Bias-Corrected Crosswise Estimators for Sensitive Inquiries*

Yuki Atsusaka[†] Randolph Stevenson[‡]

atsusaka@rice.edu randystevenson@rice.edu

August 5, 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 attentiveness. We also offer several extensions including a sensitivity analysis for conventional crosswise estimates, a strategy for weighting, and a framework for multivariate regressions in which the latent sensitive trait can be used as the outcome or as a predictor while applying the bias correction. We illustrate our methodology by simulation studies and empirical examples and further provide a practical guide for designing surveys to enable our proposed bias correction. All of our method can be easily implemented through an open-source software `cWise`.

Keyword: Crosswise model, sensitive questions, survey methodology, inattentive respondents, sensitivity analysis, weighting, regression.

Word Count: 8368

---

*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 thank 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 in the Department of Political Science at Rice University, MS-24 105 Herzstein Hall, 6100 Main St, Houston, TX 77005 (https://atsusaka.wordpress.com/).

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

1 Introduction

Social scientists and epidemiologists often use surveys to probe significantly important, but highly sensitive topics that respondents may hesitate to answer truthfully. Such topics include 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), or illicit drug use (Shamsipour et al., 2014). For these topics (and many others), we often expect many respondents to shade their answers in a socially desirable direction. To mitigate this concern, the literature has proposed various survey techniques (e.g., randomized response techniques, list experiments, and endorsement experiments) which are designed to encourage respondents to offer truthful answers by protecting their privacy (Blair, Imai and Zhou, 2015; Rosenfeld, Imai and Shapiro, 2016).

Among these designs, the *crosswise model* is an increasingly popular technique to elicit candid answers from survey respondents on sensitive questions (Yu, Tian and Tang, 2008; Jann, Jerke and Krumpal, 2011; Höglinger, Jann and Diekmann, 2016). Instead of directly asking respondents about a sensitive topic, the crosswise model shows respondents two different statements (on sensitive and non-sensitive topics) and ask whether only one of them is true. Because respondents do not need to reveal anything about the sensitive topic *per se* (and thus their privacy is perfectly protected), it is expected that they offer candid answers that they would never provide in direct questioning. Using observed responses and our prior knowledge about the non-sensitive topic, the crosswise model then enables analysts to estimate the population proportion of individuals who possess sensitive attributes. Although successful applications of the crosswise model require respondents to fully understand and follow its specific instruction, it has been reported that about 2 to 28% of respondents are not paying attention during surveys using this design (Schnapp, 2019; Höglinger and Diekmann, 2017; Höglinger and Jann, 2018). Existing studies, however, have not delineated statistical consequences of having such *inattentive respondents* and have either suggested that we drop such respondents entirely from data or adjust estimates based on a crude measure of attentiveness, leaving researchers suboptimal solutions.

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 instruction and randomly select a given choice under the design). This is highly 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 and opinion and so is "working" as expected. In other words, researchers may mistakenly conclude that, by using the crosswise model, they 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 the bias without obtaining any individual-level data for attentiveness. Our procedure only requires researchers to include what we call an *anchor question* that resembles the question asked in the crosswise model to the survey, while making several assumptions that can be easily satisfied at the design-stage of survey. In our Online Appendix, we show how to design surveys so that researchers can effectively apply our bias correction procedure.

After validating our method in simulation studies, we also offer several useful extensions of our bias-corrected estimator including a sensitivity analysis, weighting strategy, and framework for regression analysis in which the latent sensitive trait can be used either as an outcome or as a predictor. The sensitivity analysis enables researchers to apply our bias correction to surveys even when the anchor question is not available and examine how sensitive their original crosswise estimates are to the possible presence of inattentive respondents. To illustrate this method, we apply the sensitivity analysis to six published studies based on crosswise estimates (Jann, Jerke and Krumpal, 2011; Korndörfer, Krumpal and Schmukle, 2014; Shamsipour et al., 2014; Höglinger, Jann and Diekmann, 2016; Vakilian, Mousavi and Keramat, 2014), concluding that the original findings — crosswise estimates are higher than direct questioning estimates — might be mostly artifacts of inattentive respondents.

We also propose a simple strategy to incorporate sample weights to our bias-corrected estimator to help analysts extend their inferences into a larger population of interest; this is critical in the context of sensitive inquiries because most sensitive questions are asked in surveys with unrepresentative samples such as online opt-in samples. Finally, we introduce a regression framework that enables us to examine the associations between the latent or unobserved (individual-level) sensitive trait (as a response variable or an explanatory variable) and other covariates or outcome variables. Using this framework, for example, political scientists can estimate the prevalence of corruption in a legislature, analyze what kinds of politicians are more likely to engage in corruption, and whether engaging in corruption is associated with their reelection. Importantly, these extensions are based on crosswise estimates after bias correction is properly applied.

In what follows, we first introduce the basic setup of the crosswise model, discuss the presence and consequence of inattentive respondents when using the design, and describe previously proposed solutions to them. We then formally illustrate the bias in the conventional crosswise estimator, elaborate our bias correction procedure using the anchor question, and clarify several assumptions that are required to enable our bias correction. After demonstrating the method by simulations, we discuss the three extensions. 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 Crosswise Model and its Pitfall

Before introducing our bias-corrected estimator, we briefly describe how the crosswise model works and elaborate the potential problems caused by inattentive respondents.

2.1 Crosswise Model

Yu, Tian and Tang (2008) proposed the crosswise model building on a class of randomized response techniques for sensitive inquiries (Warner, 1965; Blair, Imai and Zhou, 2015; Fox, 2015). The crosswise model makes inquiries of the following form.

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 most appropriately describes your case?

- Both statements are TRUE, or both statements are FALSE $\rightsquigarrow$ **Crosswise Item**
- Otherwise

Here, Statement A is a sensitive statement that researchers would have wanted to ask directly if there had not been social desirability bias. Our quantity of interest, then, is the population proportion of individuals who agree with (or fit the description in) Statement A. Usually, this proportion is less than 0.5 and even close to 0 and direct questioning is expected to create a social desirability bias toward 0 (i.e., most people do not want to report even if they fit the description).[1] In contrast, Statement B is a non-sensitive auxiliary statement whose population prevalence is *ex ante* known to researchers.

Now, after seeing both statements, respondents are asked to answer whether they are both TRUE or both FALSE in the Question. We call this choice as the *crosswise item* since this condition corresponds to the two gray cells on diagonal of Table 1. Because they do not need to reveal whether they fit the description in Statement A *individually* (and it is never known to interviewers), it is expected that they answer truthfully to this question without being affected by social desirability bias. Importantly, the only information we obtain from this survey is the proportion at which respondents choose the crosswise item or what we call the *crosswise proportion*: $\mathbb{P}$(TRUE-TRUE ∪ FALSE-FALSE). The key idea in this design is that the crosswise proportion is a combination of the proportions for Statement A and Statement B being TRUE, respectively, and using the crosswise proportion and prior knowledge about the non-sensitive statement, researchers can "reverse-engineer" the quantity of interest.

