Bias-Corrected Crosswise Estimators for Sensitive Inquiries*

Yuki Atsusaka† Ahra Wu‡ Randolph Stevenson§

June 3, 2020

Abstract

The crosswise model is an increasingly popular survey technique to elicit candid answers from respondents on sensitive questions. We demonstrate, however, that conventional crosswise estimators for the true prevalence of sensitive attributes are biased toward 0.5 due to the presence of inattentive responses. In this article, we propose a simple bias-correction to the conventional crosswise estimators. We demonstrate that our bias-corrected estimators are more efficient, easily implemented without measuring individual attentiveness, and applicable to statistical models where crosswise estimates are used either as the outcome or a predictor. We also offer a sensitivity analysis for conventional crosswise estimates and apply it to six existing studies. Our sensitivity analysis suggests that the original findings of these studies – crosswise estimates are higher than estimates from direct questioning – may be largely artifacts of inattentive responses. Finally, we provide a practical guide for designing surveys to enable our proposed bias-correction.

Word Count: 6422

*For helpful comments, we thank Dongzhou Huang, Jeff Lewis, Michelle Torres, and members of the Rice Method Research Group (Gustavo Guajardo, Colin Jones, and Yui Nishimura). The earlier version of this project was presented at the Texas Methodology Conference in April 2019.

†Ph.D. Candidate in the Department of Political Science at Rice University. `atsusaka@rice.edu`.

‡Affiliation. `contact`.

§Professor of Political Science at Rice University. `stevenson@rice.edu`.

1 Introduction

Social scientists often use surveys to probe topics that respondents may hesitate to answer truthfully. For example, few researchers would expect respondents to faithfully report racial animus (Kuklinski, Cobb and Gilens, 1997), discriminatory attitudes and behaviors based on sex (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 druge use (Shamsipour et al., 2014). In each of these topical areas (and many others), we should expect most respondents to, at least, shade their answers in a socially desirable direction.

To combat this tendency, modern survey research has produced a number of survey techniques, collectively known as randomized response techniques (RRT), designed to encourage truthfulness by constructing questions in a way that provides complete privacy protection to respondents (Warner, 1965; Blair, Imai and Zhou, 2015; Fox, 2015). For example, the "crosswise model," which we discuss in detail below, does this by asking the respondent to answer questions based not only on the truth but also on a piece of private information known only to them (Yu, Tian and Tang, 2008). This guarantees respondent-question level anonymity, while allowing researchers to use response information aggregated across respondents to estimate population rates of sensitive attitudes and behaviors. In other words, researchers trade off the ability to know individual level responses to gain aggregate level accuracy in rates of the targeted sensitive attitude or behavior.

The statistical theory behind these estimates (Greenberg et al., 1969; Chaudhuri, 2016), as well as many validation studies (Lensvelt-Mulders et al., 2005; Wolter and Preisendörfer, 2013; Hoffmann et al., 2015; Rosenfeld, Imai and Shapiro, 2016), show that if such techniques are adequately explained and understood by respondents, they can be a useful remedy to social desirability bias. That said, all of the research into the usefulness of these measures has assumed the "if" clause above – i.e., that the techniques are "adequately explained and understood by respondents" (Böckenholt, Barlas and Van Der Heijden, 2009).

In this article, we explore the sensitivity of the crosswise model to this assumption by first pointing out that all of the randomized response survey techniques that have been proposed to combat social desirability bias require respondents to understand fairly complex instructions and to answer questions that do not look at all like the kind of traditional survey questions that respondents may have come to expect. Consequently, we argue that researchers using these techniques should be even more worried about the attentiveness and cooperation of respondents than they would be in a traditional survey situation (and even in traditional surveys, attentiveness and cooperation are perennial problems (Crump, McDonnell and Gureckis, 2013; Meade and Craig, 2012; Maniaci and Rogge, 2014)).

Indeed, below we show that for the crosswise model (which, for reasons explained below, has become the best justified and most used of these techniques), if some portion of survey respondents are confused by the instructions or do not pay careful attention to them, the usual estimator of the true prevalence of the sensitive attitude or behavior will always be biased toward 0.5, with more inattentive respondents adding more bias.[1] Unfortunately, given the low estimated prevalence of many sensitive attitudes and behaviors using traditional survey techniques, this bias is in exactly the direction that would lead researchers to conclude that the alternative crosswise estimator is showing more of the sensitive behavior and so is "working" as expected. Our research, however, suggest that some or all of this difference may simply be due to respondents not attending to the (more complex and confusing) crosswise questions.

Despite this problem, the positive message of our work is that there is a straightforward way to estimate and correct for this bias that requires only a few additional questions added to the survey, some thought about response category randomization, and a simple change to the usual calculation of prevalence estimates post-survey. We show that this bias-corrected estimator is consistent under several plausible assumptions, is easily implemented without individual data on attentiveness, and is applicable to statistical models where crosswise estimates are used either as the outcome or a predictor. Further, even when the additional survey questions necessary for the correction can not be implemented in a survey (e.g., because it was collected in the past), one can use our results to perform a simple sensitivity analysis that maps any assumed percent of inattentive responses to bounds on the amount of the bias.

