2} - \frac{1}{2\widehat{\gamma}}\right)\left(\frac{\widehat{\lambda} - \frac{1}{2}}{p - \frac{1}{2}}\right), \tag{2b}$$

where $\widehat{\gamma}$ is the estimated proportion of attentive respondents (we discuss how to obtain $\widehat{\gamma}$ below).[14]

This bias correction depends on several assumptions. First, we assume that inattentive respondents pick the crosswise item with probability 0.5.[15]

Assumption 1 (Random Pick). *Inattentive respondents choose the crosswise item with probability 0.5 (i.e., $\kappa = 0.5$).*

The survey literature tells us that this assumption may not hold in many situations because inattentive respondents will be more likely to choose a first listed item than a second listed (or lower listed) item (Krosnick, 1991; Galesic et al., 2008). Nevertheless, it is still possible to *design* a survey so that we obtain $\kappa = 0.5$ regardless of how inattentive respondents randomly choose items. We can achieve this goal by randomizing the order of the listed items in the crosswise question.

The main challenge in estimating the bias is to estimate the proportion of attentive respondents in the crosswise question (i.e., obtaining $\widehat{\gamma}$). We solve this problem by adding an *anchor question* to the survey. The anchor question is a question that looks exactly like the crosswise question. The only difference is that in the anchor question both statements are about non-sensitive topics while we *ex ante* known the population prevalence for each statement. For example, in a survey administered to a U.S. population, we might consider the following anchor question:

Instruction: Please read the two statements below
Statement C: I am taking this survey in France
Statement D: My best friend was born in January, February, or March

Anchor Question: Which of the following most appropriately describes your case?
- Both statements are TRUE, or both statements are FALSE $\rightsquigarrow$ **Crosswise Item**
- Only one of them is TRUE

[14] As shown below, $\mathbb{E}[\widehat{\lambda}] = \lambda$ and $\mathbb{E}[\widehat{\gamma}] = \gamma$ and thus, by the linearity of the expected value operator, $\mathbb{E}[\widehat{B}_{CM}] = B_{CM}$.

[15] In our sensitivity analysis, researchers can also relax this assumption by using a pre-specified value for κ as discussed in Footnote 18.

Here, $\mathsf{Statement\ C}$ is a non-sensitive anchor statement while we know that the population proportion of individuals who fit the description is (supposed to be) 0. $\mathsf{Statement\ D}$ is another non-sensitive statement whose population prevalence is also known to researchers just like $\mathsf{Statement\ B}$. Let p' be the known proportion for $\mathsf{Statement\ D}$ and let λ' be the crosswise proportion in the anchor question. Let γ' be the population proportion of attentive respondents in the anchor question.

Then, by modifying Equation (1c) for the anchor question, we obtain

$$\gamma' = \frac{\lambda' - \frac{1}{2}}{\frac{1}{2} - p'}. \tag{3a}$$

Using the observed crosswise proportion in the anchor question (which we denote $\widehat{\lambda}'$), we can then estimate the proportion of attentive respondents in the anchor question as:

$$\widehat{\gamma}' = \frac{\widehat{\lambda}' - \frac{1}{2}}{\frac{1}{2} - p'}, \tag{3b}$$

where we can show that $\mathbb{E}[\widehat{\lambda}'] = \lambda'$ (see Online Appendix A.6).

Finally, our strategy is to use $\widehat{\gamma}'$ (obtained from the anchor question) as an estimate of γ in the crosswise question and plug it in Equation (2b) to estimate the bias. For this approach to be valid, of course, we need to assume that the proportion of attentive respondents does not change across the anchor and crosswise questions.[16] Also, we need to assume that the crosswise question does not affect respondents' answers to the anchor question (and vice versa). We state these assumptions as follows:

Assumption 2 (Attention Consistency). *The proportion of attentive respondents remains unchanged across the crosswise and anchor questions (i.e., $\gamma = \gamma'$).*

Assumption 3 (No Carryover). *The crosswise question does not affect respondents' answers to the anchor question and vice versa.*

Whether these assumptions are satisfied is an empirical matter. However, researchers can design their surveys so that they can make attention consistency and no carryover hold (e.g., by randomizing the order

[16]Here, we assume that the *salience* of the question is constant across the crosswise and anchor questions. It is then critical to design a survey so that the two questions would similar as suggested in Section 6.

of the two questions and making them look alike).

To summarize, our bias-corrected estimator provides an unbiased estimate of the population prevalence of a sensitive attribute even when some respondents do not follow the instructions and instead randomly choose between the answer categories of the crosswise question. It achieves this using only two, rather weak, assumptions that can be easily satisfied in the design stage of the survey.

3.3 Inference

Yu, Tian and Tang (2008, 257) show that the population variance of the conventional crosswise estimator and its sample analog are as follows:

$$\mathbb{V}(\widehat{\pi}_{CM}) = \mathbb{V}\left[\frac{\widehat{\lambda}}{2p-1}\right] = \frac{\lambda(1-\lambda)}{n(2p-1)^2}$$
$$\widehat{\mathbb{V}}(\widehat{\pi}_{CM}) = \widehat{\mathbb{V}}\left[\frac{\widehat{\lambda}}{2p-1}\right] = \frac{\widehat{\lambda}(1-\widehat{\lambda})}{n(2p-1)^2}$$

Based on a similar derivation, we consider the population variance and its sample analog of the bias-corrected estimator. To simplify the derivation, we first introduce the following assumption, which can easily be met by design:

Assumption 4 (Independent Crosswise Proportions). *The known auxiliary probabilities for non-sensitive statements in the crosswise and anchor questions are statistically independent from each other such that the two observed crosswise proportions are also independent. (i.e.,* $p \perp\!\!\!\perp p' \Rightarrow \widehat{\lambda} \perp\!\!\!\perp \widehat{\lambda}'$*.)*

With this assumption, we derive the variance of the bias-corrected crosswise estimator and its sample analog as follows (see Online Appendix A.4):

$$\mathbb{V}(\widehat{\pi}_{BC}) = \mathbb{V}\left[\frac{\widehat{\lambda}}{\widehat{\lambda}' - \frac{1}{2}}\right] \tag{4a}$$
$$\widehat{\mathbb{V}}(\widehat{\pi}_{BC}) = \widehat{\mathbb{V}}\left[\frac{\widehat{\lambda}}{\widehat{\lambda}' - \frac{1}{2}}\right] \tag{4b}$$

Note that these variances are necessarily larger than the variances of the conventional estimator. To see why, observe that these variances are a function of two random variables ($\widehat{\lambda}$ and $\widehat{\lambda}'$), whereas the conventional variances are a function of a single random variable ($\widehat{\lambda}$). In other words, the bias-corrected estimator inevitably has more uncertainty than the naïve crosswise estimator because the former also needs to estimate the proportion of (in)attentive respondents from data. Since no analytical solution is available for Equations (4a) and (4b), in practice, we employ the bootstrap to construct confidence intervals (with 200 bootstrap replications). In Section 6, we lay out several important points to consider when designing the survey to satisfy Assumptions 1-4 and effectively apply our bias correction procedure.

4 Simulation Studies

To illustrate our bias correction procedure, we perform several simulation studies. The left panel of Figure 4 presents the result of the bias correction applied to simulated (and typical) survey responses, where we set $n = 2000$, $\pi = 0.1$ (ground truth), $p = p' = 0.15$, and $\gamma = 0.8$. It shows that the naïve point estimate is far from the ground truth while its 95% confidence interval does not capture the estimand. In contrast, the bias-corrected point estimate is fairly close to the quantity of interest and its 95% confidence interval covers the quantity. Note also that the uncertainty around the points estimate based on the bias-corrected estimator is larger than the uncertainty around the naïve estimate.

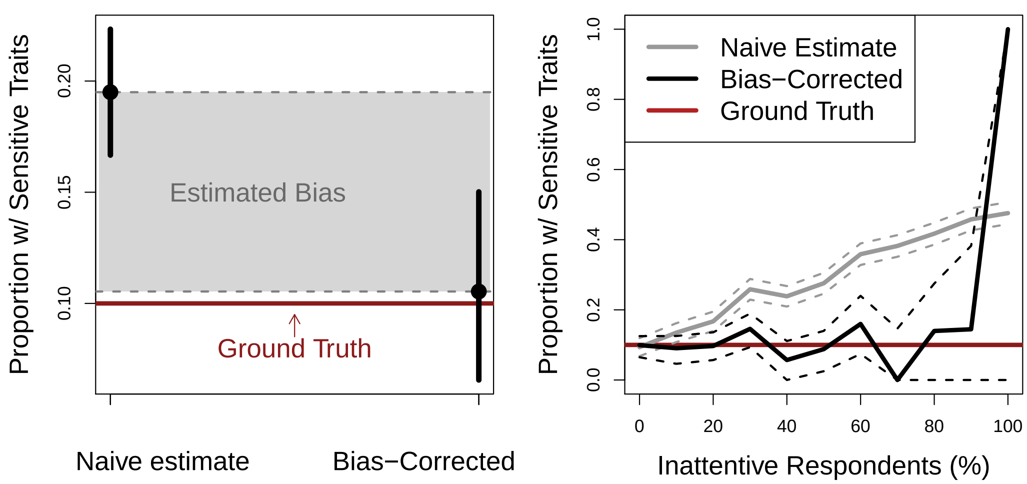

Figure 3: **Illustrations of Bias-Correction.** *Note*: The left panel illustrates that our bias-corrected estimator corrects for the estimated bias. The right panel shows that our bias-corrected estimator is robust to the presence of inattentive respondents, while the naïve estimator is increasingly biased as the percentage of inattentive respondents grows.

Using the same parameters values, we also simulate both naïve and bias-corrected estimates under varying levels of inattentive respondents. The right panel of Figure 4 demonstrates that while the naïve estimator performs increasingly poorly as more inattentive respondents are present in the survey, the bias-corrected estimator is rather robust to inattentive respondents and always captures the ground truth. It also indicates that when more than 90% of responses are inattentive, the bias-corrected estimate is no longer very informative as the 95% confidence interval covers most parts of the parameter space. However, in such surveys (in which less than 10% of respondents are paying attention), *any statistic* would be uninformative such that researchers should not use the data without precautions.

To examine the finite sample performance of our bias-corrected estimator, we further replicate the above simulations 8000 times and investigate the properties of the bias-corrected estimator more systematically. In each simulation, we choose different parameter values and examine the bias, the sample root-mean-square error (RMSE), and the coverage of the ground truth during the simulations. To apply our method to realistic contexts, we choose a set of reasonable parameter values from the parameter space (reflecting typical situations in which the crosswise model is used). Specifically, in each simulation, we draw π from a continuous uniform distribution (0.1, 0.45), p and p' from a continuous uniform distribution (0.1, 0.2), and γ from a continuous uniform distribution (0.5, 1). Finally, we repeat the set of experiments for different sample sizes of 200, 500, 1000, 2000, and 5000. We report our results in Figure 4. These results demonstrate that the bias-corrected estimator has a significantly lower bias, smaller RMSE, and higher coverage than the naïve estimator.[17] In Online Appendix B, we present another set of simulation studies to show that the bias-corrected estimator performs well even when respondents with a sensitive trait are more likely to be inattentive than respondents without the sensitive attribute. We also offer an empirical illustration in Online Appendix D.

5 Extensions

In this section, we consider three extensions of the bias-corrected estimator. First, we offer a sensitivity analysis for the crosswise model where the anchor question is not available. Second, we propose a weighting

[17]The difference between the bias-corrected estimator and naïve estimator is especially remarkable concerning the coverage of the true parameters. While the coverage of the native estimator's 95% confidence intervals rapidly deteriorates as the sample size increases, the bias-corrected estimator captures the true parameter approximately 95% of the time regardless of the sample size.