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

Table 1: **The Crosswise Model.** *Note*: The crosswise model ask respondents to reveal whether they correspond to the gray cells (on diagonal) or white cells (off diagonal).

To demonstrate how this design works, we provide a simple numerical example. Suppose that the population proportion for Statement B is known to be 0.25 and the true crosswise proportion is 0.65. We can then express that: $\mathbb{P}(\text{TRUE-TRUE} \cup \text{FALSE-FALSE}) = \mathbb{P}(\text{A=TRUE}) \times \mathbb{P}(\text{B=TRUE}) + \mathbb{P}(\text{A=FALSE}) \times \mathbb{P}(\text{B=FALSE}) = \mathbb{P}(\text{A=TRUE}) \times 0.25 + \mathbb{P}(\text{A=FALSE}) \times 0.85 = 0.65$. Consequently, if we know the probability for Statement B and the crosswise proportion, we can easily reverse engineer the proportion for Statement A as $\mathbb{P}(\text{A=TRUE}) = \frac{0.65-0.85}{2\times 0.25-1} = 0.2$. The next section offers a full description of the design more formally.

Among other survey methods for sensitive inquiries, the crosswise model is considered to be the most

[1] 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 than 0.5 (e.g., most people say Yes to the statement), we can always flip the direction of inquiries before (i.e., including "NOT" clause) or after the surveys (i.e., subtracting the quantity of interest from 1).

effective in terms of reducing social desirability bias because, under this design, respondents do not need to confront the sensitive topic *per se* (Yu, Tian and Tang, 2008). This is because respondents are never asked directly about the sensitive statement and respondents who understand the design have no incentive to answer untruthfully. It is worth noting, however, that this theoretical feature of the crosswise model is only attained when all respondents carefully read, fully understand, and follow the somewhat complex instruction.

2.2 Inattentive Respondents (and Why We Should Care)

In surveys, inattentive respondents are pervasive, and many researchers rely on different attention checks to detect any careless and inattentive respondents during surveys. This is especially true in opt-in online surveys where it is estimated that at least 8% to 12% of survey takers are inattentive (Crump, McDonnell and Gureckis, 2013; Meade and Craig, 2012; Maniaci and Rogge, 2014). Researchers have found that inattentive responses are also common in surveys using the crosswise model (Schnapp, 2019), estimating that 2% and 12% (Höglinger and Diekmann, 2017), 28% (Höglinger and Jann, 2018), and 13% (Enzmann, 2017, cited in Schnapp (2019)) of respondents chose the crosswise item randomly.[2] In this article, we explicitly define inattentive respondents as *respondents who randomly pick the crosswise item or guess their answers in the crosswise question*. We assume that no respondent always or never chooses the crosswise item regardless of their true state and that no respondent willfully chooses the "wrong" item against the instruction. For this reason, we do not consider attentiveness of respondents during the entire survey, but focus on inattentive respondents in the crosswise question specifically.

Our first contribution to the literature is to point out that point estimates from the crosswise model are highly sensitive to inattentive respondents (respondents who answer randomly or guess their answers); more specifically, the presence of inattentive respondents creates a bias in point estimates toward 0.5.[3] This is highly problematic in many sensitive inquiries (where we expect social desirability bias toward 0) because this bias could make researchers falsely conclude that they successfully induced higher estimates (i.e., more candid answers) from respondents even when the sensitive attributes of question are in fact less pervasive.

[2]Inattentive respondents could also be present and lead to misleading inferences in other methods for sensitive inquiries (e.g., Blair, Chou and Imai, 2019).

[3]An intuition behind this is that the presence of inattentive respondents makes the crosswise proportion closer to 0.5, and to make logical ends meet it also requires $\mathbb{P}(\mathsf{A}{=}\mathsf{TRUE})$ to approach 0.5. Consider the case where every respondent randomly chooses the crosswise item such that the crosswise proportion becomes 0.5. In this extreme situation, we can show that $\mathbb{P}(\mathsf{A}{=}\mathsf{TRUE}) = \frac{0.5-1+\mathbb{P}(\mathsf{B}{=}\mathsf{TRUE})}{2\mathbb{P}(\mathsf{B}{=}\mathsf{TRUE})-1} = \frac{\mathbb{P}(\mathsf{B}{=}\mathsf{TRUE})-0.5}{2(\mathbb{P}(\mathsf{B}{=}\mathsf{TRUE})-0.5)} = 0.5$.

To illustrate this problem, Figure 1 plots the estimated proportion of sensitive attributes (with its 95% confidence interval) against the percentage of inattentive respondents based on hypothetical (and typical) data from the crosswise model. It portrays that there is always a positive bias in the crosswise estimate and the bias increases as the percentage of inattentive respondents grows. While this "naïve" crosswise estimator is still unbiased under the (rare) condition that every respondent properly follows the instruction (the left arrow), this condition does not meet in almost all survey data. Our primary contribution in this article 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 valid inferences about sensitive topics.

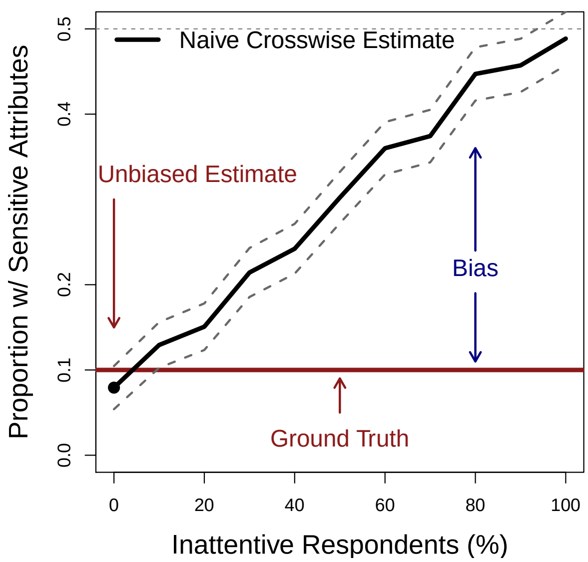

Figure 1: **Impact of Inattentive Respondents on Crosswise Estimates.** *Note*: This plot illustrates that the bias increases as the percentage of inattentive responses grows. Gray dashed lines show confidence intervals.