In the rest of this article, we first review the setup of the crosswise model and derive an expression for bias in the conventional estimator under varying degrees of respondent inattentiveness, as well as plausible assumptions about what inattentive respondents do when forced to provide some answer. Importantly, we also show how simple changes in the design of the crosswise answer categories can make these assumptions

[1] As is usual in the crosswise design, we assume that respondents are forced to make a choice for each question. Allowing respondents to skip questions should not be done in these designs since it introduces a non-random pattern of missing data that will almost certainly bias results in unknown ways. Instead, if one forces all respondents to answer and some do not cooperate by answering randomly, this will cause a biases that we can combat with the techniques described this article. Finally, in a forced choice context, the crosswise design is often preferred by researchers over other randomized response designs exactly because it provides no obvious response strategy for individuals who do not wish to cooperate or who – despite the question format – seek to deceive. Consequently, for this design, it is very reasonable to assume that non-cooperators will either answer randomly or in ways (like always choosing the first response) that can be made random by design (i.e., randomizing the order or response choices).

almost certain to hold. Next, we propose a simple, design-based, correction to the usual crosswise estimator by adding information to the problem in the form of crosswise questions for which prevalence rates are known because they are almost certainly 0 or 1. By choosing such "anchor questions" in the right way, we show that it is possible to estimate and correct for the bias caused by inattentive or confused respondents.

Next, we consider several extensions of our bias-corrected estimator. First, we provide a framework for building regression models where the crosswise estimate is used either as the outcome or a predictor. Second, we use our expression for the bias-corrected estimator to show how researchers who can not make the design corrections we suggest can nevertheless conduct a sensitivity analysis to understand how consequential bias from inattentiveness is for the inferences they want to make. We apply this sensitivity analysis to six existing studies comparing crosswise estimates to estimates based on direct questioning and find that many of the original findings showing that crosswise estimates of prevalence rates are higher than those form direct questioning are plausibly due to the presence of careless and inattentive respondents. Finally, we present a weighting method for bias-corrected crosswise estimates.

In Online Appendix, we illustrate the proposed methodology by applying it to a data set about unethical behaviors of paid survey takers. We also provide a practical guide for researchers about design-stage considerations when applying the crosswise model with our method. We also provide an R package that implements the bias-corrected estimator and its extensions.

2 Crosswise model for Sensitive Inquiries

To understand why bias-correction is critical in the crosswise model, we first describe how the crosswise model work and elaborate the potential problems caused by careless and inattentive responses. Due to their complicated instructions, the crosswise model may be more likely to produce careless and inattentive responses than direct inquiries on sensitive questions and non-sensitive questions. We show that previously suggested solutions will not work unless researchers make unreasonably strong assumptions about careless and inattentive responses.

2.1 Crosswise model

Yu, Tian and Tang (2008) proposed the crosswise model building on a class of randomized response techniques (RRT) in 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
Question: Which of the following most appropriately describes your case?
• Both statements are TRUE, or both statements are FALSE **[the crosswise item]**
• Only one of the statements is TRUE

Here, Statement A is a sensitive statement of interest, and our quantity of interest is the proportion of individuals who agree with (or fit the description in) Statement A (often called as *prevalence rate*). Usually, the prevalence rate of Statement A is less than 0.5 and even close to 0 (e.g., the proportion of workers who have committed serious crime before) and the 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).[2] In contrast, Statement B is a non-sensitive statement whose population prevalence is *ex ante* known to researchers.

Now, after considering *both* statements, respondents are asked to answer whether they are both true or false in Question. This case corresponds to the two open circles (○) in 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. Given the observed *crosswise proportion*, $\mathbb{P}(\text{TRUE-TRUE} \cup \text{FALSE-FALSE})$, researchers can then estimate the quantity of interest via a simple plug-in estimator.

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

Table 1: **Crosswise Model**

To demonstrate how the crosswise model works, let us work through a simple numerical example. Sup-

[2]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).

pose that the true probability for $\mathsf{Statement\ B}$ is (known to be) 0.25 and the observed crosswise proportion is 0.65. As we elaborate below, the crosswise proportion is a combination of the probabilities of $\mathsf{Statement\ A}$ and $\mathsf{Statement\ B}$ being true, respectively. In our example, we can show that $\mathbb{P}$(TRUE-TRUE $\cup$ FALSE-FALSE) = $\mathbb{P}(\mathsf{Statement\ A{=}TRUE}) * 0.25 + \mathbb{P}(\mathsf{Statement\ A{=}FALSE}) * 0.85 = 0.65$. Consequently, if we know the probability for $\mathsf{Statement\ B}$ and the crosswise proportion, we can easily reverse engineer the probability for $\mathsf{Statement\ A}$ as $\mathbb{P}(\mathsf{Statement\ A{=}TRUE}) = \frac{0.65-0.85}{2*0.25-1} = 0.2$.[3] The next section introduces the mechanics more formally.

Among other RRTs, the crosswise model is known to be the most effective in terms of reducing social desirability bias because, under this design, there is no clear choice to make even when respondents still want to "lie" about their answers (Yu, Tian and Tang, 2008). This is true 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 only works when all respondents carefully read, fully understand, and follow thw complex instruction. Indeed, as shown in the next subsection, "random picks" by respondents in the crosswise model not only increase standard errors, but also lead to biased point estimates.

2.2 Inattentive Responses

In surveys, inattentive responses 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 the crosswise model (Schnapp, 2019), estimating 2% and 12% (Höglinger and Diekmann, 2017), 28% (Höglinger and Jann, 2018), and 13% (Enzmann, 2017, cited in Schnapp (2019)) of respondents chose survey answers randomly.

As we demonstrate below, point estimates from the crosswise model are highly sensitive to respondents who answer randomly, and the presence of such respondents generate an estimate of the prevalence rate that is biased toward 0.5. This is highly problematic in the context of sensitive inquiries (where we expect social deniability bias toward 0) because such random answers might make researchers falsely conclude that they successfully induced more candid (and thus true) answers — free from social desirability bias — from

[3] To "re-reverse" engineer this, see that $\mathbb{P}$(TRUE-TRUE $\cup$ FALSE-FALSE) $= 0.2 * 0.25 + (1-0.2) * (1-0.25) = 0.65$.

respondents even when the true prevalence of the sensitive attribute is less pervasive.