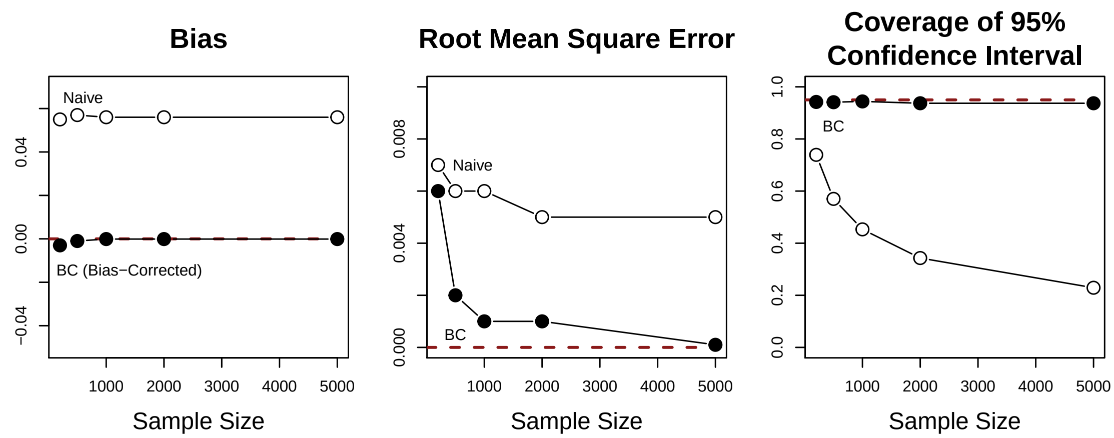

Figure 4: **Finite Sample Performance of the Naïve and Bias-Corrected Estimators** *Note*: This figure displays the bias, root mean square error, and the coverage of 95% confidence interval of naïve and bias-corrected estimators.

strategy to estimate the population prevalence of sensitive traits among the population of interest using unrepresentative samples. Finally, we consider a framework for multivariate regressions in which the latent sensitive trait can be used either as an outcome or a predictor.

5.1 Sensitivity Analysis

While our proposed bias correction requires researchers to estimate the proportion of inattentive respondents using the anchor question, it may be the case that researchers do not have access to such information (e.g., because they already used the crosswise model in the past or they could not include the anchor question due to cost constraints). To enable analysts to take advantage of our bias correction even in such a situation, we propose a sensitivity analysis that helps them clarify how sensitive their crosswise estimates are to the presence of inattentive respondents and what assumptions they must make to preserve their original conclusions. Specifically, we offer a set of sensitivity bounds for original crosswise estimates by applying the bias correction to them under varying levels of inattentive respondents. With our sensitivity bounds, researchers can ask *to what extent they can tolerate* the presence of inattentive respondents to keep their substantive conclusions.

To illustrate this procedure, we apply the sensitivity analysis to six published studies that used the crosswise model to study sensitive behaviors, including partial and severe plagiarism (Jann, Jerke and Krumpal, 2011), tax evasion (Korndörfer, Krumpal and Schmukle, 2014), illicit drug use (Shamsipour et al., 2014), academic misconduct (Höglinger, Jann and Diekmann, 2016), and sexual behavior (Vakilian, Mousavi and

Keramat, 2014). Figure 5 visualizes the sensitivity bounds. For each study, we plot the bias-corrected estimates of the quantity of interest against varying percentages of inattentive respondents under Assumption 1. We also plot the point and interval estimates based on direct questioning (if available) because many studies attempt to show (and indeed claim) that the crosswise estimator *performs better* (i.e., leads to higher estimates) than direct questioning. The original point estimates are shown by $\times$.

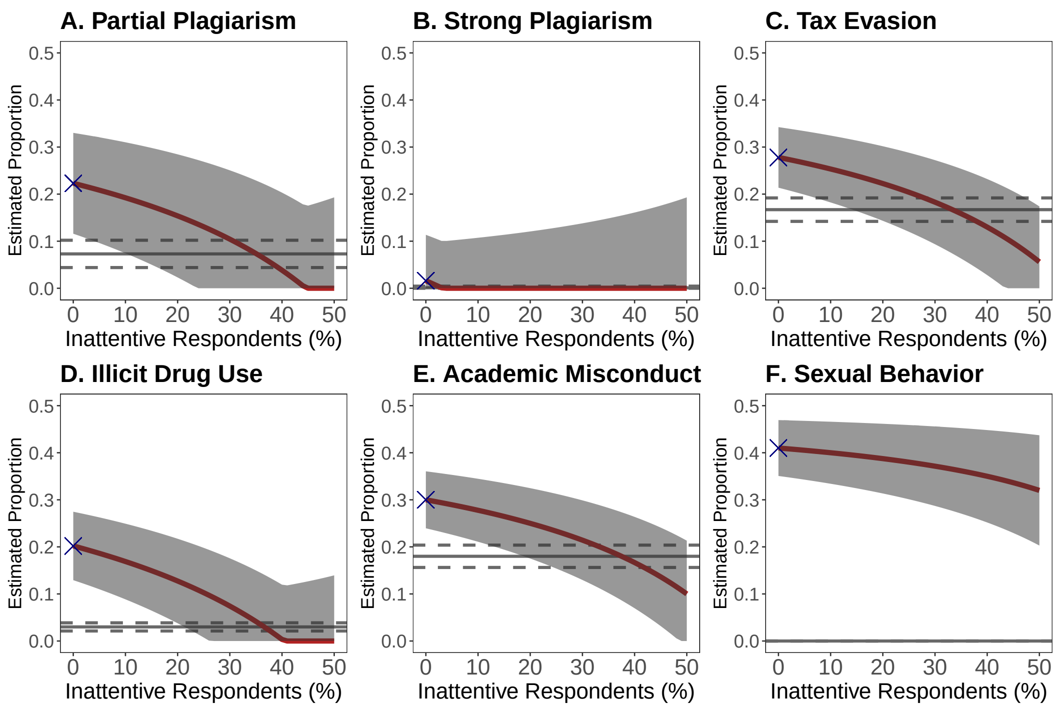

Figure 5: **Sensitivity Analysis of Previous Crosswise Estimates.** *Note*: This figure shows the results of sensitivity analysis for six crosswise estimates. For each estimate, the bias correction is applied with varying levels of the attentive rate under Assumption 1 ($\kappa = 0.5$). The solid and dashed lines show point and interval estimates based on direct questioning (except for Study F). The original point estimates are marked by $\times$.

The results suggest that, in many cases, high estimates that previous studies reported may have been an artifact of inattentive respondents. Our sensitivity analysis implies that most of these studies do not find any statistically significant difference between direct questioning and the crosswise model unless they assume that less than 20% of the respondents were inattentive. In other words, to claim that the crosswise estimator does any better than direct questioning in these studies, researchers must assume that more than 80% of the respondents were attentive and properly followed the instructions. While we cannot verify this assumption with available data, we must look at the original estimates with precautions given the previous findings that between 2% and 28% may have been inattentive in surveys in which the crosswise model was employed (as discussed in Section 2.2).[18] Although we cannot make a similar comparison for Study F, the sensitivity analysis is informative on its own as it shows that the original estimate is relatively robust to the presence of inattentive respondents.

5.2 Weighting

While the literature of sensitive inquiries usually assumes that survey respondents are obtained via simple random sampling from a finite population of interest, a growing share of surveys is administered with unrepresentative samples such as online opt-in samples (Franco et al., 2017; Mercer, Lau and Kennedy, 2018).[19] Online opt-in samples are known to be often unrepresentative of the entire population that researchers want to study, and analysts using such samples may wish to use weighting to extend their inferences into the population of real interest. The benefit of using weighting is even larger in sensitive inquiries since sensitive questions are not usually asked to (e.g., nationally) representative samples in large-scale public surveys, and researchers often need to conduct surveys on their own to study these questions. To date, however, no research has provided a practical guide for how to include sample weights in the crosswise model. Fortunately, our bias-corrected crosswise estimator can straightforwardly incorporate sample weights.

Recall that only sample statistics we observe in our framework are $\hat{\lambda}$ and $\hat{\lambda}'$, which are observed proportions of respondents choosing the crosswise item in the crosswise and anchor questions, respectively. The key idea here is that we can apply a Horvitz-Thompson-type estimator of the mean (and thus the inverse probability weighting more generally) to the crosswise proportions, where weights are the inverse of the probabilities that respondents in different strata will be in the sample. Namely, we can apply a weight $w_i = \frac{1}{\Pr(S_i=1|\mathbf{X_i})}$, where $S_i = \{0, 1\}$ is a binary variable denoting if respondent i is in the sample and $\mathbf{X_i}$ is a vector of the respondent's background characteristics.[20]

Let $Y_i \in \{0, 1\}$ be a binary variable denoting if respondent i chooses the crosswise item in the crosswise question and $A_i \in \{0, 1\}$ be a binary variable denoting if the same respondent chooses the crosswise item in the anchor question. We then estimate the weighted crosswise proportion in the crosswise question and

[18]Our R software also allows researchers to set different values of κ other than 0.5 depending on the nature of their surveys. The effect of κ on the bounds is, however, context-dependent. This is because the relative size of bias is determined by a distance between κ and an estimated crosswise proportion as indicated by Equation (2b). In practice, the true value of κ is unknown to researchers unless they design their surveys so that Assumption 1 holds. We thus recommend that analysts consider multiple values of κ if Assumption 1 could be violated under their designs.

[19]One important exception is the statistical model for endorsement experiments developed by Bullock, Imai and Shapiro (2011) in which the authors consider weighting via poststratification.

[20]In this article, we only consider the base weight for the conceptual simplicity, but one can naturally include other weights such as non-response weights to construct the final survey weights)

the weighted proportion of attentive respondents in the following way:

$$\widehat{\lambda}_w = \frac{\sum_{i=1}^n w_i Y_i}{\sum_{i=1}^n w_i} \tag{5a}$$

$$\widehat{\gamma}_w = \frac{\frac{\sum_{i=1}^n w_i A_i}{\sum_{i=1}^n w_i} - \frac{1}{2}}{\frac{1}{2} - p'}, \tag{5b}$$

To illustrate this strategy, Figure 6 compares bias-corrected crosswise estimates based on simulated unrepresentative samples with and without sample weights (Online Appendix C.1 provides details in our simulation). It demonstrates that while the unweighted estimator always overestimates the ground truth, the weighted estimator captures the population-level quantity of interest.

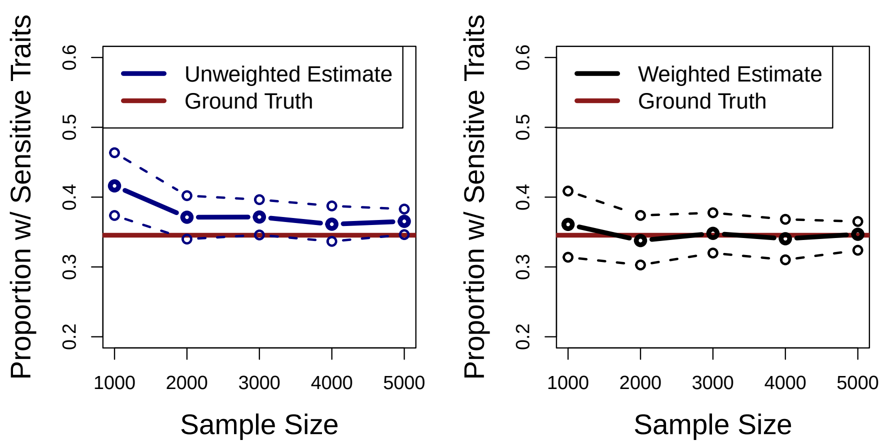

Figure 6: **Weighting in the Crosswise Estimator.** *Note*: This figure shows bias-corrected crosswise estimates without weighting (left panel) and estimates with weighting (right panel) based on simulated unrepresentative samples.