Although previous research has not actually shown what the bias actually looks like as in Figure 1, 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, 2017; Höglinger and Jann, 2018). However, this approach leads to a biased estimate of the quantity of interest unless researchers assume that attentive respondents are a simple random subsample from the original sample. This is an unreasonably strong assumption in most situations and needs to be empirically tested. Even if the assumption holds, this approach necessarily affects inferences by decreasing relative efficiency.

The second solution is to detect whether respondents answered crosswise questions 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 the direct question is itself not susceptible to the inattentiveness or social desirability bias. But such assumptions are largely 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.

3 The Proposed Methodology

Now, 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 possess sensitive attributes. The main 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 reasonable assumptions.

3.1 The Setup

To simplify the problem, suppose that we consider a single sensitive question of interest in a survey with n respondents who are drawn from a finite population via simple random sampling. Suppose also that we apply the crosswise model to estimate the prevalence of the sensitive traits 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 perform weighting when using non-representative samples.

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

Given these quantities, 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 identify with respect to the quantity of interest yields $\pi = \frac{\lambda+p-1}{2p-1}$. Yu, Tian and Tang (2008) then derived 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 pick the crosswise item or guess their answers in this design. Again, it has been estimated that about 2 to 28% of respondents are inattentive respondents in previous surveys using the crosswise model, and this is a practical survey issue that we need to take seriously. The novelty of our approach is to explicitly incorporate these respondents in the above identity. That is, 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), where γ is assumed to be 1. While the naïve crosswise estimator is unbiased as long as $\gamma = 1$ (Online Appendix A.4), it is no longer an unbiased estimator whenever $\gamma < 1$. More specifically, we can define and quantify the bias in the conventional crosswise estimator as follows (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 the quantity of interest caused by inattentive respondents. Importantly, under several regularity conditions (i.e., typical crosswise settings), the bias term is always positive

(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 2 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 and as the quantity of interest approaches 0, but that it does not change regardless of the value of p.

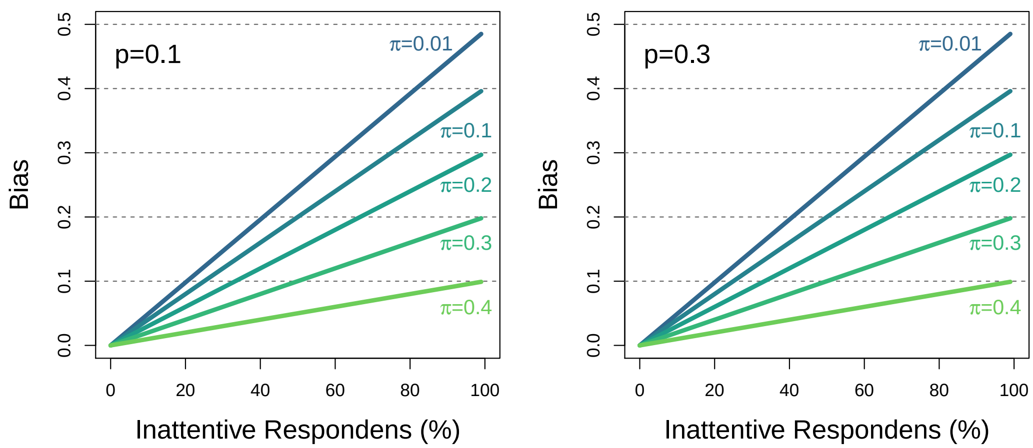

Figure 2: **Bias in the Crosswise Estimator.** *Note*: Both panels display the (theoretical) bias in the conventional crosswise estimator with varying levels of inattentive respondents. The relative size of the bias increases as the ground truth approaches 0.5, while the value of p does not affect the size of bias.

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 observed proportion of attentive respondents; here $\mathbb{E}[\widehat{\lambda}] = \lambda$ and $\mathbb{E}[\widehat{\gamma}] = \gamma$ and thus $\mathbb{E}[\widehat{B}_{CM}] = B_{CM}$.

Now, for our bias-correction to work, we need to establish few assumptions. First, we assume $\kappa = 0.5$; that is, inattentive respondents pick the crosswise item with probability of 0.5. We formally state this assumption as follows:

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

The survey literature tells us that this assumption may not hold in most situations because inattentive respondents (in a more general sense) are 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 careless respondents pick items. This is easily achieved by randomizing the order of the listed items in the Question.

The main challenge in estimating the bias is to obtain an estimated proportion of attentive respondents ($\widehat{\gamma}$). We solve this problem by employing what we call an *anchor question* along with the original crosswise question. The anchor question is a question that looks exactly like the crosswise question. The only difference is that, this time, both statements are about non-sensitive topics and we *ex ante* known the population prevalence for each statement. For example, in a survey administered to a U.S. population (e.g., Qualtrics survey), we can 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**
- Otherwise

Here, Statement C is a non-sensitive anchor statement and we know that the population proportion of individuals who fit the description therein is (supposed to be) 0. Statement D is another non-sensitive statement whose population prevalence is also known to researchers just like 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 $\widehat{\gamma}'$), 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 $\mathbb{E}[\widehat{\lambda}'] = \lambda'$ (Online Appendix A.5).

Finally, our strategy is to use $\widehat{\gamma}'$ (obtained from the anchor question) as a valid 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 two questions. We state this assumption as follow:

Assumption 2 (Attention Consistency). *The population proportion of attentive respondents is constant across the crosswise and anchor questions ($\gamma = \gamma'$).*

While whether this assumption is satisfied is an empirical matter, researchers can design their surveys so that the attention consistency can be achieved (e.g., by randomizing the order of the two questions and making them look alike).

To summarize, our bias-corrected estimator provides an unbiased estimate of the population prevalence of sensitive attributes even when some respondents do not follow the instruction and randomly pick the crosswise item, and it does so with only two and weak assumptions that can be easily satisfied in the design stage of surveys.

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 for the bias-corrected estimator. To simplify the derivation, we first introduce the following assumption:

Assumption 3 (Independent Crosswise Proportions). *The population proportions 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. Formally,* $p \perp\!\!\!\perp p' \Rightarrow \widehat{\lambda} \perp\!\!\!\perp \widehat{\lambda}'$.

With this assumption, we derive the variance of the bias-corrected crosswise estimator and its sample analog as follows (Online Appendix A.3):

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

Note that these variances are necessarily larger than the variances of the conventional estimator. To see why, simply 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 (in addition to the quantity of interest). Since no analytical solution is available for Equations (4a) and (4b), in practice, we employ bootstrapping to construct confidence intervals and perform hypothesis testing.

In Online Appendix E, we lay out several important points to consider when designing the survey in order to satisfy Assumptions 1-3 and effectively apply our bias correction procedure.

4 Simulation Studies

To illustrate and our bias corrected 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 and its 95% confidence interval does not capture it. In contrast, the bias-corrected point estimate is fairly close to the quantity of interest and its 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.