To see how problematic this can get, Figure 1 plots the estimated prevalence rate of the sensitive attribute (with its 95% confidence interval) against the percentage of inattentive responses based on hypothetical (and typical) data from the crosswise model. It portrays that there is always a positive bias in the crosswise estimate (i.e., divergence between the true proportion of the sensitive attribute and its estimate) and the bias increases as the percentage of inattentive responses grows. While the original crosswise estimate is still unbiased under the (rare) condition that every respondent follows the instruction (the left arrow), this condition does not meet in almost all survey data. Our contribution here is to present a simple way to quantify the bias caused by inattentive responses (the right arrows) and correct such bias so that the true prevalence rate can be captured by crosswise estimates.

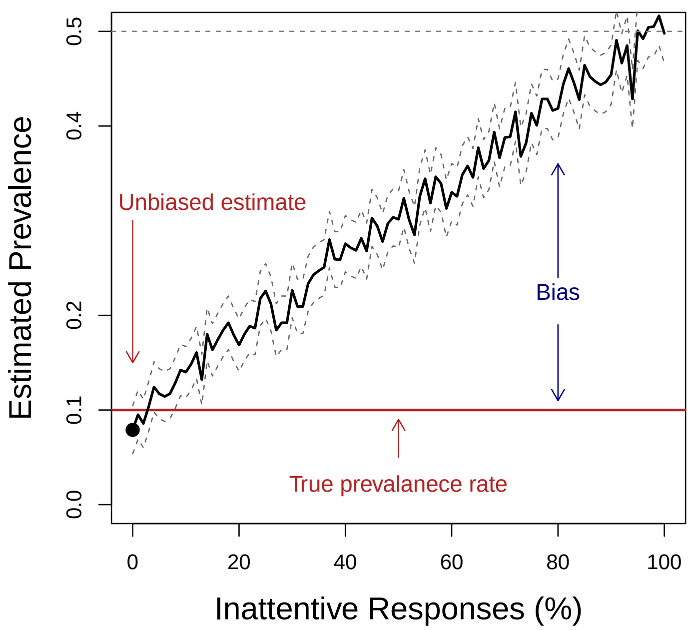

Figure 1: **Consequences of Inattentive Responses on Crosswise Estimates: An Illustration**
Note: This plot illustrates that the bias (i.e., divergence between the true proportion of the sensitive attribute and its estimate) increases as the percentage of inattentive responses grows. Gray dashed lines show confidence intervals. The results are based on hypothetical data where $\pi = 0.1$, $p = 0.2$, and $\widehat{\lambda} = 0.6$.

To date, several solutions to the presence of inattentive respondents have been proposed (even though previous research has not actually shown what the bias actually looks like as in Figure 1 and in the proof as follows). 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 true prevalence rate 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 efficiency (i.e., losing statistical power).

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 when 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 of and inattentive responses by offering a bias-corrected crosswise estimator. Although our bias-corrected estimator is less efficient than an uncorrected or naïve estimator, it yields unbiased estimates of the true prevalence rates with a remarkably weaker set of assumptions than in existing solutions.

3 Bias-Corrected Crosswise Estimators

We now quantify the degree of bias caused by careless and inattentive respondents and then propose a correction for it. The core idea is that the amount of bias can be estimated from the parameters in the crosswise model and the "proportion of inattentive responses." While the latter quantity is unknown in the conventional crosswise model, we show that it is still possible to estimate this proportion by slightly modifying the design. One of our contributions is to point out the exact property of the crosswise model that enables us to obtain such bias-corrected estimates with a set of reasonable assumptions.

3.1 Setting and Notation

For simplicity, suppose that we consider one sensitive question of interest and that N respondents are drawn from a finite population of interest via simple random sampling. Suppose that we apply the crosswise model to estimate the prevalence of the sensitive item and there are no missing data.

Let π be the true prevalence rate of interest (population proportion of individuals who fit the description in Statement A), and let λ denote the crosswise proportion (proportion of respondents who choose TRUE-TRUE or FALSE-FALSE in the crosswise question). Let us then define p as the *randomization probability*, which is the known proportion of individuals who fit the description in Statement B.

Based on Table 1, Yu, Tian and Tang (2008) show that the (true) crosswise proportion is a function of the true prevalence rate and the randomization probability:

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

When inattentive responses are present, however, the crosswise proportion becomes a function of the true prevalence rate, the randomization probability, and the proportion of attentive responses:

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

where γ is the proportion of attentive respondents and κ is the probability that inattentive respondents pick the crosswise item (TRUE-TRUE or FALSE-FALSE).

Note that Equation (1b) is a strict generalization of Equation (1a). Assuming $\gamma = 1$, conventional the crosswise model expresses the true prevalence rate as $\pi = \frac{\lambda+p-1}{2p-1}, p \neq 0.5$. The naïve crosswise estimator then is $\widehat{\pi}_{CM} = \frac{\widehat{\lambda}+p-1}{2p-1}, p \neq 0.5$ and unbiased, where $\widehat{\lambda}$ is the observed crosswise proportion.[4] When $\gamma < 1$, however, $\widehat{\pi}_{CM}$ is no longer an unbiased estimator because it does not take into account the presence of inattentive respondents. Rearranging Equation (1a) and taking expectation, we can quantify and define the bias in this situation (Online Appendix A.1) as:

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

Here, B_{CM} is a bias with respect to the true prevalence that is caused by inattentive responses. Importantly, whenever the true prevalence is less than 0.5 (as we assume), the bias term is always positive (Online Appendix A.2). This means that the conventional crosswise estimator always *overestimates* the true prevalence rate in the presence of inattentive responses. This property of the bias yields a highly problematic

[4]To see this, 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 $\mathbb{E}[\widehat{\lambda}] = \mathbb{E}\big[\frac{k}{N}\big] = \frac{1}{N}\mathbb{E}[k] = \frac{1}{N}N\lambda = \lambda$. Following the parameterization invariance property of MLEs, $\mathbb{E}[\widehat{\pi}_{CM}] = \pi$.

consequence as discussed above.