5.3 Multivariate Regressions

In many studies, researchers may wish to go beyond estimating the prevalence of sensitive attributes and draw statistical inferences about the associations between the sensitive attributes and respondents' characteristics in multivariate regressions. Political scientists, for example, may wish to estimate the prevalence of corruption in a legislature by using the crosswise model, analyze what kinds of politicians are more likely to engage in corruption, and probe whether engaging in corruption is associated with reelection. The main challenge in drawing such inferences is that analysts do not observe any individual-level information about sensitive attributes in the crosswise model. Despite the difficulty, we demonstrate that it is possible to use a regression framework to study these individual-level associations only with aggregate-level information about sensitive topics (obtained through the crosswise model) along with individual-level covariates while accounting for the bias.

Our proposed regression models are similar to (and build upon) several regression models developed for other survey techniques for sensitive inquiries including RRT (e.g., Blair, Imai and Zhou, 2015), list experiments (e.g., Imai, 2011; Blair and Imai, 2012; Imai, Park and Greene, 2015), and endorsement experiments (e.g., Bullock, Imai and Shapiro, 2011). While a simple regression model for the crosswise model, where the latent sensitive trait is used as an outcome, has been considered in several studies (Jann, Jerke and Krumpal, 2011; Vakilian, Mousavi and Keramat, 2014; Korndörfer, Krumpal and Schmukle, 2014), our contribution is to further extend such a framework by enabling analysts to use the latent sensitive trait either as an outcome or a predictor, while simultaneously applying our bias correction procedure. Our software can easily implement these regressions while also offering simple ways to perform post-estimation simulation with which researchers can generate predicted values for their quantities of interest along with uncertainty quantification via the parametric bootstrap.

5.3.1 Using the Latent Sensitive Trait as an Outcome

We first introduce multivariate "crosswise" regressions in which the latent (unobserved) variable for having a sensitive trait is used as an outcome variable while applying our bias correction. Let $Z_i \in \{0, 1\}$ be a binary variable denoting if respondent i has a sensitive trait and $T_i \in \{0, 1\}$ be a binary variable denoting if the same respondent is attentive. Both of these quantities are unobserved *latent* variables.

We define the regression model (conditional expectation) of interest as:

$$\mathbb{E}[Z_i|\mathbf{X_i} = \mathbf{x}] = \mathbb{P}(Z_i = 1|\mathbf{X_i} = \mathbf{x}) = \pi_{\boldsymbol{\beta}}(\mathbf{x}), \tag{6a}$$

where $\mathbf{X_i}$ is a random vector of respondent i's characteristics, $\mathbf{x}$ is a vector of specific values of such covariates, and $\boldsymbol{\beta}$ is a vector of unknown parameters (our quantities of interest) that associate these characteristics with the probability of having the sensitive trait. Our goal is to make inferences about these unknown parameters and to use estimated coefficients in predictions (e.g., Are young respondents more likely to support a terrorist organization than old respondents?).

To apply our bias correction along with the above regression, we also introduce the following conditional expectation for being attentive:

$$\mathbb{E}[T_i|\mathbf{X_i} = \mathbf{x}] = \mathbb{P}(T_i = 1|\mathbf{X_i} = \mathbf{x}) = \gamma_{\boldsymbol{\theta}}(\mathbf{x}), \tag{6b}$$

where $\boldsymbol{\theta}$ is a vector of unknown parameters that associate the same respondent's characteristics with the probability of being attentive.

According to our bias correction, we can naturally consider the crosswise proportions conditional upon covariates as functions of Equations (6a) and (6b) as follows:

$$\lambda_{\boldsymbol{\beta},\boldsymbol{\theta}}(\mathbf{X_i}) = \Big(\pi_{\boldsymbol{\beta}}(\mathbf{X_i})p + (1 - \pi_{\boldsymbol{\beta}}(\mathbf{X_i}))(1 - p)\Big)\gamma_{\boldsymbol{\theta}}(\mathbf{X_i}) + \frac{1}{2}\Big(1 - \gamma_{\boldsymbol{\theta}}(\mathbf{X_i})\Big) \tag{7a}$$

$$\lambda'_{\boldsymbol{\theta}}(\mathbf{X_i}) = \Big(\frac{1}{2} - p'\Big)\gamma_{\boldsymbol{\theta}}(\mathbf{X_i}) + \frac{1}{2} \tag{7b}$$

Assuming that Y_i and A_i are statistically independent conditional on $\mathbf{X_i}$, our approach is to model the joint probability distribution of the observed crosswise data and the covariates. Under Assumptions 1-4, we construct the following likelihood function:

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\beta},\boldsymbol{\theta}|\{\mathbf{X_i}, Y_i, A_i\}_{i=1}^n, p, p') &= \prod_{i=1}^n \Big\{\lambda_{\boldsymbol{\beta},\boldsymbol{\theta}}(\mathbf{X_i})\Big\}^{Y_i}\Big\{1 - \lambda_{\boldsymbol{\beta},\boldsymbol{\theta}}(\mathbf{X_i})\Big\}^{1-Y_i}\Big\{\lambda'_{\boldsymbol{\theta}}(\mathbf{X_i})\Big\}^{A_i}\Big\{1 - \lambda'_{\boldsymbol{\theta}}(\mathbf{X_i})\Big\}^{1-A_i} \\
&= \prod_{i=1}^n \Big\{\Big(\pi_{\boldsymbol{\beta}}(\mathbf{X_i})p + (1 - \pi_{\boldsymbol{\beta}}(\mathbf{X_i}))(1 - p)\Big)\gamma_{\boldsymbol{\theta}}(\mathbf{X_i}) + \frac{1}{2}\Big(1 - \gamma_{\boldsymbol{\theta}}(\mathbf{X_i})\Big)\Big\}^{Y_i} \\
&\quad \times \Big\{1 - \Big[\Big(\pi_{\boldsymbol{\beta}}(\mathbf{X_i})p + (1 - \pi_{\boldsymbol{\beta}}(\mathbf{X_i}))(1 - p)\Big)\gamma_{\boldsymbol{\theta}}(\mathbf{X_i}) + \frac{1}{2}\Big(1 - \gamma_{\boldsymbol{\theta}}(\mathbf{X_i})\Big)\Big]\Big\}^{1-Y_i} \\
&\quad \times \Big\{\Big(\frac{1}{2} - p'\Big)\gamma_{\boldsymbol{\theta}}(\mathbf{X_i}) + \frac{1}{2}\Big\}^{A_i} \\
&\quad \times \Big\{1 - \Big[\Big(\frac{1}{2} - p'\Big)\gamma_{\boldsymbol{\theta}}(\mathbf{X_i}) + \frac{1}{2}\Big]\Big\}^{1-A_i} \\
&= \prod_{i=1}^n \Big\{\Big((2p - 1)\pi_{\boldsymbol{\beta}}(\mathbf{X_i}) + \Big(\frac{1}{2} - p\Big)\Big)\gamma_{\boldsymbol{\theta}}(\mathbf{X_i}) + \frac{1}{2}\Big\}^{Y_i} \\
&\quad \times \Big\{1 - \Big[\Big((2p - 1)\pi_{\boldsymbol{\beta}}(\mathbf{X_i}) + \Big(\frac{1}{2} - p\Big)\Big)\gamma_{\boldsymbol{\theta}}(\mathbf{X_i}) + \frac{1}{2}\Big]\Big\}^{1-Y_i} \\
&\quad \times \Big\{\Big(\frac{1}{2} - p'\Big)\gamma_{\boldsymbol{\theta}}(\mathbf{X_i}) + \frac{1}{2}\Big\}^{A_i} \\
&\quad \times \Big\{1 - \Big[\Big(\frac{1}{2} - p'\Big)\gamma_{\boldsymbol{\theta}}(\mathbf{X_i}) + \frac{1}{2}\Big]\Big\}^{1-A_i}
\end{aligned} \tag{8a}$$

The final step is to determine the functional forms for $\pi_{\boldsymbol{\beta}}(\mathbf{X_i})$ and $\gamma_{\boldsymbol{\theta}}(\mathbf{X_i})$, respectively. Building on van den Hout, van der Heijden and Gilchrist (2007), we apply a logistic regression approach to model the conditional expectations in the crosswise and anchor questions, respectively:

$$\pi_{\boldsymbol{\beta}}(\mathbf{X_i}) = \text{logit}^{-1}(\boldsymbol{\beta}\mathbf{X_i}) \quad \text{and} \quad \gamma_{\boldsymbol{\theta}}(\mathbf{X_i}) = \text{logit}^{-1}(\boldsymbol{\theta}\mathbf{X_i}) \tag{8b}$$

Substituting these into Equation (8a) yields the likelihood function that contains every information we need to estimate our quantities of interest. For estimation, we maximize this likelihood function (after taking its natural log) by an iterative maximization method.[21] For inference, we use the negative inverse of the Hessian matrix of the log-likelihood evaluated at the maximum likelihood estimates to compute standard errors of the estimated coefficients.

5.3.2 Using the Latent Sensitive Trait as a Predictor

Next, we propose crosswise regressions in which the latent sensitive trait is used as a predictor (or independent variable) while applying our bias correction. To the best of our knowledge, this type of regression has been underdeveloped for the crosswise model. Thus, we begin by describing the model for the naïve crosswise estimator and then extend it to our bias-corrected estimator.

Let V_i be a continuous or discrete outcome variable for respondent i.[22] We define the regression model (conditional expectation) of interest as:

$$g_{\boldsymbol{\Theta}}(V_i|\mathbf{X_i}, Z_i), \tag{9a}$$

where $\boldsymbol{\Theta}$ is a vector of parameters that associate a set of predictors as well as an indicator for having a sensitive trait ($\mathbf{X_i}$, Z_i) and the response variable (V_i). For example, for a normally distributed outcome variable, we can consider $g_{\boldsymbol{\Theta}}(V_i|\mathbf{X_i}, Z_i) = \mathcal{N}(\alpha + \boldsymbol{\gamma}^\top \mathbf{X_i} + \delta Z_i, \sigma^2)$ with $\boldsymbol{\Theta} = (\alpha, \boldsymbol{\gamma}, \delta, \sigma^2)$. Similarly, for a binary response variable, we can consider $g_{\boldsymbol{\Theta}}(V_i|\mathbf{X_i}, Z_i) = \text{Bernoulli}(\phi)$, where $\frac{\phi}{1-\phi} = \alpha + \boldsymbol{\gamma}^\top \mathbf{X_i} + \delta Z_i$ and $\boldsymbol{\Theta} = (\alpha, \boldsymbol{\gamma}, \delta)$. Our goal is to make inferences about the association between the latent sensitive attribute (Z_i) and the response variable (V_i) after controlling for other covariates (e.g., Are people supporting a

[21] Another possible *estimation* strategy is based on the expectation-maximization (EM) algorithm in the spirit of Blair, Imai and Zhou (2015). However, we do not consider the EM-algorithm here as our simulation results imply that the direct maximization (iterative maximization of the entire likelihood) would suffice.

[22] Although other types of outcome variables can be easily incorporated into our framework, we leave to future research the development of such regressions.

terrorist organization more likely to donate money to a local political institution?). Thus, δ is our primary quantity of interest here.