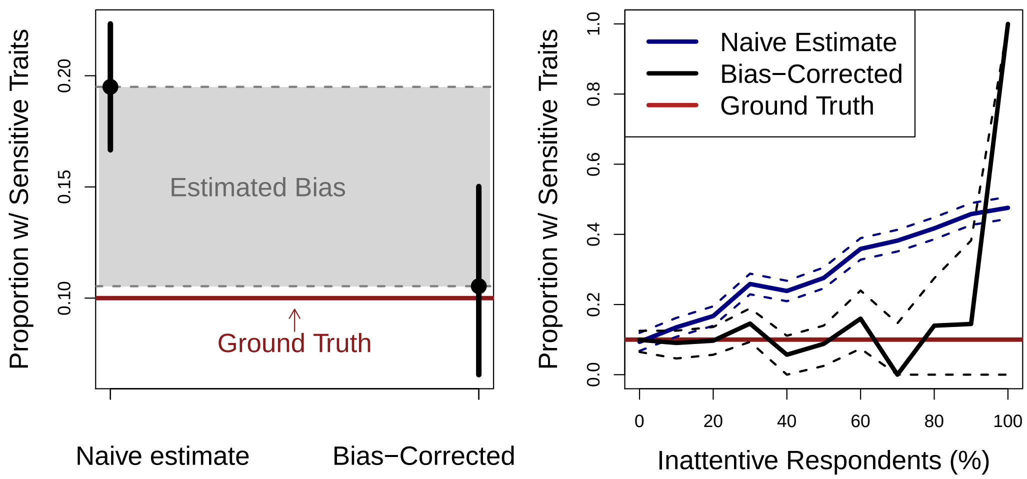

Figure 3: **Illustrations of Bias-Correction.** *Note*: The left panel illustrates that our bias-corrected estimator corrects for the estimated bias. The right panel shows that our bias-corrected estimator is robust to the presence of inattentive respondents, while the naïve estimator is increasingly 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 the proportion of inattentive respondents. The right panel of Figure 4 demonstrates that while the naïve point and interval estimate does increasingly poorly as more inattentive respondents are present in the survey, the bias-corrected estimate is rather robust to inattentive respondents and always captures the ground truth. It also indicates that when over 90% of responses are inattentive, the bias-corrected estimate is no longer very informative as the confidence interval is very wide. However, in such surveys (in which only 10% of respondents are paying attention), *any statistic* would be uninformative and researchers should not use the data without precautions.

To examine the finite sample properties of our bias-corrected estimator, we further replicate the above simulations 8000 times to investigate overall performance of the bias-corrected estimator more systematically. In each simulation, we set different values of parameters and examine the coverage of the true prevalence as well as the bias and the sample root-mean-square error (RMSE). To apply our method to realistic contexts, we choose a set of reasonable values from a parameter space, reflecting the usual situations in which crosswise estimates are applied. Specifically, in each simulation, we draw the true prevalence rate from a continuous uniform distribution (0.1, 0.45), the two auxiliary probabilities from a continuous uniform distribution (0.1, 0.2), and the attentive rate 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.

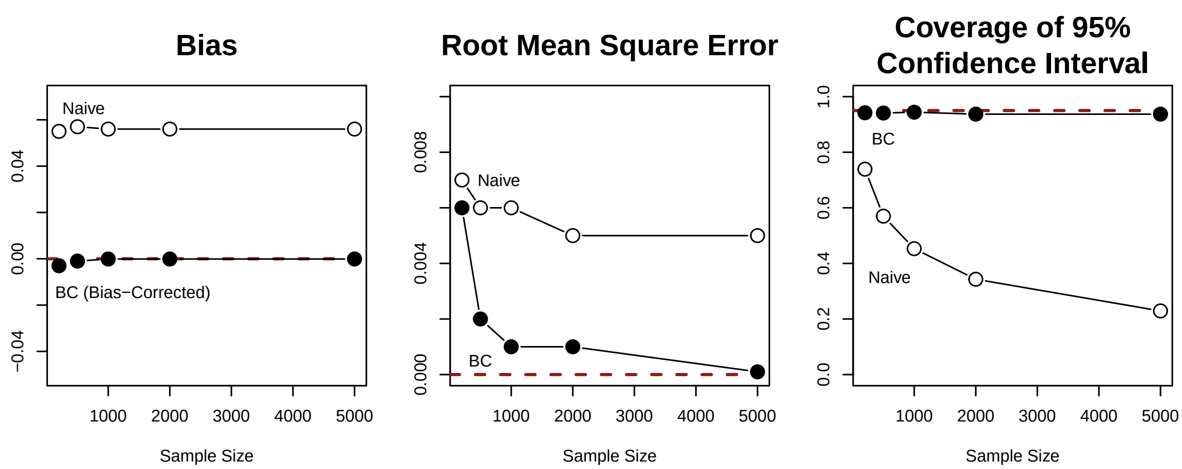

Figure 4: **Simulation Results with Varying Sample Size.** *Note*: This figure displays the bias, and root mean square error, and the coverage of 95% confidence interval of naïve and bias-corrected estimators.

These results demonstrate that the bias-corrected estimator has higher coverage, significantly lower bias, and smaller RMSE than the naïve estimator. The difference between the bias-corrected estimator and naïve estimator is especially remarkable with respect to 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. In Online Appendix B, we provide another set of simulation studies and show that the bias-corrected estimator performs well even when respondents with sensitive traits are more likely to be inattentive than respondents without sensitive attributes.

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 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 the outcome or a predictor.

5.1 Sensitivity Analysis

While our proposed bias correction requires researchers to obtain the estimated proportion of inattentive respondents using the anchor question, it may be the case that researchers do not possess such information (e.g., because the crosswise model was already applied in the past or the anchor question cannot be included 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 enables them to clarify how sensitive their crosswise estimates are to inattentive respondents and what assumptions they must make in order to preserve their original substantive conclusions.

Specifically, we offer a set of sensitivity bounds for original crosswise estimates by applying the bias correction to them under varying levels of attentive respondents. With our sensitivity bounds, researchers can then ask: In order to claim the original magnitude of the estimated prevalence, how many attentive respondents should we have? Or, to what extent can we tolerate the presence of inattentive respondents in order to keep the initial claim?

To illustrate this procedure, we apply the sensitivity analysis to six published studies on 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 responses under Assumption 1. We also plot the point and interval estimate 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 marked by $\times$.