3.2 Bias-Corrected Estimator

To address the problem, we propose the following bias-corrected crosswise estimator:

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

where we estimate the bias term as:

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

where $\mathbb{E}[\widehat{\lambda}] = \lambda$ and $\mathbb{E}[\widehat{\gamma}] = \gamma$ and thus $\mathbb{E}[\widehat{B}_{CM}] = B_{CM}$.

For our bias-correction to work, we need few assumptions. First, we assume $\kappa = 0.5$, that is, inattentive respondents pick the crosswise item with probability of 0.5. We 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 that this assumption may not hold in most situations because inattentive respondents 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 randomization of the order of the listed items.

When estimating the bias term, another challenge is to obtain an estimated proportion of attentive respondents ($\widehat{\gamma}$). We solve this problem by employing what we call an *anchor question*, where ask about a non-sensitive item whose prevalence is *a priori* known in the crosswise format. For example, for a survey administered to a U.S. population in the U.S., an anchor question may look like:

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

Question: Which of the following most appropriately describes your case?

- Both statements are TRUE, or both statements are FALSE **[the crosswise item]**
- Only one of the statements is TRUE

Here, Statement C is a non-sensitive anchor statement and we know that the true prevalence of individuals who fit its description is (supposed to be) 0 or 1 (in this case 0). Statement D is another non-sensitive (randomization) statement whose population prevalence is known to researchers just like Statement B. Similarly to the target sensitive question, the crosswise proportion in the anchor question is a function of the true prevalence rate for Statement C, the randomization probability for Statement D, and the proportion of attentive responses (here we assume $\kappa = \frac{1}{2}$).

Now, because we know the first two probabilities, we can recover the proportion of attentive responses from the crosswise proportion. Namely, rearrange Equation (1a) yields,

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

where λ' is the crosswise proportion and p' is the randomization probability in the anchor question, respectively. We can estimate this quantity as:

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

where $\widehat{\lambda}'$ the observed crosswise proportion in the anchor question. We can show that $\mathbb{E}[\widehat{\lambda}'] = \lambda'$.[5]

Our strategy then is to plug-in $\widehat{\lambda}$ (obtained from the anchor question) into Equation (2b). For such a plug-in estimator to be valid, we need the following assumption:

Assumption 2 (Constant Attentive Rate). *The population proportion of attentive responses is constant*

[5]To see this, 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 $\Pr(\widehat{\lambda}' = \frac{n}{N}) = \binom{N}{n}(k')^n(1-k')^{N-n}$.

across the crosswise and anchor questions.

This assumption enables us to apply the information about inattentive responses obtained from the anchor question to our estimator for the sensitive question of interest.

In short, our bias-corrected estimator provides an unbiased estimate of the true prevalence rate of sensitive attributes even with the presence of inattentive responses, 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 Randomization). *The randomization probability in the sensitive question is statistically independent from the randomization probability in the anchor question such that the two observed crosswise probabilities are also independent. Formally,* $p \perp\!\!\!\perp p' \Rightarrow \widehat{\lambda} \perp\!\!\!\perp \widehat{\lambda}'$.

With this assumption, we now 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 have more uncertainty than the native crosswise estimator because the former also needs to estimate the attentive rate from data (in addition to the prevalence rate of interest). Since no analytical form 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.

4 Simulation

To illustrate our bias-correction procedure, we perform several simulated studies. The left panel of Figure 2 presents the results of the bias-correction applied to simulated (and typical) survey responses, where $N = 2000$, $\pi = 0.1$, $p = p' = 0.15$, and $\gamma = 0.8$. It shows that the naïve point estimate is far from the true prevalence rate and its 95% confidence interval does not capture the ground truth. In contrast, the bias-corrected point estimate is close to the true prevalence rate and its confidence interval covers the quantity of interest. Note that, as expected, the uncertainty around the bias-corrected estimate is larger than the uncertainty around the naïve estimate.

Using the same parameters values, we also simulate both naïve and bias-corrected estimates of the true prevalence under varying levels of the proportion of inattentive responses $(1 - \gamma)$. The right panel of Figure 2 demonstrates that while the naïve point and interval estimate does increasingly poorly as more inattentive responses are included in the survey, the bias-corrected estimate is rather stable and (always) captures the true prevalence rate (in this simulation). It also indicates that when over 90% of responses are inattentive responses even the bias-corrected estimate is no longer very informative as the confidence interval is wide. In such surveys (in which only 10% of respondents are following the instruction), *any statistic* would be uninformative and researchers should not use the data without precautions.

In Online Appendix B, we report further results of our simulation study and confirm that the bias-corrected estimator performs as expected in more general conditions. Online Appendix D also provides an empirical illustration of our methodology by applying it to data about ethics in survey taking.