To simplify the derivation, we assume that $V_i \perp\!\!\!\perp Y_i|\mathbf{X_i}$ (the outcome variable and the binary indicator for choosing the crosswise item are statistically independent conditional upon covariates). Using all the available information from data, we can construct the following likelihood function:

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\Theta}|\{V_i, \mathbf{X_i}, Y_i\}_{i=1}^n, p) &= \prod_{i=1}^n g_{\boldsymbol{\Theta}}(V_i|\mathbf{X_i}, Z_i)\mathbb{P}(Y_i = 1, Z_i|\mathbf{X_i}) \\
&= \prod_{i=1}^n \Big\{ g_{\boldsymbol{\Theta}}(V_i|\mathbf{X_i}, 1)p^{Y_i}(1-p)^{1-Y_i}\pi_{\boldsymbol{\beta}}(\mathbf{X_i}) \\
&\qquad + g_{\boldsymbol{\Theta}}(V_i|X_i, 0)(1-p)^{Y_i}p^{Y_i}(1-\pi_{\boldsymbol{\beta}}(\mathbf{X_i}))\Big\} \qquad \text{(9b)}
\end{aligned}$$

Here, the first part inside the bracket is $g_{\boldsymbol{\Theta}}(V_i|\mathbf{X_i}, Z_i = 1)\mathbb{P}(Y_i = 1, Z_i = 1|\mathbf{X_i}) = g_{\boldsymbol{\Theta}}(V_i|\mathbf{X_i}, Z_i = 1)\mathbb{P}(Y_i = 1|Z_i = 1)\mathbb{P}(Z_i = 1|\mathbf{X_i})$ and the second part is $g_{\boldsymbol{\Theta}}(V_i|\mathbf{X_i}, Z_i = 0)\mathbb{P}(Y_i = 1, Z_i = 0|\mathbf{X_i}) = g_{\boldsymbol{\Theta}}(V_i|\mathbf{X_i}, Z_i = 0)\mathbb{P}(Y_i = 1|Z_i = 0)\mathbb{P}(Z_i = 0|\mathbf{X_i})$.

Finally, we extend this framework by incorporating our bias correction. The observed data likelihood function then becomes:

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{\Theta}|\{V_i, \mathbf{X_i}, Y_i, A_i\}_{i=1}^n, p, p') &= \prod_{i=1}^n g_{\boldsymbol{\Theta}}(V_i|\mathbf{X_i}, Z_i, T_i)\mathbb{P}(Y_i = 1, Z_i, T_i|\mathbf{X_i})\mathbb{P}(A_i = 1, Z_i, T_i|\mathbf{X_i}) \\
&= \prod_{i=1}^n \Big\{ g_{\boldsymbol{\Theta}}(V_i|\mathbf{X_i}, 1, 1)p^{Y_i}(1-p)^{1-Y_i}\pi_{\boldsymbol{\beta}}(\mathbf{X_i})(1-p')^{A_i}p'^{1-A_i}\gamma_{\boldsymbol{\theta}}(\mathbf{X_i}) \\
&\qquad + g_{\boldsymbol{\Theta}}(V_i|\mathbf{X_i}, 0, 1)(1-p)^{Y_i}p^{1-Y_i}(1-\pi_{\boldsymbol{\beta}}(\mathbf{X_i}))(1-p')^{A_i}p'^{1-A_i}\gamma_{\boldsymbol{\theta}}(\mathbf{X_i}) \\
&\qquad + g_{\boldsymbol{\Theta}}(V_i|\mathbf{X_i}, 1, 0)\frac{1}{2}\pi_{\boldsymbol{\beta}}(\mathbf{X_i})\frac{1}{2}(1-\gamma_{\boldsymbol{\theta}}(\mathbf{X_i})) \\
&\qquad + g_{\boldsymbol{\Theta}}(V_i|\mathbf{X_i}, 0, 0)\frac{1}{2}(1-\pi_{\boldsymbol{\beta}}(\mathbf{X_i}))\frac{1}{2}(1-\gamma_{\boldsymbol{\theta}}(\mathbf{X_i}))\Big\}, \qquad \text{(9c)}
\end{aligned}$$

where each part is $g_{\boldsymbol{\Theta}}(V_i|\mathbf{X_i}, z, t)\mathbb{P}(Y_i = 1|Z_i = z, T_i = t)\mathbb{P}(Z_i = z|\mathbf{X_i})\mathbb{P}(A_i = 1|Z_i = z, T_i = t)\mathbb{P}(T_i = 1|\mathbf{X_i})$, where $z = \{0, 1\}$ and $t = \{0, 1\}$. To link the above likelihood function with a vector of covariates, we use the same model specification as in Equation (8b).

Here, we assume that Assumptions 1-4 hold as well as $V_i \perp\!\!\!\perp Y_i|\mathbf{X_i}$, $V_i \perp\!\!\!\perp A_i|\mathbf{X_i}$, and $Y_i \perp\!\!\!\perp A_i|\mathbf{X_i}$. The key idea is that, under these assumptions, we can rewrite the entire likelihood of the observed crosswise data as a product of three conditional probabilities. We can then marginalize the product over the two latent variables Z_i and T_i by summing up the conditional probabilities that we could in principle obtain for all possible combinations of the latent variables.[23] To estimate the unknown parameters, including δ (our primary quantity of interest), we use an iterative maximization of the above entire observed likelihood function (after taking its natural log). To compute standard errors of the estimated coefficients, we use the negative inverse of the Hessian matrix of the log likelihood as before.

To summarize, with our method, researchers can make valid statistical inferences about the population prevalence of sensitive attributes and use such estimates in multivariate analysis for further exploration. For both regressions, it is straightforward to compute predicted values with 95% confidence intervals via the parametric bootstrap (and our software offers a routine to implement it). We provide simulation studies of these extensions in Online Appendix C.

6 Practical Guide: How to Design a Crosswise Survey in the Presence of Inattentive Respondents

In this section, we offer a practical guide for researchers designing surveys that will use our proposed methodology. The validity of our bias correction and its extensions hinges upon the four assumptions discussed in Section 3. Below, we clarify several points that researchers should consider at the survey design stage to satisfy these assumptions.

6.1 How to Ensure that Inattentive Respondents Randomly Pick Items? (Assumption 1)

The random pick assumption states that inattentive respondents choose "both are TRUE or both are FALSE" at the probability of 0.5. This assumption can be satisfied by ensuring that inattentive respondents do not distinguish two available options (i.e., "both are TRUE or both are FALSE" and "one of them is TRUE")

[23] For example, the third component inside the bracket represents $g_{\boldsymbol{\Theta}}(V_i|\mathbf{X_i}, 1, 0)\mathbb{P}(Y_i = 1|Z_i = 1, T_i = 0)\mathbb{P}(Z_i = 1|\mathbf{X_i})\mathbb{P}(A_i = 1|Z_i = 1, T_i = 0)\mathbb{P}(T_i = 1|\mathbf{X_i})$. Here, $\mathbb{P}(Y_i = 1|Z_i = 1, T_i = 0)$ is the conditional probability that respondents choose the crosswise item when they actually have a sensitive trail *and* do not provide attentive responses. Because they do not follow the instruction, Assumption 1 states that this probability is $\frac{1}{2}$ (regardless of Z_i). Next, $\mathbb{P}(Z_i = 1|\mathbf{X_i})$ is the conditional probability that respondents have a sensitive trait, and we defined this quantity as $\pi_{\boldsymbol{\beta}}(\mathbf{X_i})$. Now, $\mathbb{P}(A_i = 1|Z_i = 1, T_i = 0)$ is the conditional probability that respondents choose the crosswise item in the anchor question when they actually have a sensitive trait *and* do not provide attentive responses. Because they do not follow the instruction, Assumption 1 states that this probability is $\frac{1}{2}$ (regardless of Z_i). Finally, $\mathbb{P}(T_i = 1|\mathbf{X_i})$ is the conditional probability that respondents *do not* provide attentive responses, and we defined this quantity as $1 - \gamma_{\boldsymbol{\theta}}(\mathbf{X_i})$. Hence, the joint probability for this component is $g_{\boldsymbol{\Theta}}(V_i|\mathbf{X_i}, 1, 0)\frac{1}{2}\pi_{\boldsymbol{\beta}}(\mathbf{X_i})\frac{1}{2}(1 - \gamma_{\boldsymbol{\theta}}(\mathbf{X_i}))$.

and they pick one of the two choices randomly. A simple approach to make this assumption plausible is to randomize the ordering of the two answer-choices both in the crosswise and anchor questions.

6.2 How to Achieve Attention Consistency and No Carryover (Assumptions 2-3)

Attention consistency is satisfied when the crosswise and anchor questions have the same *proportion* of attentive respondents. Importantly, this assumption does not require that the same respondents remain attentive across the two questions. To satisfy this assumption, we suggest that researchers design their surveys so that respondents see the crosswise and anchor questions in the same way. If respondents, on average, perceive one question to be somehow different from the other question, attention consistency could be violated. Thus, we recommend that researchers design both the crosswise and anchor questions to look quite similar (e.g., the anchor question has the same length of wording as the crosswise question). Moreover, randomizing the position of the anchor question in the survey relative to the crosswise question will be helpful to guarantee that there is no "carryover effect" from one type of question to another.

6.3 How to Make Independent Crosswise Proportions (Assumption 4)

The last assumption requires that auxiliary probabilities (p and p') used in the crosswise and anchor questions are statistically independent such that the two crosswise proportions are also statistically independent from each other. This assumption will be satisfied when researchers carefully choose p and p' based on different topics. For example, when p is selected based on one's mother's birth month and p' is based on her father's birth month we are more or less confident that this assumption holds (assuming that marriage is not a function of birth months of partners).[24] We recommend that researchers always *ex ante* ask respondents to think of two (or more) different topics (e.g., friends, friend and parent, friend and sibling, etc.) and then use the information for p and p'.[25] This strategy also helps researchers by separating the respondents' tasks of coming up with topics and thinking about questions.

[24] Importantly, this assumption will be violated when researchers only use a single topic (e.g., mother's birth month) with two different "cut-off points" (e.g., January to March and October to December). It is because the probability that one's mother was born in the first period contains information about the probability that she was born in the second period.

[25] Using multiple siblings in the birth month-type randomization can be problematic since two siblings' birth months may not necessarily be statistically independent. Another option is to use virtual dice rolls as explained in footnote 2

Concluding Remarks

The crosswise model is a powerful survey-based tool for investigating the prevalence of sensitive attitudes and behaviors. The presence of inattentive respondents, however, may jeopardize statistical inferences based on the conventional crosswise estimator. In particular, we showed that the bias caused by inattentive respondents often leads researchers to conclude that the method induced more candid responses even when such conclusions are an artifact of bias. We have proposed a simple design-based solution to this often overlooked problem using the anchor question, which can estimate and correct for the bias without having to measure the attentiveness of individual respondents. We have also provided several useful extensions of our bias correction, including a sensitivity analysis, weighting strategy, and multivariate regressions, which allow researchers to analyze sensitive topics with the crosswise model in ways that were not available before.

To maximize the utility of the proposed methodology, we recommend that applied researchers conduct an extensive pilot study to try different wordings and values for p and p' for the crosswise and anchor questions. In the pilot study, it is critical to pay attention to the practical guide in Section 6 and examine how such design-based efforts can improve the empirical validity of the assumptions. Researchers can also use our sensitivity analysis as a helpful tool to assist the pilot study by visually inspecting the degree of bias in sensitive topics that they study. Importantly, we do not advocate that researchers implement a "poor survey" and then correct for the bias with our method. We hope that our discussion and mathematical results can encourage researchers to write better instructions and questions for respondents so that they can minimize the proportion of inattentive respondents.

We believe that, when carefully implemented, our method will be a valuable addition to the toolkit for analyzing sensitive questions along with the randomized response technique (Blair, Imai and Zhou, 2015), list experiments (Imai, 2011; Blair and Imai, 2012; Imai, Park and Greene, 2015), and endorsement experiments (Bullock, Imai and Shapiro, 2011). How to combine data from the (bias-corrected) crosswise model and other indirect questioning techniques and improve our inferences will be a fruitful topic in future research.