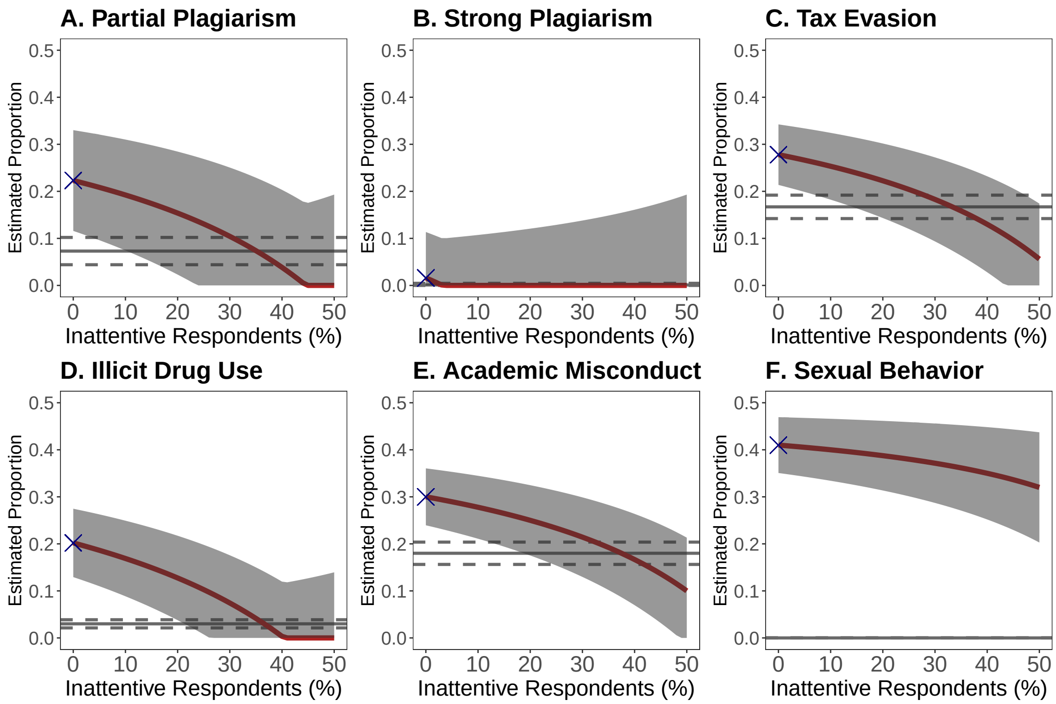

Figure 5: **Sensitivity Analysis of Previous Estimates.** *Note*: This figure shows the results of sensitivity analysis for six crosswise estimates. For each estimate, the bias correction is applied with varying levels of the attentive rate under Assumption 1 ($\kappa = 0.5$). The solid and dashed lines 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, originally presented higher estimates may have been artifacts of inattentive respondents rather than the property of the crosswise model that mitigates social desirability bias. 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 make an assumption that the true proportion of inattentive respondents is less than 0.2. In other words, in order to claim that the crosswise estimator does any better than direct questioning in these studies, researchers must assume that more than 85-90% of the respondents were attentive and followed the instruction. While we cannot verify this with given data, the original estimates must be looked with precautions given the previous findings that 2 to 28% may have been inattentive in surveys in which the crosswise model was employed (Section 2.2).[4]

[4]Our R software `cWise` also allows researchers to set different values of κ other than 0.5 depending on the nature of their already implemented 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 might be violated under their designs.

5.2 Weighting

While the literature of sensitive inquiries usually assumes that survey respondents are obtained by simple random sampling from a finite population of interest, a growing share of surveys are administered with unrepresentative samples such as online opt-in samples (Franco et al., 2017; Mercer, Lau and Kennedy, 2018). Online opt-in samples are known to be often unrepresentative of the entire population of interest, and researchers 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 for sensitive inquiries since sensitive questions are not usually asked to nationally representative samples on large public surveys and researchers often need to conduct their own surveys 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 incorporate sample weights in a straightforward way.

Recall that what we only observe in our framework are λ and λ', 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 $w_i = \frac{1}{\text{Pr}(S_i=1|\mathbf{X})}$, where $S_i = \{0, 1\}$ is a binary variable denoting if unit i is in the sample and $\mathbf{X}$ is a covariate vector. (Here we only consider the base weight for the simplicity, but one can include other weights such as non-response weights).

Let $Y_i \in \{0, 1\}$ be a binary indicator for choosing the crosswise item in the crosswise question and $A_i \in \{0, 1\}$ be a binary indicator for choosing the crosswise item in the anchor question. We then propose to include sample weights 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, Figure 6 compares unweighted and weighted bias-corrected crosswise estimates based on simulated unrepresentative samples (Online Appendix C.1 provides details in our simulation). It demonstrates that while the unweighted estimates overestimate the ground truth, the weighted estimates properly captures the population-level quantity of interest.

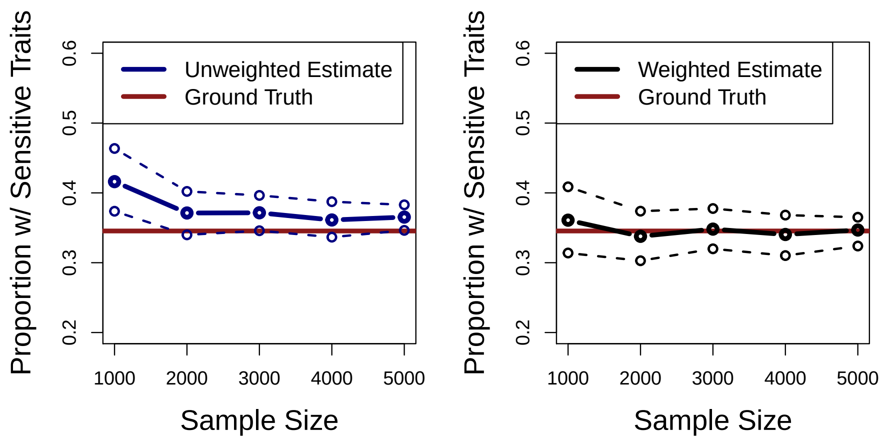

Figure 6: **Weighting in the Crosswise Estimator.** *Note*: The left panel shows bias-corrected crosswise estimates without weighting, whereas the right panel displays bias-corrected estimates with weighting based on simulated unrepresentative samples.

5.3 Crosswise Regressions

In many empirical studies, researchers may wish to go beyond estimating the proportion of individuals with sensitive attributes and to use individual-level information about the possession of sensitive traits as the outcome or a predictor 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 whether engaging in corruption is associated with their 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 still possible to construct a framework for regression analysis which only requires aggregate-level information obtained through the crosswise model.

The application of logistic regression to the latent sensitive variables as the outcome is not new and has been considered in several studies such as Jann, Jerke and Krumpal (2011), Vakilian, Mousavi and Keramat (2014), and Korndörfer, Krumpal and Schmukle (2014), while a generalized linear model framework using the latent sensitive trait as a predictor has been offered by Blair, Imai and Zhou (2015) and Imai, Park and Greene (2015). Meanwhile, our contribution is to propose a framework of multivariate "crosswise" regressions, in which the latent sensitive trait can be used 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 estimates.