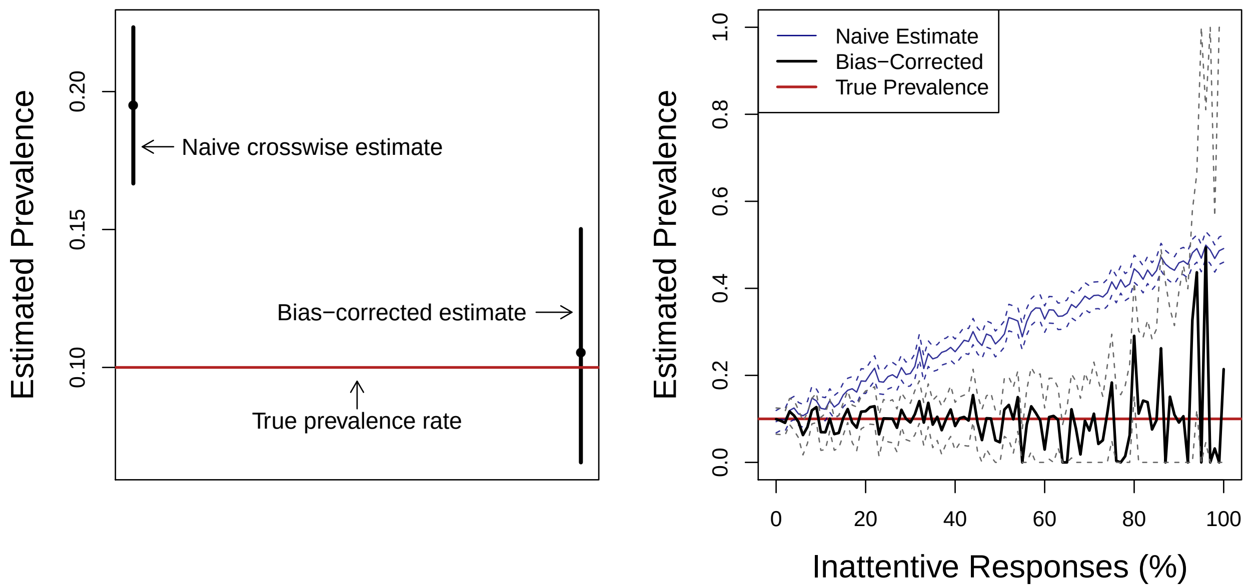

Figure 2: **Illustration of Bias-Correction with Bootstrap Confidence Intervals**
Note: This figure illustrates the bias-corrected estimate along with the native crosswise estimate. The right plot shows that bias-corrected estimates are stable in most range of the inattentiveness, while naïve estimates are not.

5 Extensions

Now, we consider three extensions of the proposed bias-corrected estimator. First, we offer a sensitivity analysis for the crosswise model where the estimated proportion of attentive responses is not available. Second, we consider a framework for multivariate regressions in which crosswise estimates are incorporated either as the outcome or a predictor. Finally, we propose a weighting strategy to estimate the true prevalence among the population of interest using survey answers which are not based on random sampling.

5.1 Sensitivity Analysis

While our proposed bias correction requires researchers to obtain the estimated proportion of inattentive responses ($\widehat{\gamma}$) using the anchor question, it may be the case that researchers do not possess such information (e.g., because the crosswise model was applied in the past or an anchor question cannot be included due to cost). Here, we propose a sensitivity analysis that enables analysts to clarify how sensitive their crosswise estimates are to inattentive responses and what assumptions they must make in order to preserve their original substantive conclusions.

Specifically, we offer a set of sensitivity bounds for crosswise estimates by applying the bias correction to them under varying proportions of attentive responses. With the sensitive bounds, researchers can ask: In order to see the original size of the estimated population prevalence, how many attentive respondents must we have? In other words, 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 proposed sensitivity analysis to six published studies on sensitive behavior, 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 3 visualizes the sensitivity analysis. For each study, we plot the bias-corrected estimates of the true prevalence rate against the possible proportion 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., produce higher estimated prevalence) than direct questioning.

The results suggest that, in many cases, higher prevalence estimates may have been artifacts of the presence of inattentive responses rather than the property of the crosswise model that mitigates social deniability 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 responses is less than 0.2. In other words, in order to claim that crosswise estimator does any better than direct questioning in these studies, the researchers must assume that more than 85-90% of the respondents were attentive and followed the instruction and met Assumption 1. 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).[6]

[6]Our R program for performing this sensitivity analysis 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.