References

Aronow, Peter M, Jonathon Baron and Lauren Pinson. 2019. "A note on dropping experimental subjects who fail a manipulation check." *Political Analysis* 27(4):572–589.

Blair, Graeme and Kosuke Imai. 2012. "Statistical analysis of list experiments." *Political Analysis* 20(1):47–77.

Blair, Graeme, Kosuke Imai and Yang-Yang Zhou. 2015. "Design and analysis of the randomized response technique." *Journal of the American Statistical Association* 110(511):1304–1319.

Blair, Graeme, Winston Chou and Kosuke Imai. 2019. "List experiments with measurement error." *Political Analysis* 27(4):455–480.

Bullock, Will, Kosuke Imai and Jacob N Shapiro. 2011. "Statistical analysis of endorsement experiments: Measuring support for militant groups in Pakistan." *Political Analysis* pp. 363–384.

Crump, Matthew JC, John V McDonnell and Todd M Gureckis. 2013. "Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research." *PloS one* 8(3):e57410.

De Jong, Martijn G, Rik Pieters and Stefan Stremersch. 2012. "Analysis of sensitive questions across cultures: An application of multigroup item randomized response theory to sexual attitudes and behavior." *Journal of personality and social psychology* 103(3):543.

Franco, Annie, Neil Malhotra, Gabor Simonovits and LJ Zigerell. 2017. "Developing standards for post-hoc weighting in population-based survey experiments." *Journal of Experimental Political Science* 4(2):161–172.

Galesic, Mirta, Roger Tourangeau, Mick P Couper and Frederick G Conrad. 2008. "Eye-tracking data: New insights on response order effects and other cognitive shortcuts in survey responding." *Public Opinion Quarterly* 72(5):892–913.

Gingerich, Daniel W, Virginia Oliveros, Ana Corbacho and Mauricio Ruiz-Vega. 2016. "When to protect? Using the crosswise model to integrate protected and direct responses in surveys of sensitive behavior." *Political Analysis* pp. 132–156.

Höglinger, Marc and Andreas Diekmann. 2017. "Uncovering a blind spot in sensitive question research: false positives undermine the crosswise-model RRT." *Political Analysis* 25(1):131–137.

Höglinger, Marc and Ben Jann. 2018. "More is not always better: An experimental individual-level validation of the randomized response technique and the crosswise model." *PloS one* 13(8):e0201770.

Höglinger, Marc, Ben Jann and Andreas Diekmann. 2016. Sensitive questions in online surveys: An experimental evaluation of different implementations of the randomized response technique and the crosswise model. In *Survey Research Methods*. Vol. 10 pp. 171–187.

Imai, Kosuke. 2011. "Multivariate regression analysis for the item count technique." *Journal of the American Statistical Association* 106(494):407–416.

Imai, Kosuke, Bethany Park and Kenneth F Greene. 2015. "Using the predicted responses from list experiments as explanatory variables in regression models." *Political Analysis* pp. 180–196.

Jann, Ben, Julia Jerke and Ivar Krumpal. 2011. "Asking sensitive questions using the crosswise model: an experimental survey measuring plagiarism." *Public opinion quarterly* 76(1):32–49.

Korndörfer, Martin, Ivar Krumpal and Stefan C Schmukle. 2014. "Measuring and explaining tax evasion: Improving self-reports using the crosswise model." *Journal of Economic Psychology* 45:18–32.

Krosnick, Jon A. 1991. "Response strategies for coping with the cognitive demands of attitude measures in surveys." *Applied cognitive psychology* 5(3):213–236.

Kuklinski, James H, Michael D Cobb and Martin Gilens. 1997. "Racial attitudes and the" New South"." *The Journal of Politics* 59(2):323–349.

Lyall, Jason, Graeme Blair and Kosuke Imai. 2013. "Explaining support for combatants during wartime: A survey experiment in Afghanistan." *American Political Science Review* 107(4):679–705.

Maniaci, Michael R and Ronald D Rogge. 2014. "Caring about carelessness: Participant inattention and its effects on research." *Journal of Research in Personality* 48:61–83.

Meade, Adam W and S Bartholomew Craig. 2012. "Identifying careless responses in survey data." *Psychological methods* 17(3):437.

Mercer, Andrew, Arnold Lau and Courtney Kennedy. 2018. "For Weighting Online Opt-In Samples, What Matters Most." *Pew Research Center* .

Reinikka, Ritva and Jakob Svensson. 2003. *Survey techniques to measure and explain corruption*. The World Bank.

Rosenfeld, Bryn, Kosuke Imai and Jacob N Shapiro. 2016. "An empirical validation study of popular survey methodologies for sensitive questions." *American Journal of Political Science* 60(3):783–802.

Schnapp, Patrick. 2019. "Sensitive Question Techniques and Careless Responding: Adjusting the Crosswise Model for Random Answers." *methods, data, analyses* 13(2):13.

Shamsipour, Mansour, Masoud Yunesian, Akbar Fotouhi, Ben Jann, Afarin Rahimi-Movaghar, Fariba Asghari and Ali Asghar Akhlaghi. 2014. "Estimating the prevalence of illicit drug use among students using the crosswise model." *Substance Use & Misuse* 49(10):1303–1310.

Vakilian, Katayon, Seyed Abbas Mousavi and Afsaneh Keramat. 2014. "Estimation of sexual behavior in the 18-to-24-years-old Iranian youth based on a crosswise model study." *BMC research notes* 7(1):28.

van den Hout, Ardo, Peter GM van der Heijden and Robert Gilchrist. 2007. "The logistic regression model with response variables subject to randomized response." *Computational Statistics & Data Analysis* 51(12):6060–6069.

Warner, Stanley L. 1965. "Randomized response: A survey technique for eliminating evasive answer bias." *Journal of the American Statistical Association* 60(309):63–69.

Yu, Jun-Wu, Guo-Liang Tian and Man-Lai Tang. 2008. "Two new models for survey sampling with sensitive characteristic: design and analysis." *Metrika* 67(3):251.

Online Appendix

For "Bias-Corrected Crosswise Estimators for Sensitive Inquiries"

Yuki Atsusaka and Randolph T. Stevenson

Table of Contents

A Additional Discussion on the Bias-Corrected Estimator

A.1 Derivation of the Bias

Here, we derive the bias in the naïve crosswise estimator based on the argument in Section 3.

$$
\begin{aligned}
B_{CM} &\equiv \mathbb{E}[\widehat{\pi}_{CM}] - \pi \\
&= \mathbb{E}\left[\frac{\widehat{\lambda} + p - 1}{2p - 1}\right] - \frac{\lambda - (1-p)\gamma - \kappa(1-\gamma)}{(2p-1)\gamma} \\
&= \frac{\gamma(\lambda + p - 1) - (\lambda - \gamma + p\gamma - \kappa + \kappa\gamma)}{(2p-1)\gamma} \\
&= \frac{\lambda\gamma + p\gamma - \gamma - \lambda + \gamma - p\gamma + \kappa - \kappa\gamma}{(2p-1)\gamma} \\
&= \frac{\lambda\gamma - \kappa\gamma - \lambda + \kappa}{(2p-1)\gamma} \\
&= \frac{\lambda - \kappa}{(2p-1)} - \frac{\lambda - \kappa}{(2p-1)\gamma} \\
&= \frac{1}{2}\left(\frac{\lambda - \kappa}{p - \frac{1}{2}}\right) - \frac{1}{2\gamma}\left(\frac{\lambda - \kappa}{p - \frac{1}{2}}\right) \\
&= \left(\frac{1}{2} - \frac{1}{2\gamma}\right)\left(\frac{\lambda - \kappa}{p - \frac{1}{2}}\right)
\end{aligned}
$$

A.2 Behavior of the Bias

By definition, the bias vanishes when the proportion of attentive respondents is 1 ($\lambda = 1$). To see this, simply observe the following limit:

$$
\begin{aligned}
&\lim_{\lambda \to 1}\left(\frac{1}{2} - \frac{1}{2\gamma}\right)\left(\frac{\lambda - \kappa}{p - \frac{1}{2}}\right) \\
&= \left(\frac{1}{2} - \frac{1}{2}\right)\left(\frac{\lambda - \kappa}{p - \frac{1}{2}}\right) \\
&= 0
\end{aligned}
$$

In contrast, as the proportion of attentive approaches 0 (from the side of 1), the bias term explodes and approaches the positive infinity. To see this, observe that the multiplier $\left(\frac{1}{2} - \frac{1}{2\lambda}\right)$ is always negative and the multiplicand $\left(\frac{\lambda-\kappa}{p-\frac{1}{2}}\right)$ is also negative under some regularity conditions. These conditions state that $\lambda > \kappa$ and $p < \frac{1}{2}$. The regularity conditions hold in most surveys with the crosswise model. Since the absolute value of the multiplier grows as λ approaches 0, the bias term increases as the proportion of attentive responses decreases.

However, the limit itself does not exist as:

$$\lim_{\lambda \to 0} \left(\frac{1}{2} - \frac{1}{2\gamma} \right) \left(\frac{\lambda - \kappa}{p - \frac{1}{2}} \right)$$
$$= \text{Undefined}$$

A.3 Simulation of the Bias

The following figure visualizes the property of the bias in the conventional crosswise estimator (assuming $\kappa = 0.5$). It shows that the size of the bias increases as the percentage of inattentive respondents increases, while it also increases as the quantity of interest approaches 0, but that it does not change regardless of the value of p.

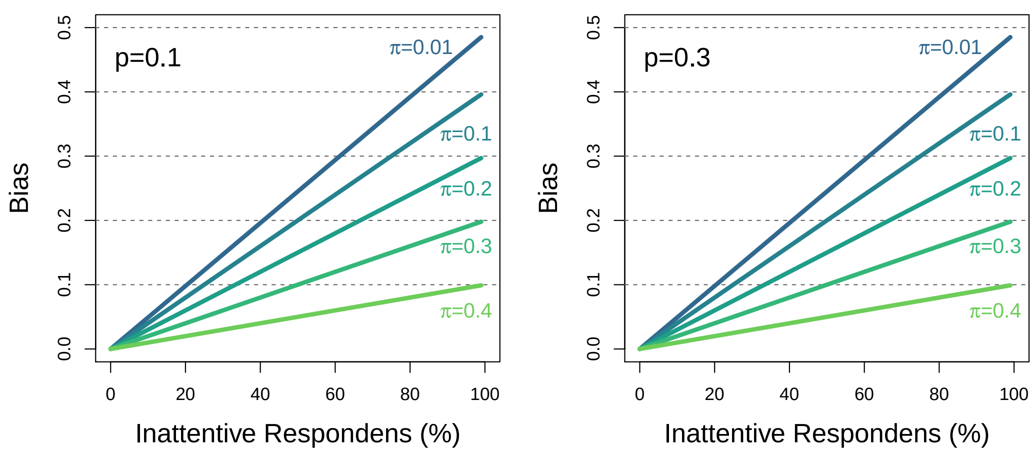

Figure A.1: **Bias in the Naïve Crosswise Estimator.** *Note*: Both panels display the (theoretical) bias in the conventional crosswise estimator with varying levels of inattentive respondents. The the bias increases as the ground truth approaches 0.5, while the value of p does not affect the size of bias.