5.3.1 Using the Latent Sensitive Trait as an Outcome

We first introduce crosswise regressions in which the latent (unobserved) variable for having a sensitive trait is used as an outcome variable, while applying our bias correction procedure. Let $Z_i \in \{0, 1\}$ be a binary indicator for having a sensitive trait and $T_i \in \{0, 1\}$ be a binary indicator for being attentive for respondent i. Note that both of these quantities are unobserved in reality and thus they are latent variables.

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

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

where $\boldsymbol{\beta}$ is a vector of unknown parameters and $\mathbf{X_i}$ is a vector of characteristics for respondent i. Our goal is to make inferences about the associations between these covariates and the possession of the sensitive trait (and use estimated coefficients in predictions) (e.g., Are young respondents more likely to support a terrorist organization than old respondents?). Thus, $\boldsymbol{\beta}$ are our quantities of interest.

To apply our bias correction inside the above regression, we also introduce the following conditional probability for being attentive:

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

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

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

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

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

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 1-3, we construct the following likelihood function:

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

Building on van den Hout, van der Heijden and Gilchrist (2007), we apply a generalized linear model approach, or logistic regressions more specifically to model the conditional expectations in the crosswise and anchor questions, respectively:

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

Substituting these into Equation (8a) yields the likelihood function that contains every information we need to make estimate our quantities of interest. For estimation, we maximize this likelihood function (after taking its natural log) by an iterative maximization method.[5]

[5] 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.

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 not yet been proposed with respect to 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. (Although other types of outcome variables can be easily incorporated into our framework, we leave to future research the development of such regressions.) We define the regression model (conditional expectation) of interest as

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

where $\mathbf{\Theta}$ is a vector of parameters that associate a series of predictors ($\mathbf{X_i}$, Z_i) and the response variable (V_i). For example, for a normally distributed outcome variable, we can consider $g_{\mathbf{\Theta}}(V_i|\mathbf{X_i}, Z_i) = \mathcal{N}(\alpha + \boldsymbol{\gamma}^\top \mathbf{X_i} + \delta Z_i, \sigma^2)$ with $\mathbf{\Theta} = (\alpha, \boldsymbol{\gamma}, \delta, \sigma^2)$. Similarly, for a binary response variable, we can consider $g_{\mathbf{\Theta}}(V_i|\mathbf{X_i}, Z_i) = \text{Bernoulli}(\phi)$, where $\frac{\phi}{1-\phi} = \alpha + \boldsymbol{\gamma}^\top \mathbf{X_i} + \delta Z_i$ and $\mathbf{\Theta} = (\alpha, \boldsymbol{\gamma}, \delta)$. Our goal is to make inferences about the association between the latent sensitive attribute and the response variable after controlling for other covariates (e.g., Are people supporting a terrorist organization more likely to donate money to a local political institution?). Thus, δ is our primary quantity of interest.

To simplify the derivation, we assume that $V_i \perp\!\!\!\perp Y_i|\mathbf{X_i}$ (the outcome variable and whether the respondents choose the crosswise item are statistically independent conditional upon covariates). Then, using all the available information from data, we construct the following likelihood function:

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

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

Finally, we extend this framework by incorporating the bias correction procedure. With our bias correction, the observed data likelihood function becomes:

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

where each part is $g_{\boldsymbol{\Theta}}(V_i|\mathbf{X_i}, z, t)\Pr(Y_i = 1|Z_i = z, T_i = t)\Pr(Z_i = z|\mathbf{X_i})\Pr(A_i = 1|Z_i = z, T_i = t)\Pr(T_i = 1|\mathbf{X_i})$, 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 1-3 hold and $V_i \perp\!\!\!\perp Y_i|\mathbf{X_i}$, $V_i \perp\!\!\!\perp A_i|\mathbf{X_i}$, 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 (the first equality). 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.[6] 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 the natural log).

To summarize, with our method, researchers can make valid statistical inferences about the population prevalence of sensitive attributes and use such estimate in multivariate analysis for further exploration. We provide both simulation studies and empirical illustration of these extensions in Online Appendices C and D, respectively.

[6]For example, the third component represents $g_{\boldsymbol{\Theta}}(V_i|\mathbf{X_i}, 1, 0)\Pr(Y_i = 1|Z_i = 1, T_i = 0)\Pr(Z_i = 1|\mathbf{X_i})\Pr(A_i = 1|Z_i = 1, T_i = 0)\Pr(T_i = 1|\mathbf{X_i})$. Here, $\Pr(Y_i = 1|Z_i = 1, T_i = 0)$ is the conditional probability that respondents choose the crosswise item when they actually have sensitive trails *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, $\Pr(Z_i = 1|\mathbf{X_i})$ is the conditional probability that respondents have sensitive traits, and we defined this quantity as $\pi_{\boldsymbol{\beta}}(\mathbf{X_i})$. Now, $\Pr(A_i = 1|Z_i = 1, T_i = 0)$ is the conditional probability that respondents choose the crosswise item in the anchor question when they actually have sensitive traits *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, $\Pr(T_i = 1|\mathbf{X_i})$ 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 $g_{\boldsymbol{\Theta}}(V_i|\mathbf{X_i}, 1, 0)\frac{1}{2}\pi_{\boldsymbol{\beta}}(\mathbf{X_i})\frac{1}{2}(1-\gamma_{\boldsymbol{\theta}}(\mathbf{X_i}))$.

Concluding Remarks

We proposed a bias-corrected crosswise estimator for sensitive inquiries. The presence of inattentive responses jeopardizes our statistical inference on sensitive attributes under the conventional crosswise estimator. In particular, the bias caused by inattentive respondents may lead researchers to draw incorrect conclusions that the crosswise model induced more candid answers from survey respondents even when such conclusions are artifact of the bias. We proposed a simple design-based solution to this substantially important problem and demonstrated our strategy both in simulations and an empirical example. Several extensions including sensitivity analysis, multivariate regressions, and weighting method were also provided. We also offered a practical guide on how to design crosswise surveys in order to enable our bias-correction.

References

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.

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.

Fox, James Alan. 2015. *Randomized response and related methods: Surveying sensitive data*. Vol. 58 SAGE Publications.

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.

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, 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.

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"

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.

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

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 few regularity conditions. These conditions state that $\lambda > \kappa$ and $p < \frac{1}{2}$. The regularity conditions hold in usual surveys with the crosswise model. Since 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:

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

A.3 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:

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

Now, by Assumption 3, 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:

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

Combining both results yields the population variance,

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

and its sample analog,

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

No analytical form is available for these functions.