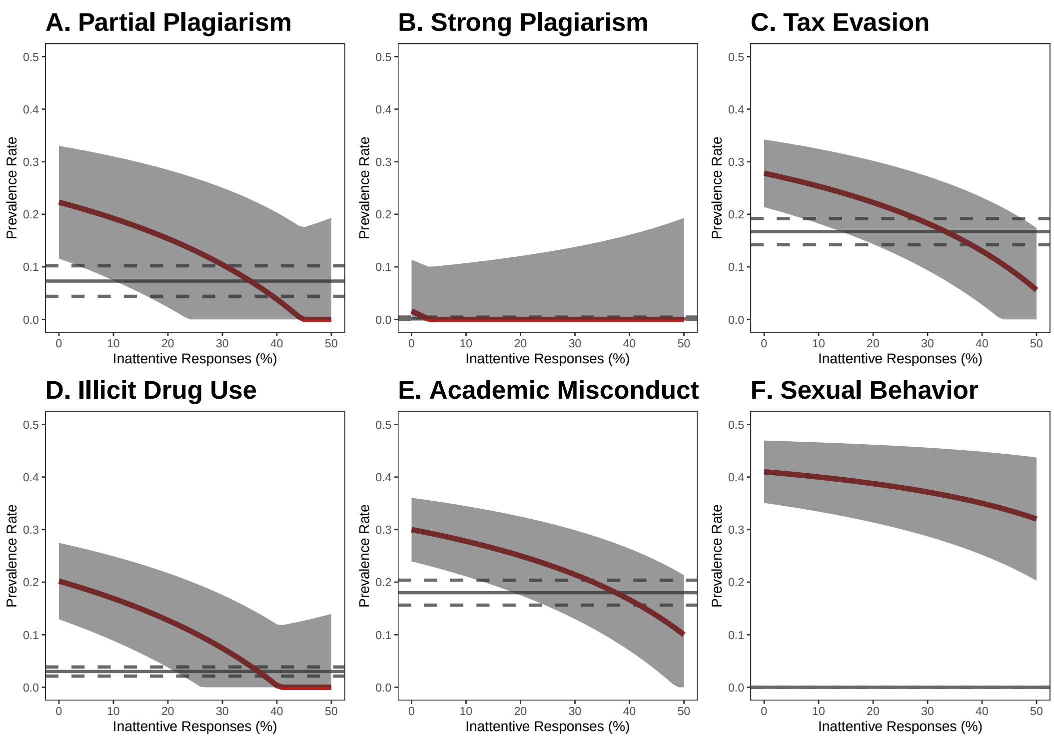

Figure 3: **Sensitivity Analysis of Previous Estimates**
Note: This figure shows the results of sensitivity analysis for six crosswise estimates. For each estimate, the bias correction is applied with varying levels of the attentive rate under Assumption 1 ($\kappa = 0.5$). The confidence intervals are obtained via a normal approximation assuming that the population level attentive rate is known. The solid and dashed lines show point and interval estimates based on direct questioning (except for Study F).

5.2 Multivariate Regressions

In many cases, researchers seek to use the obtained crosswise estimate (prevalence of sensitive attribute) as the outcome or a predictor in multivariate regressions. The application of logistic regression to crosswise outcome variables 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 with a crosswise predictor has been offered by Blair, Imai and Zhou (2015). Here, we generalize the previously proposed framework of regressions used with the crosswise model by incorporating our bias-correction procedure. Online Appendix C.1 details our framework and estimation strategy.

5.3 Weighting

Weighting is a technique to estimate population level quantities based on sample statistics obtained through unrepresentative samples. While the literature on sensitive inquiries usually assumes that samples are generated by simple random sampling, 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 interest.

The benefit of using weighting is even larger for sensitive (like crosswise) 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 randomized response estimates. In Online Appendix C.2, we show that our bias-corrected crosswise estimator can incorporate sample weights in a straightforward way.

Concluding Remarks

We proposed a bias-correction to the conventional 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 responses make researchers draw incorrect conclusions that the usage of the crosswise model induced more candid answers from survey respondents. We proposed a 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.

Future research....

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.

Böckenholt, Ulf, Sema Barlas and Peter GM Van Der Heijden. 2009. "Do randomized-response designs eliminate response biases? An empirical study of non-compliance behavior." *Journal of Applied Econometrics* 24(3):377–392.

Chaudhuri, Arijit. 2016. *Randomized response and indirect questioning techniques in surveys*. Chapman and Hall/CRC.

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.

Greenberg, Bernard G, Abdel-Latif A Abul-Ela, Walt R Simmons and Daniel G Horvitz. 1969. "The unrelated question randomized response model: Theoretical framework." *Journal of the American Statistical Association* 64(326):520–539.

Hoffmann, Adrian, Birk Diedenhofen, Bruno Verschuere and Jochen Musch. 2015. "A strong validation of the crosswise model using experimentally-induced cheating behavior." *Experimental Psychology* .

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.

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.

Lensvelt-Mulders, Gerty JLM, Joop J Hox, Peter GM Van der Heijden and Cora JM Maas. 2005. "Meta-analysis of randomized response research: Thirty-five years of validation." *Sociological Methods & Research* 33(3):319–348.

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.

Wolter, Felix and Peter Preisendörfer. 2013. “Asking sensitive questions: An evaluation of the randomized response technique versus direct questioning using individual validation data.” *Sociological Methods & Research* 42(3):321–353.

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 naïve crosswise estimates 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:

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

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.

B Additional Simulation Studies

We replicate the simulation discussed in Section 4 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-squared 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 randomization 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 Table B.1. 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.

	N=200	*N=500*	*N=1000*	*N=2000*	*N=5000*
Coverage (%)					
naïve estimator	73.9	57.0	45.3	34.3	22.9
Bias-Corrected	94.2	94.1	94.4	93.7	93.7
Bias					
naïve estimator	0.055	0.057	0.056	0.056	0.056
Bias-Corrected	-0.003	-0.001	-0.000	-0.000	-0.000
RMSE					
naïve estimator	0.007	0.006	0.006	0.005	0.005
Bias-Corrected	0.006	0.002	0.001	0.001	0.000

Table B.1: **Performance of naïve and Bias-Corrected Estimators**
Note: This table reports the coverage, bias, and root mean squared error (RMSE) of naïve and bias-corrected estimates. The results are based on 8000 simulated data sets where parameter values are drawn from the following distributions. $\pi \sim \text{unif}(0.1, 0.45), p \sim \text{unif}(0.1, 0.2), p' \sim \text{unif}(0.1, 0.2), \gamma \sim \text{unif}(0.5, 1)$. Each column lists results from a different sample size.

Moreover, 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.[1] 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.