A.4 Derivation of the Variance

Here, we derive the population and sample variance of the bias-corrected crosswise estimator discussed in Section 3. Rearranging Equation (2b), we obtain

$$
\begin{aligned}
\mathbb{V}(\widehat{\pi}_{BC}) &= \mathbb{V}\left[\frac{\widehat{\lambda} - (1-p)\widehat{\gamma} - \frac{1}{2}(1-\widehat{\gamma})}{(2p-1)\widehat{\gamma}}\right] \\
&= \frac{1}{(2p-1)^2}\mathbb{V}\left[\frac{\widehat{\lambda} - (1-p)\widehat{\gamma} - \frac{1}{2}(1-\widehat{\gamma})}{\widehat{\gamma}}\right] \\
&= \frac{1}{(2p-1)^2}\mathbb{V}\left[\frac{\widehat{\lambda}}{\widehat{\gamma}} - (1-p) - \frac{1}{2\widehat{\gamma}} + \frac{1}{2}\right] \\
&= \frac{1}{(2p-1)^2}\mathbb{V}\left[\frac{2\widehat{\lambda}-1}{2\widehat{\gamma}}\right] \\
&= \frac{1}{(2p-1)^2}\mathbb{V}\left[(2\widehat{\lambda}-1)\left(\frac{\frac{1}{2}-p}{2\widehat{\lambda}'-1}\right)\right] \quad \text{(By Equation (3b))} \\
&= \frac{(\frac{1}{2}-p)^2}{(2p-1)^2}\mathbb{V}\left[\frac{2\widehat{\lambda}-1}{2\widehat{\lambda}'-1}\right] \\
&= \frac{4(\frac{1}{2}-p)^2}{(2p-1)^2}\mathbb{V}\left[\frac{\widehat{\lambda}}{\widehat{\lambda}'-\frac{1}{2}}\right] \\
&= \mathbb{V}\left[\frac{\widehat{\lambda}}{\widehat{\lambda}'-\frac{1}{2}}\right]
\end{aligned}
$$

To see that no analytical form is available, observe that the variance term can be rewritten as:

$$
\begin{aligned}
&\mathbb{V}\left[\frac{\widehat{\lambda}}{\widehat{\lambda}'-\frac{1}{2}}\right] \\
=&\mathbb{E}\left[\frac{\widehat{\lambda}^2}{(\widehat{\lambda}'-\frac{1}{2})^2}\right] - \left(\mathbb{E}\left[\frac{\widehat{\lambda}}{\widehat{\lambda}'-\frac{1}{2}}\right]\right)^2
\end{aligned}
$$

Now, by Assumption 4, we can expand the first term as:

$$
\begin{aligned}
\mathbb{E}\left[\frac{\widehat{\lambda}^2}{(\widehat{\lambda}'-\frac{1}{2})^2}\right] &= \mathbb{E}[\widehat{\lambda}^2] \times \mathbb{E}\left[\frac{1}{(\widehat{\lambda}'-\frac{1}{2})^2}\right] \\
&= \left(\frac{\lambda(1-\lambda)}{N} + \lambda\right)\sum_{n=0}^{N}\frac{1}{\left(n-\frac{1}{2}\right)^2}\binom{N}{n}(k')^n(1-k')^{N-n} \\
&= \left(\frac{\lambda(1-\lambda)}{N} + \lambda\right)(1-k')^N\sum_{n=0}^{N}\binom{N}{n}\frac{1}{\left(n-\frac{1}{2}\right)^2}\left(\frac{k'}{1-k'}\right)^n
\end{aligned}
$$

Similarly, the second term can be expanded as:

$$
\begin{aligned}
\left(\mathbb{E}\left[\frac{\widehat{\lambda}}{\widehat{\lambda}' - \frac{1}{2}}\right]\right)^2 &= \left(\mathbb{E}[\widehat{\lambda}] \times \mathbb{E}\left[\frac{1}{\widehat{\lambda}' - \frac{1}{2}}\right]\right)^2 \\
&= \left(\lambda \sum_{n=0}^{N} \frac{1}{(n-\frac{1}{2})} \binom{N}{n} (k')^n (1-k')^{N-n}\right)^2 \\
&= \left(\lambda (1-k')^N \sum_{n=0}^{N} \binom{N}{n} \frac{1}{(n-\frac{1}{2})} \left(\frac{k'}{1-k'}\right)^n\right)^2
\end{aligned}
$$

Combining both results yields the population variance,

$$
\begin{aligned}
\mathbb{V}(\widehat{\pi}_{BC}) = &\left(\frac{\lambda(1-\lambda)}{N}\lambda\right)(1-k')^N \sum_{n=0}^{N} \binom{N}{n} \frac{1}{(n-\frac{1}{2})^2} \left(\frac{k'}{1-k'}\right)^n \\
&+ \left(\lambda (1-k')^N \sum_{n=0}^{N} \binom{N}{n} \frac{1}{(n-\frac{1}{2})} \left(\frac{k'}{1-k'}\right)^n\right)^2
\end{aligned}
$$

and its sample analog,

$$
\begin{aligned}
\widehat{\mathbb{V}(\widehat{\pi}_{BC})} = &\left(\frac{\widehat{\lambda}(1-\widehat{\lambda})}{N}\lambda\right)(1-k')^N \sum_{n=0}^{N} \binom{N}{n} \frac{1}{(n-\frac{1}{2})^2} \left(\frac{k'}{1-k'}\right)^n \\
&+ \left(\widehat{\lambda} (1-k')^N \sum_{n=0}^{N} \binom{N}{n} \frac{1}{(n-\frac{1}{2})} \left(\frac{k'}{1-k'}\right)^n\right)^2
\end{aligned}
$$

No analytical form is available for these functions.

A.5 Unbiasedness of the Naïve Estimator

To see that the naïve estimator is unbiased when $\gamma = 1$, let Y_i be a binary random variable denoting whether respondent i chooses the crosswise item (i.e., TRUE-TRUE or FALSE-FALSE) and its realization $y_i \in \{0, 1\}$. Let the number of respondents choosing the crosswise item be $k = \sum_{i=1}^{N} y_i$, where $k < n$. Then, the likelihood function for λ given any observed k is $L(\lambda|n, k) = \binom{n}{k}\lambda^k(1-\lambda)^{n-k}$. Applying the first-order condition yields a maximum likelihood estimate (MLE) of λ, $\widehat{\lambda} = \frac{k}{n}$, where $\mathbb{E}[\widehat{\lambda}] = \lambda$. The unbiasedness follows from the fact that $\mathbb{E}[\widehat{\lambda}] = \mathbb{E}\left[\frac{k}{n}\right] = \frac{1}{n}\mathbb{E}[k] = \frac{1}{n}n\lambda = \lambda$. Following the parameterization invariance property of MLEs, $\mathbb{E}[\widehat{\pi}_{CM}] = \pi$.

A.6 Unbiasedness of $\widehat{\lambda}'$

To see that $\widehat{\lambda}'$ is an unbiased estimator of λ', let us define that $\widehat{\lambda}'$ is a binomial random variable (like $\widehat{\lambda}$) with parameters n, λ' and $\widehat{\lambda}' = k'/n$, where k' is the number of people who choose the crosswise item in the anchor question. This is because $k' \sim \text{Binom}(n, \lambda')$ and $\widehat{\lambda}' = k'/n$ suggests $\widehat{\lambda}' \sim \text{Binom}(n, k')$. The probability mass function that $\widehat{\lambda}'$ taking n'/n is given by $\Pr(\widehat{\lambda}' = \frac{n'}{n}) = \binom{n}{n'}(k')^{n'}(1-k')^{n-n'}$.

B Additional Simulation Studies

We perform another simulation study to confirm that the bias-corrected estimator is not susceptible to a potential correlation between inattentiveness and possession of sensitive attributes. The concern is that if respondents with sensitive attributes are more likely to be inattentive, correcting for bias caused by inattentive responses might affect the estimate of the prevalence rate of sensitive attributes in the *population* (which includes both attentive and inattentive respondents). The results shown in Figure B.1 resolve this concern. In each panel, we plot the bias-corrected point and interval estimates over the true prevalence rate against varying levels of inattentiveness in hypothetical data. Here, we fix the true prevalence rate among attentive respondents ($\pi_{\text{attentive}} = 0.1$) while varying the true prevalence rates among inattentive respondents in four panels ($\pi_{\text{inattentive}} \in \{0.2, 0.3, 0.4, 0.5\}$). The true prevalence rates are plotted in red straight lines. Notice that the true prevalence rate in each panel is now a linear combination (convex combination) of $\pi_{\text{attentive}}$ and $\pi_{\text{inattentive}}$ which are denoted by red and blue dots. (That is, $\pi = \pi_{\text{attentive}} * \gamma + \pi_{\text{inattentive}} * (1 - \gamma)$, where γ is the proportion of attentive responses.) Nevertheless, Figure B.1 show that our bias-corrected estimator properly captures the true prevalence rate of sensitive attributes regardless of the degrees of correlation between inattentiveness and possession of sensitive attributes.

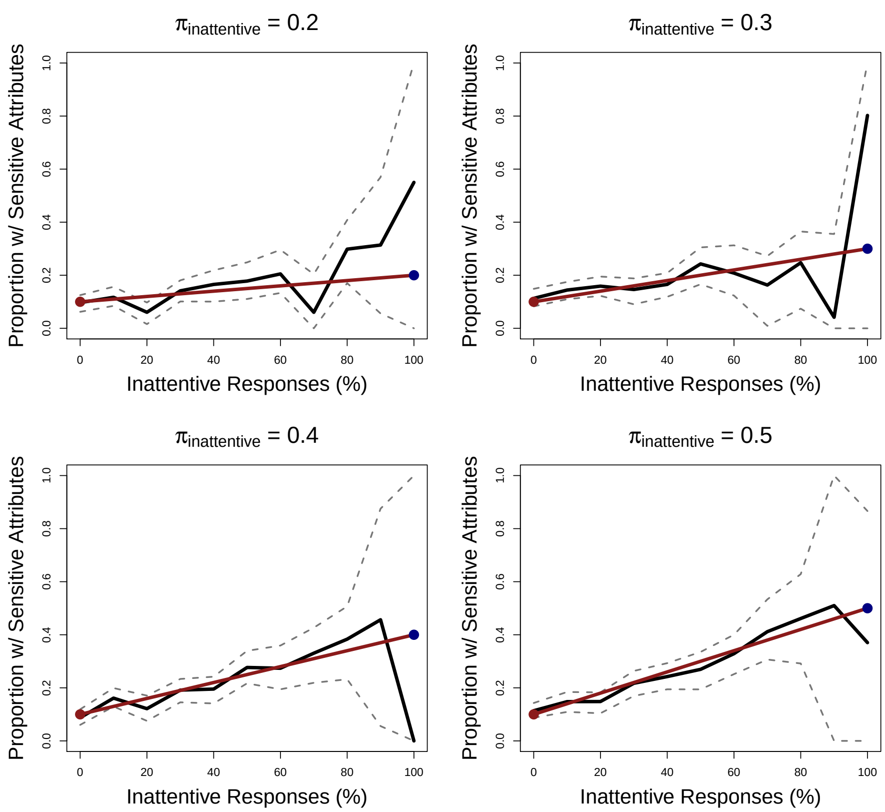

Figure B.1: **Correlation between Attentiveness and Possession of Sensitive Attributes.** *Note*: This graph illustrates the bias-corrected estimates (black lines with gray dashed lines) with the true prevalence rates (red straight lines) when the true prevalence rates are convex combinations of $\pi_{\text{attentive}}$ and $\pi_{\text{inattentive}}$, denoted by red and blue dots, respectively. The bias-corrected estimates properly capture the true prevalence rates regardless of the degrees of deviation of $\pi_{\text{inattentive}}$ from $\pi_{\text{attentive}}$. While $\pi_{\text{attentive}}$ is fixed at 0.1, $\pi_{\text{inattentive}}$ varies. The data is generated by setting $n = 2000, p = 0.15, p' = 0.15$.