A.4 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.5 Unbiasedness of $\widehat{\lambda}'$

To 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 $k' \sim$ Binom(N, λ') and $\widehat{\lambda}' = k'/N$ suggests $\widehat{\lambda}' \sim$ 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 dashed 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.

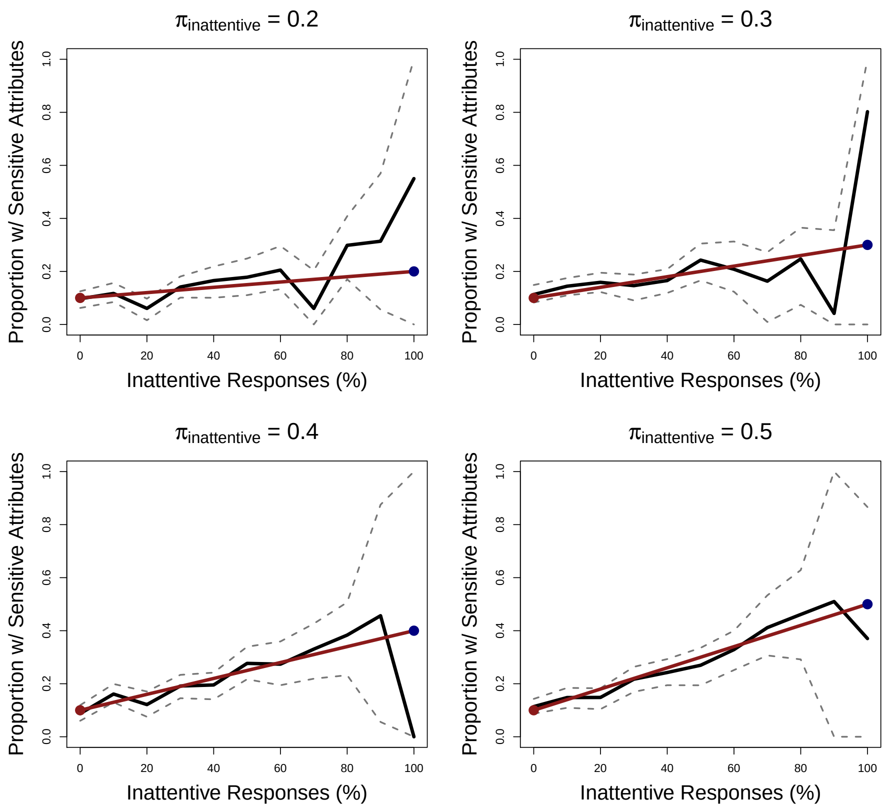

Figure B.1: **Correlation between Attentiveness and Possession of Sensitive Attributes.** *Note*: This graph illustrates the bias-corrected estimates (black lines with gray dashed lines) with the true prevalence rates (red dashed lines) when the true prevalence rates are convex combinations of $\pi_{\text{attentive}}$ 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 discuss the details in 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 $Y_i S_i$:

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

Similarly, we can show that

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

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. For the purpose of choosing 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 the probability of 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 shown in Figure 6.

C.2 Simulations for Crosswise Regressions: Sensitive Trait as the 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 size. The results suggest that the proposed model and estimation strategy are able to recover the true parameters (asymptotically).

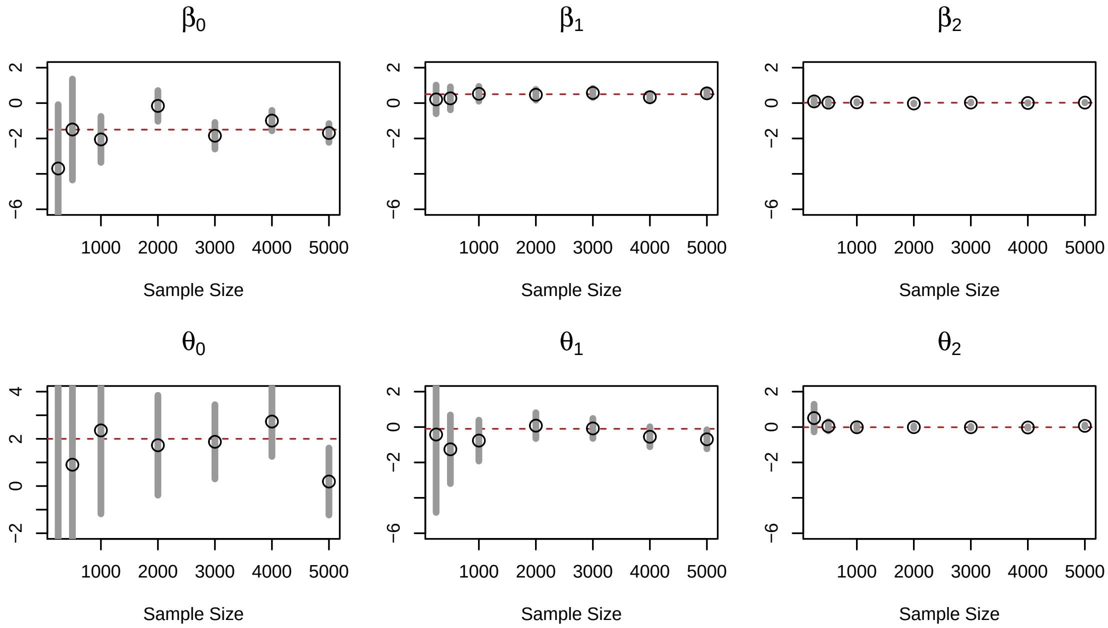

Figure C.1: **Finite Sample Performance of Regression Estimator (Sensitive Trait as a Predictor).** *Note*: Regression estimates of six parameters in simulated data. The dashed lines indicate the true values for the parameters.

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.2 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 association between the latent sensitive trait and the response variable.

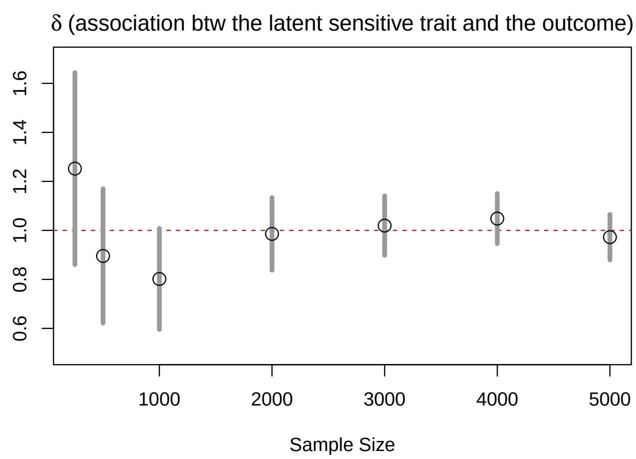

Figure C.2: **Finite Sample Performance of Regression Estimator (Sensitive Trait as a Predictor)**. *Note*: The dashed lines indicate the true values for the parameters.