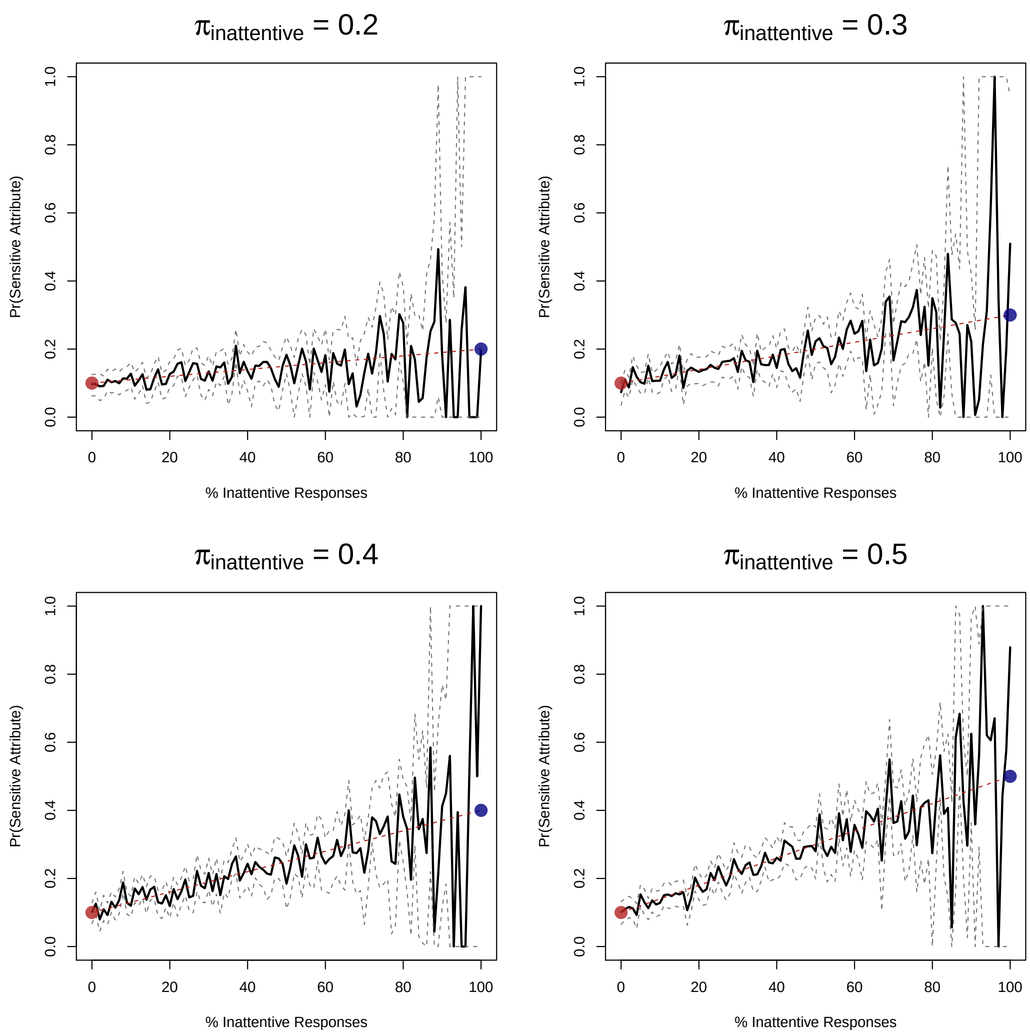

Figure B.1: **Correlation between Attentiveness and Possession of Sensitive Attributes**
Note: This graph illustrates the bias-corrected estimates (black lines with gray dashed lines) with the true prevalence rates (red dashed lines) when the true prevalence rates are convex combinations of $\pi_{\text{attentive}}$ 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 $N = 2000, p = 0.15, p' = 0.15$.

[1] That is, $\pi = \pi_{\text{attentive}} * \gamma + \pi_{\text{inattentive}} * (1 - \gamma)$, where γ is the proportion of attentive responses.

C Details in Extensions

In this section, we discuss the details in the proposed extensions of the bias-corrected estimator.

C.1 Multivariate Regressions with the Crosswise Model

C.1.1 Using Bias-Corrected Estimates as Outcomes

We define the regression model of interest as

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

where Z_i is a binary indicator for having a sensitive attribute of interest, $\boldsymbol{\beta}$ is a vector of unknown parameters, and $\mathbf{X_i}$ is a vector of respondent characteristics. We also introduce the following conditional probability for attentive responses:

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

where $T_i=1$ indicates an attentive respondent and $\boldsymbol{\theta}$ is a vector of unknown parameters that associate a set of background characteristics to the probability of giving careless and inattentive responses.

Let Y_i and A_i be binary indicators that respondent i selects (TRUE-TRUE or FALSE-FALSE) in the crosswise and anchor questions with randomization probabilities p and p', respectively. 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. Under Assumptions 1-3, we can construct the following likelihood function:

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

Our goal here is to maximize this likelihood function to estimate the vector of unknown parameters ($\boldsymbol{\beta}$ and $\boldsymbol{\theta}$).

Building on the existing literature, we consider the generalized linear model (GLM) framework proposed by van den Hout, van der Heijden and Gilchrist (2007) and especially its logistic regression variant. Here we define

$$\pi_{\boldsymbol{\beta}}(\mathbf{X_i}) \equiv \left((2p-1)\pi_{\boldsymbol{\beta}}(\mathbf{X_i})+\left(\frac{1}{2}-p\right)\right)\gamma_{\boldsymbol{\theta}}(\mathbf{X_i})+\frac{1}{2} \quad \text{and} \quad g(\pi_{\boldsymbol{\beta}}(\mathbf{X_i})) = \text{logit}^{-1}(\mu_i) = \boldsymbol{\beta}\mathbf{X_i},$$

where $g(\cdot)$ is a monotonic and differentiable link function. Similarly, we model the response for the anchor question as

$$\gamma_{\boldsymbol{\theta}}(\mathbf{X_i}) \equiv \Big(\frac{1}{2} - p\Big)\gamma_{\boldsymbol{\theta}}(x) + \frac{1}{2} \quad \text{and} \quad g(\gamma_{\boldsymbol{\theta}}(\mathbf{X_i})) = \text{logit}^{-1}(\mu_i') = \boldsymbol{\theta}\mathbf{X_i}$$

Substituting these into Equation (2), the multivariate logistic regression is estimated via the iterative optimization algorithms of choice. Another estimation strategy is based on the expectation-maximization (EM) algorithm in the spirit of Blair, Imai and Zhou (2015).