C Additional Information and Simulations for Extensions

In this section, we present additional information for the proposed extensions of the bias-corrected estimator.

C.1 Weighting Method

The proof is straightforward. Assuming that $Y_i \perp\!\!\!\perp S_i|X$ (choosing the crosswise item and being in the sample are statistically independent conditional upon a covaraite), weighting can recover the population crosswise proportion λ from the sample crosswise response Y_iS_i:

$$
\begin{aligned}
&\mathbb{E}\left[\frac{Y_iS_i}{\Pr(S_i=1|X)}\right]\\
=&\mathbb{E}\left[\mathbb{E}\left[\frac{Y_iS_i}{\Pr(S_i=1|X)}\middle|X\right]\right] \quad \text{(Iterative Expectation)}\\
=&\mathbb{E}\left[\frac{\mathbb{E}[Y_i|X]\mathbb{E}[S_i|X]}{\Pr(S_i=1|X)}\right] \quad \text{(Conditional Independence)}\\
=&\mathbb{E}\left[\frac{\mathbb{E}[Y_i|X]\Pr(S_i=1|X)}{\Pr(S_i=1|X)}\right] \quad \text{(Definition of Expectation)}\\
=&\mathbb{E}[\mathbb{E}[Y_i|X]]\\
=&\mathbb{E}[Y_i] \quad \text{(Iterative Expectation)}\\
=&\lambda
\end{aligned}
$$

Similarly, we can show that

$$
\begin{aligned}
&\mathbb{E}\left[\frac{A_iS_i}{\Pr(S_i=1|X)}\right]\\
=&\mathbb{E}\left[\mathbb{E}\left[\frac{A_iS_i}{\Pr(S_i=1|X)}\middle|X\right]\right]\\
=&\mathbb{E}[A_i]\\
=&\lambda'
\end{aligned}
$$

In practice, researchers can calculate weights using their favorite weighting techniques such as raking (or iterative proportional fitting), matching, propensity score weighting, or sequential applications of these. Recent research shows that "when it comes to accuracy, choosing the right variables for weighting is more important than choosing the right statistical method" (Mercer, Lau and Kennedy, 2018, 4). Thus, we recommend that researchers think carefully about the association between the sensitive attribute of interest and basic demographic and other context-dependent factors when using weighting. To choose the "right" variables, our proposed regression models can also be useful exploratory aids. When generalizing the results on sensitive attributes to a larger population, however, it is strongly advised to elaborate on how weights are constructed and what potential bias may exist (Franco et al., 2017).

Another possible approach to deal with highly selected samples is to employ multilevel regression and post-stratification (MRP) (Downes et al., 2018). While we do not consider MRP with crosswise estimates in this article, future research should explore the optimal strategy to use MRP in sensitive inquiries.

To illustrate our weighting strategy, we simulate crosswise data with two covariates: $X_1 \sim \text{Binomial}(0.5)$ and $X_2 \sim \text{Poisson}(30)$ for, let's say, 100,000 voters. Specifically, we simulate the true prevalence rates in the crosswise and anchor questions according to the following generative models:

$$\pi = \frac{\exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2)}{1 + \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2)}$$
$$\gamma = \frac{\exp(\theta_0 + \theta_1 X_1 + \theta_2 X_2)}{1 + \exp(\theta_0 + \theta_1 X_1 + \theta_2 X_2)},$$

where we set $\beta_0 = -1.5$, $\beta_1 = 0.5$, $\beta_2 = 0.02$ and $\theta_0 = 2$, $\theta_1 = -0.1$, $\theta_2 = -0.01$. We then simulate the crosswise data according to Equation (1c). For example, we can consider X_1 as a binary indicator for being female (as opposed to non-female) and female voters are more likely to have sensitive traits than non-female voters (i.e., $\beta_1 = 0.5$).

Under this generative model, the population-level proportion of individuals with sensitive traits is **0.35** (and the population proportion is **0.40** for female and **0.29** for non-female voters). Now, from the population of 100,000 voters, we sample 1000, 2000, 3000, 4000, and 5000 individuals. In this process, we intentionally oversample female voters with probability 0.7. Consequently, we obtain sample weights 1.43 for female voters and 3.33 for non-female voters. We then generate bias-corrected crosswise estimates with and without incorporating the sample weights. The results are presented in Figure 6.

C.2 Simulations for Crosswise Regressions: Sensitive Trait as an Outcome

To validate this regression framework, we simulate crosswise data with two covariates: $X_1 \sim \text{Binomial}(0.5)$ and $X_2 \sim \text{Poisson}(30)$. Specifically, we simulate the true prevalence rates in the crosswise and anchor questions according to the following generative models:

$$\pi = \frac{\exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2)}{1 + \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2)}$$
$$\gamma = \frac{\exp(\theta_0 + \theta_1 X_1 + \theta_2 X_2)}{1 + \exp(\theta_0 + \theta_1 X_1 + \theta_2 X_2)},$$

where we set $\beta_0 = -1.5$, $\beta_1 = 0.5$, $\beta_2 = 0.02$ and $\theta_0 = 2$, $\theta_1 = -0.1$, $\theta_2 = -0.01$. We then simulate the crosswise data according to Equation (1c).

Finally, we estimate the crosswise regression with the latent sensitive trait as the outcome variable. Figure C.1 displays the estimated parameters and confidence intervals with different sample sizes. The results suggest that the proposed model and estimation strategy can recover the true parameters (asymptotically). It is also straightforward to compute predicted probabilities of having a sensitive attribute with 95% confidence intervals using the parametric bootstrap. Figure C.2 displays the predicted probabilities of having sensitive attributes in this particular simulation with different values for X_1.

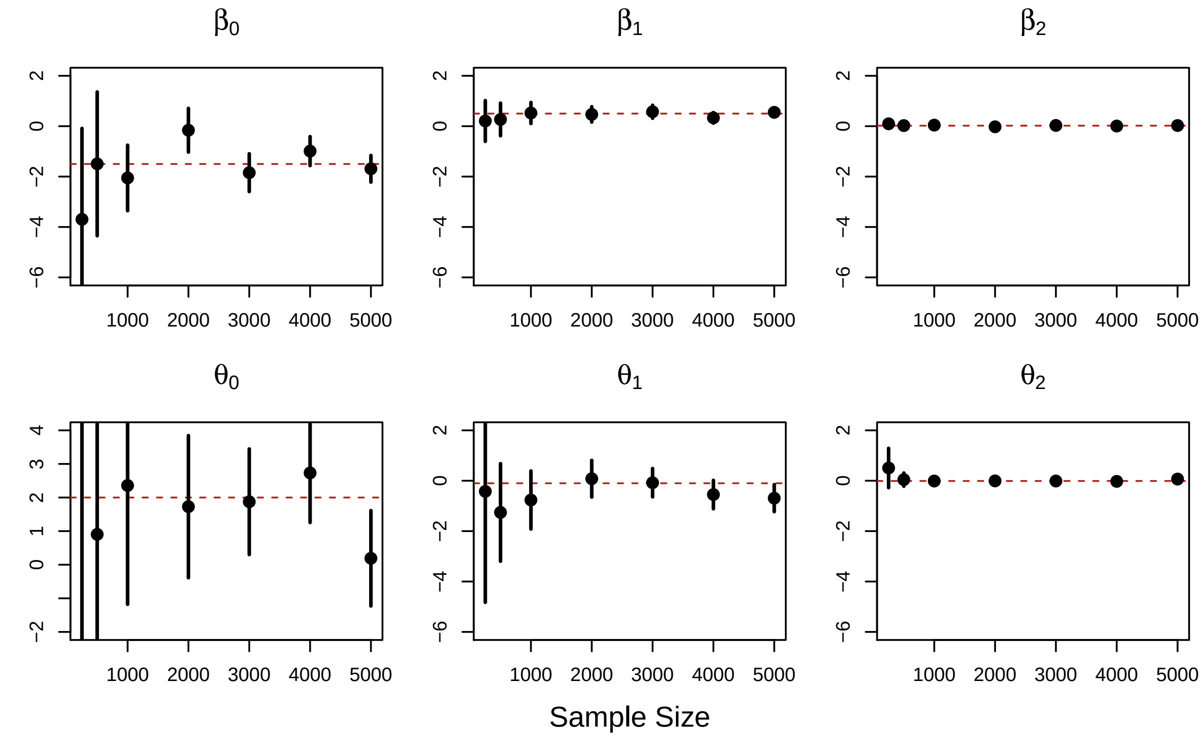

Figure C.1: **Finite Sample Performance of Regression Estimator (Sensitive Trait as a Predictor).** *Note*: Regression estimates of six parameters in simulated data. The dashed lines indicate the true values for the parameters.

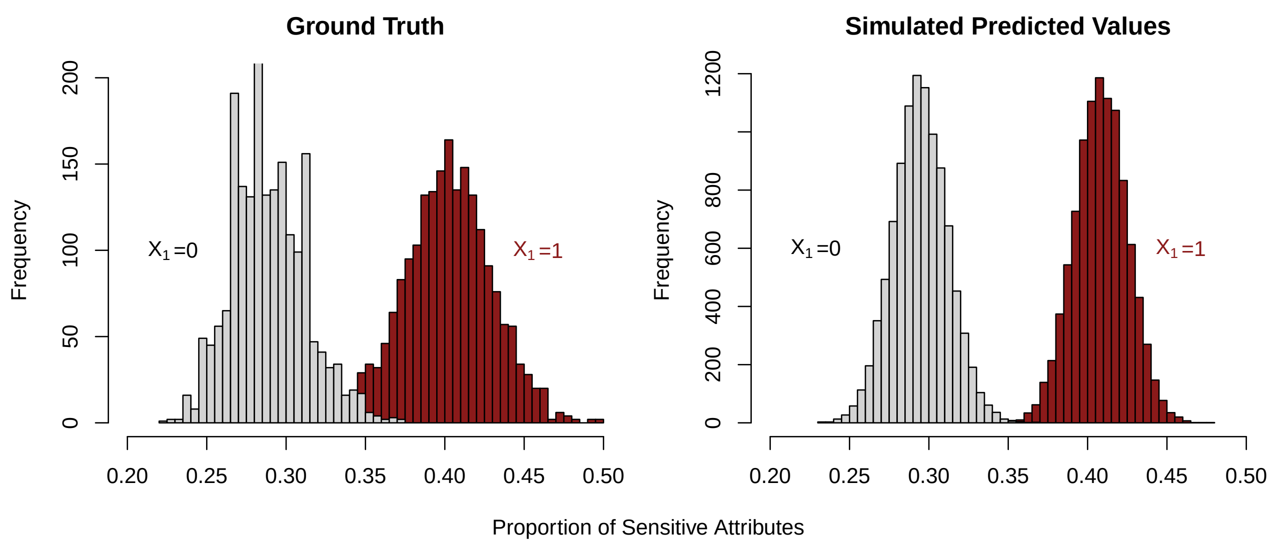

Figure C.2: **True and Predicted Proportions of Sensitive Attributes** *Note*: This graph visualizes the empirical distribution of the probability of having sensitive attributes for $X_1 = 0$ and $X_1 = 1$ based on the ground truth with $N = 4000$ (left panel) and simulated predicted values (right panel).

C.3 Simulations for Crosswise Regressions: Sensitive Trait as a Predictor

To validate the proposed framework, we simulate crosswise data with two covariates as in Online Appendix C.2. We then simulate the response variable according to the following generative model:

$$V_i = \gamma_0 + \gamma_1 X_1 + \gamma_2 X_2 + \delta Z_i + \epsilon_i,$$

where we set $\gamma_0 = 0$, $\gamma_1 = 0.3$, $\gamma_2 = 0.01$, $\delta = 1$, and $\epsilon_i \sim N(0, 1)$. Recall that Z_i is a latent variable for having a sensitive trait and we cannot observe its value directly (and thus crosswise data do not contain Z_i).