D Empirical Illustration

In this section, we illustrate the proposed methodology by using a 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 in order to create a normative environment that admitting the behaviors in (1) to (3) becomes a fairy 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.

Next, we employ our proposed regression model framework to examine whether there exists any covariate that predicts sensitive attributes among respondents. For this illustration, we focus one unethical behavior: lying about qualifications on taking surveys. Studying the false qualification is substantively crucial in survey research because when some groups of individuals tend to lie about their qualifications and participate in surveys it may significantly bias substantive conclusions from the survey. For potential predictors, we included variables denoting for age, gender, and the level of general trust. We estimate the logistic-type regression using crosswise responses with the randomization probability of 0.25. Column 1 of Table D.1 report the estimated regression coefficients. We find that none of the included variables have coefficients that are statistically significantly different from zero. The results suggest that false qualifications are not associated with the three variables and might happen randomly.

Moreover, we use the same variable denoting false qualification as a predictor in a regression model.

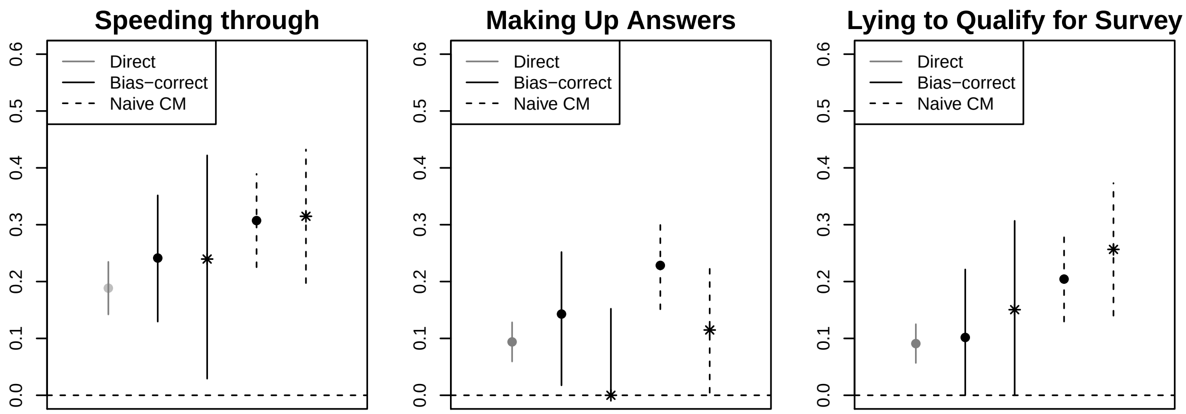

Figure D.1: **Comparison of Prevalence Estimates.** *Note*: This graph visualizes the estimated prevalence of sensitive attributes based on direct inquiry, bias-corrected estimator, and naïve crosswise estimator. For crosswise estimates, dots (second and fourth from the left) are based on low auxiliary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

	Qualification	*Numeracy*
Lie to qualify		5.091**
		(1.131)
Age	-0.031	0.112***
	(0.022)	(0.023)
Female	0.965	-2.765**
	(0.680)	(0.759)
Trust	-0.137	
	(0.178)	
Intercept	0.040	-17.74**
	(1.086)	(0.749)
p	0.25	0.086
p'	0.25	0.086
N	274	196

Table D.1: **Results of Regression Analysis with Bias-Corrected Crosswise Estimates**. *Note*: This table the results of two regression estimates.

Here, we consider subjective numeracy as our dependent variable (mean=-14.98, sd=5.25). Subjective numeracy measures respondents' perceived levels of numeracy or skills to understand numeric information. For predictors, we include variables denoting for age, gender, and false qualification. Note that we do not observe individual level value for the false qualification variable in our crosswise survey data. Nevertheless, as we discussed in Section 5, we can still estimate the coefficient on the latent variable through the joint likelihood function. To estimate the regression, we use crosswise responses with the randomization probability of 0.086. Column 2 of Table D.1 report the results for the regression. The results suggest that individuals who lie to qualify in survey works tend to have higher subjective numeracy. In addition, older respondents and female survey takers seem to have higher subjective numeracy.

E Practical Guide: How to Effectively Design the Crosswise Survey?

In this section, we offer a practical guide for researchers when they apply our proposed methodology. Importantly, the validity of our bias-correction and its extensions hinge upon the three assumptions discussed in Section 3. In the following, we clarify several important points that researchers must consider at the survey design stage in order to satisfy these assumptions.

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

The random pick assumption states that inattentive respondents choose TRUE-TRUE/FALSE-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.,TRUE-TRUE/FALSE-FALSE or otherwise) and they pick one of the two choices randomly. A simple approach is to randomize the ordering of the two choices both in the sensitive question of interest and its anchor question.

E.2 How to Achieve Attention Consistency (Assumption 2)

The constant attentive rate assumption is satisfied when the sensitive question and its anchor question have the same population proportion of attentive *responses*. It must be emphasized that this assumption does not require that the same *respondents* remain inattentive across questions. An important part of this is that researchers must make sure that respondents see both the sensitive and anchor questions in the same way. If respondents, on average, perceive one question to be somehow different from another question, the assumption could be violated. Thus, we recommend that researchers design both the sensitive and anchor questions to look quite similar. Specifically, we suggest that anchor questions be from the same topic and have the same length of wording as sensitive questions. Moreover, randomizing the position of anchor questions in the survey relative to sensitive questions will be helpful to guarantee that there is no carryover effect from one type of question to another.

E.3 How to Make Independent Auxiliary Probabilities (Assumption 3)

The independent randomization assumption claims that auxiliary probabilities used in the sensitive and anchor questions are statistically independent or $p \perp\!\!\!\perp p'$. This assumption will be relatively easily satisfied when researchers carefully choose two auxiliary probabilities (and not sensitive and anchor statements) based on different randomization topics. For example, when the first randomization probability is based on one's mother's birth month and the second probability is based on her father's birth month we are more or less confident that the independent randomization assumption holds (assuming that marriage is not a function of birth months of partners). Importantly, this assumption will be violated when researchers only use a single randomization topic (e.g., mother's birth month) with two different "cut-off points" (e.g., January to March and October to December). This 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. Our recommendation is 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 topics for randomization.[1] This strategy also helps

[1] Using multiple siblings in the birth month-type randomization can be problematic since two siblings' birth months may not necessarily be statistically independent.

researchers by separating the respondents' tasks of coming up with topics and thinking about questions.

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 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* .

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.