We then estimate the above crosswise regression model with the simulated observed outcome and crosswise data. Figure C.3 shows the estimates for our quantity of interest with different sample size. It demonstrates that the proposed regression model and estimation strategy can recover the latent magnitude of the association between the latent sensitive trait and the response variable ($\delta = 1$). It also shows that other ten parameters can be properly estimated by the proposed regression model. It is also straightforward to compute predicted values of the outcome variable with 95% confidence intervals using the parametric bootstrap. Figure C.4 displays the predicted values of the outcome in this particular simulation with different values for Z.

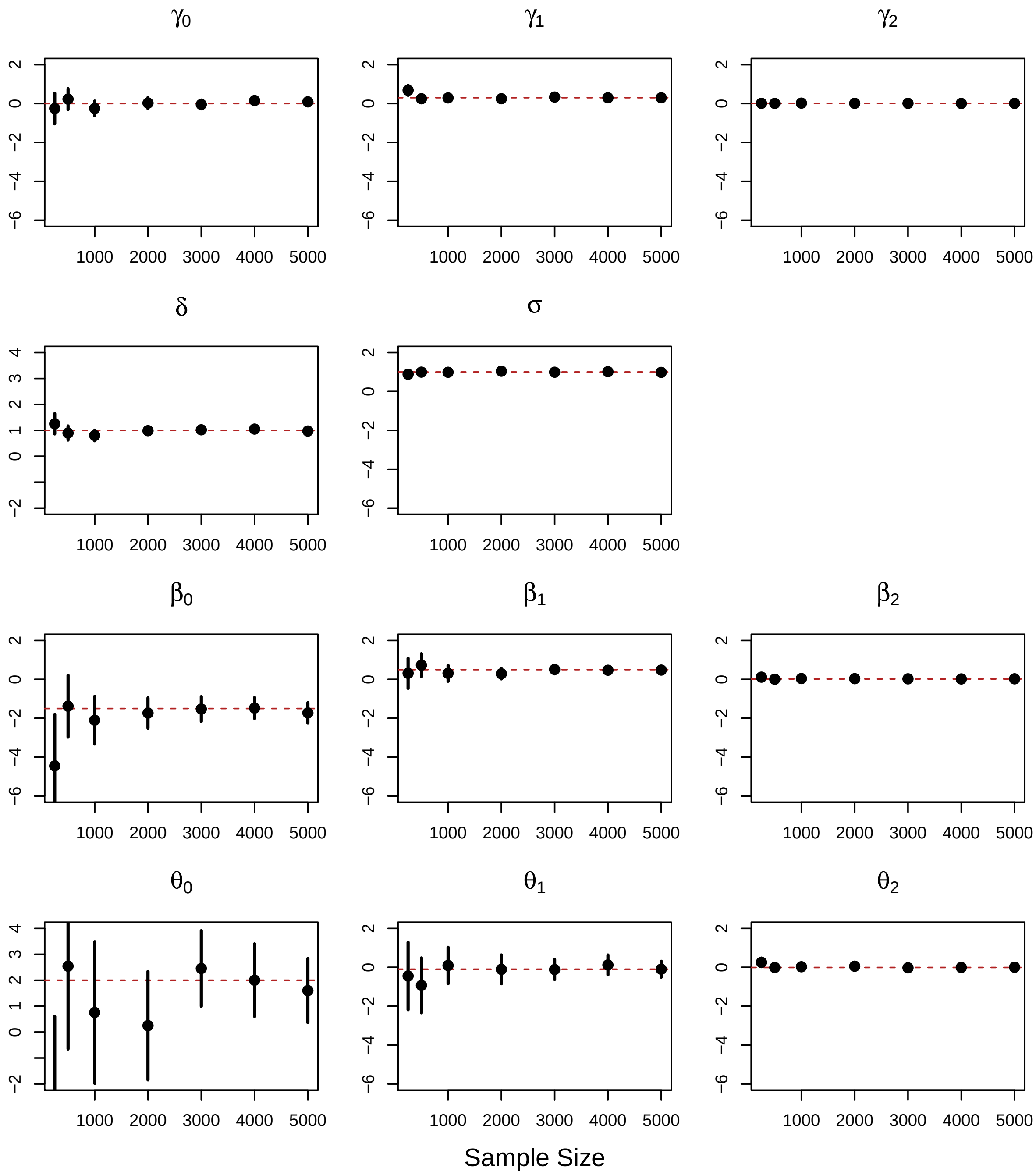

Figure C.3: **Finite Sample Performance of Regression Estimator (Sensitive Trait as a Predictor)**. *Note*: The dashed lines indicate the true values for the parameters.

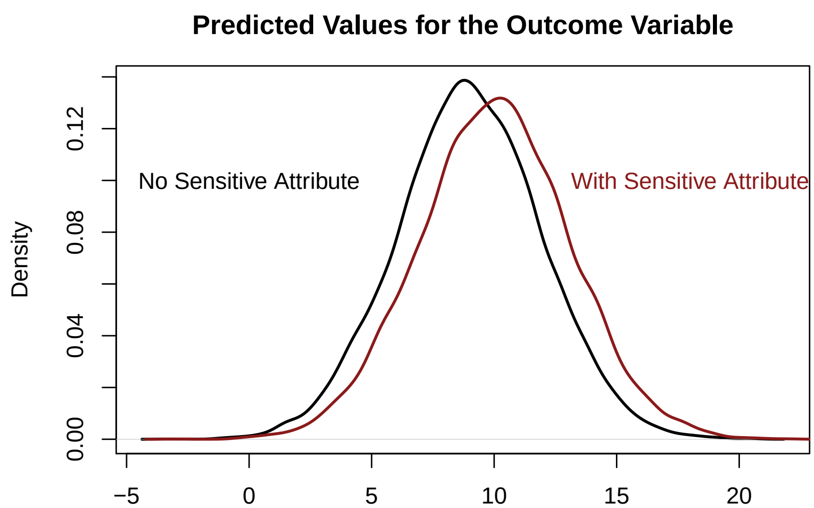

Figure C.4: **Simulated Outcome Values with and without the Sensitive Attribute.** *Note*: This graph visualizes the density of simulated (predicted) values for the outcome variable in the absence (left density) and presence (right density) of the sensitive attribute.

D Empirical Illustration

In this section, we illustrate the proposed methodology by using survey data about the behavior of paid survey takers. We ran an online survey through Qualtrics asking respondents about their past behavior as paid survey takers. Specifically, we asked whether they have (1) speeded through questions without reading, (2) made up answers, and (3) lied about their qualifications. It was emphasized that the survey was specifically about the behavior of paid survey takers. We did so to create a normative environment that admitting the behaviors in (1) to (3) becomes a sensitive response because as paid survey takers they are not supposed to do any of the three "unethical" items.

For our anchor question, we asked whether respondents were taking the current survey somewhere outside the United States. We chose this anchor item because we know that all survey takers in Qualtrics are sampled from survey takers who are living in the U.S. and the topic is closely related to our sensitive items of interest. For auxiliary probabilities, we asked respondents to list five people they know as well as their birth months in the beginning of the crosswise questions. We took this approach to make sure that respondents will not be distracted from answering the crosswise questions of interest by performing these additional tasks simultaneously. We then randomly assign respondents different auxiliary probabilities of 0.086 and 0.25, which we call *low* and *moderate* auxiliary probabilities. Along with the crosswise model, we also performed "direct questioning" on the same sensitive items.

We first apply the proposed bias-correction to our data and obtain point and uncertainty estimates for the prevalence proportions of interest. We also estimate the prevalence rates based on direct inquiry and the naïve crosswise estimator. The results are demonstrated in Figure D.1. For crosswise estimates, dots (second and fourth from the left) are based on low auxiliary probabilities ($p = 0.086$) and asterisks are based on moderate auxiliary probabilities ($p = 0.25$). It is shown that bias-corrected estimates are generally higher than direct inquiry estimates, but lower than naïve crosswise estimates. Estimated standard errors are wider for bias-corrected estimates than for naïve crosswise estimates due to the additional uncertainty for estimating attentive rates. By the construction of crosswise estimates, uncertainty is larger for estimates based on higher auxiliary probabilities, which suggests that researchers will be benefited from using low auxiliary probabilities whenever possible.

Importantly, without bias-correction, researchers may mistakenly infer that the crosswise model induced more candid answers on sensitive items (i.e., direct inquiry and naïve estimates are statistically significantly different in most cases) even though such differences are artificially caused by the presence of inattentive responses. Our methodology exactly prevents this form of incorrect inferences.

References for Online Appendix

Blair, Graeme, Kosuke Imai and Yang-Yang Zhou. 2015. "Design and analysis of the randomized response technique." *Journal of the American Statistical Association* 110(511):1304–1319.

Downes, Marnie, Lyle C Gurrin, Dallas R English, Jane Pirkis, Dianne Currier, Matthew J Spittal and John B Carlin. 2018. "Multilevel Regression and Poststratification: A Modeling Approach to Estimating Population Quantities From Highly Selected Survey Samples." *American journal of epidemiology* 187(8):1780–1790.

Franco, Annie, Neil Malhotra, Gabor Simonovits and LJ Zigerell. 2017. "Developing standards for post-hoc

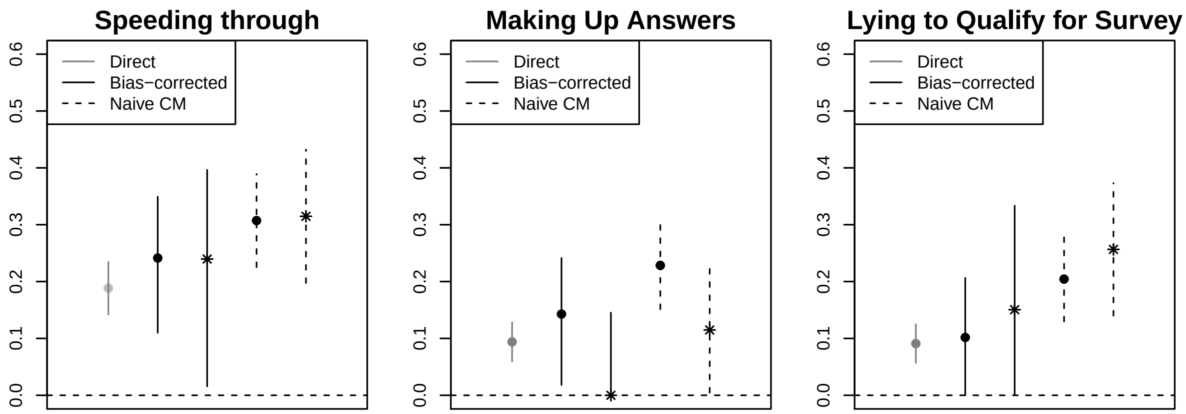

Figure D.1: **Comparison of Prevalence Estimates.** *Note*: This graph visualizes the estimated prevalence of sensitive attributes based on direct inquiry, bias-corrected estimator, and naïve crosswise estimator. For crosswise estimates, dots (second and fourth from the left) are based on low auxiliary probabilities ($p = 0.086$) and asterisks (third and fifth from the left) are based on moderate auxiliary probabilities ($p = 0.25$).

weighting in population-based survey experiments." *Journal of Experimental Political Science* 4(2):161–172.

Mercer, Andrew, Arnold Lau and Courtney Kennedy. 2018. "For Weighting Online Opt-In Samples, What Matters Most." *Pew Research Center* .