C.1.2 Using Bias-Corrected Estimates as Predictors

We also consider regression models where bias-corrected estimates are used as predictors. To our knowledge, this type of regression model has not yet been discussed with respect to crosswise estimators. Thus, we begin by describing the model for naïve crosswise estimators.

Define V_i as a continuous or discrete response variable. Researchers may wish to consider the following regression to study how the sensitive attribute is associated with the response variable of interest and other covariates.

$$g_{\boldsymbol{\theta}}(V_i|\mathbf{X_i}, Z_i), \tag{3}$$

where $\boldsymbol{\theta}$ is a vector of associated unknown parameters. For example, for a normally distributed outcome variable, we can consider $g_{\boldsymbol{\theta}}(V_i|\mathbf{X_i}, Z_i) = \mathcal{N}(\alpha + \boldsymbol{\gamma}^\top\mathbf{X_i} + \delta Z_i, \sigma^2)$ and $\boldsymbol{\theta} = (\alpha, \boldsymbol{\gamma}, \delta, \sigma^2)$. 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 $\boldsymbol{\theta} = (\alpha, \boldsymbol{\gamma}, \delta)$.

Using all the available information from data, the observed likelihood function is written as

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\theta}|\{V_i, \mathbf{X_i}, Y_i\}_{i=1}^N, p) &= \prod_{i=1}^N g_{\boldsymbol{\theta}}(V_i, \mathbf{X_i}, Z_i) \\
&= \prod_{i=1}^N \Bigg\{ g_{\boldsymbol{\theta}}(V_i, X_i, 1)\pi_{\boldsymbol{\beta}}(\mathbf{X_i})p^{Y_i}(1-p)^{1-Y_i} \\
&\qquad + g_{\boldsymbol{\theta}}(V_i, X_i, 0)(1 - \pi_{\boldsymbol{\beta}}(\mathbf{X_i}))(1-p)^{Y_i}p^{Y_i} \Bigg\}
\end{aligned} \tag{4}$$

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

Now we employ a similar modeling strategy to use a bias-corrected crosswise estimate as a predictor.

With bias-correction, the observed likelihood function is written as follows:

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\theta} | \{V_i, \mathbf{X_i}, Y_i, A_i\}_{i=1}^N, p, p') &= \prod_{i=1}^N g_{\boldsymbol{\theta}}(V_i, \mathbf{X_i}, Z_i, T_i) \\
&= \prod_{i=1}^N \Bigg\{ g_{\boldsymbol{\theta}}(V_i, \mathbf{X_i}, 1, 1)\pi_{\boldsymbol{\beta}}(\mathbf{X_i}) p^{Y_i}(1-p)^{1-Y_i}(1-p')^{A_i} p'^{1-A_i} \\
&\quad + g_{\boldsymbol{\theta}}(V_i, \mathbf{X_i}, 0, 1)(1-\pi_{\boldsymbol{\beta}}(\mathbf{X_i}))(1-p)^{Y_i} p^{1-Y_i}(1-p')^{A_i} p'^{1-A_i} \\
&\quad + g_{\boldsymbol{\theta}}(V_i, \mathbf{X_i}, 1, 0)\pi_{\boldsymbol{\beta}}(\mathbf{X_i}) \frac{1}{2} \times \frac{1}{2} \\
&\quad + g_{\boldsymbol{\theta}}(V_i, \mathbf{X_i}, 0, 0)(1-\pi_{\boldsymbol{\beta}}(\mathbf{X_i})) \frac{1}{2} \times \frac{1}{2} \Bigg\} \qquad (5)
\end{aligned}$$

Here, each part is $g_{\boldsymbol{\theta}}(V_i, \mathbf{X_i}, z, t)\Pr(Z_i = z | \mathbf{X_i})\Pr(Y_i = 1 | Z_i = z, T_i = t)\Pr(A_i = 1 | T_i = t)$, where $z = \{0, 1\}$ and $t = \{0, 1\}$. Parameter estimates are then obtained via iterative optimization algorithms. Alternatively, the model can be estimated via the EM algorithm as in Blair, Imai and Zhou (2015).

C.2 Weighting in the Crosswise Model

Recall that what we actually observe in the crosswise model are λ and λ', which are observed proportions of respondents choosing (TRUE-TRUE/FALSE-FALSE) in the crosswise and anchor questions. The idea is that we calculate weighed averages of these quantities, instead of simple arithmetic means, where weights reflect sample weights of sensitive and anchor responses. For each respondent, the same weight is used for the two different questions. We can then obtain the proportion of inattentive responses as a deterministic function of λ'.

Formally, we propose to include sample weights in the following way:

$$\widehat{\lambda}_w = \sum_{i=1}^n w_i Y_i \qquad (6)$$

$$\widehat{\gamma}_w = \frac{\sum_{i=1}^n w_i A_i - \frac{1}{2}}{\frac{1}{2} - p'}, \qquad (7)$$

where w_i is a weight for respondent i. The above quantities are then used to calculate our bias-corrected crosswise estimates in Equation (3).

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.

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 randomization 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 randomization probabilities of 0.086 and 0.25, which we call *low* and *moderate* randomization probabilities. Along with the crosswise model, we also performed "direct inquiries" 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 randomization probabilities ($p = 0.086$) and asterisks are based on moderate randomization 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 randomization probabilities, which suggests that researchers will be benefited from using low randomization 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.

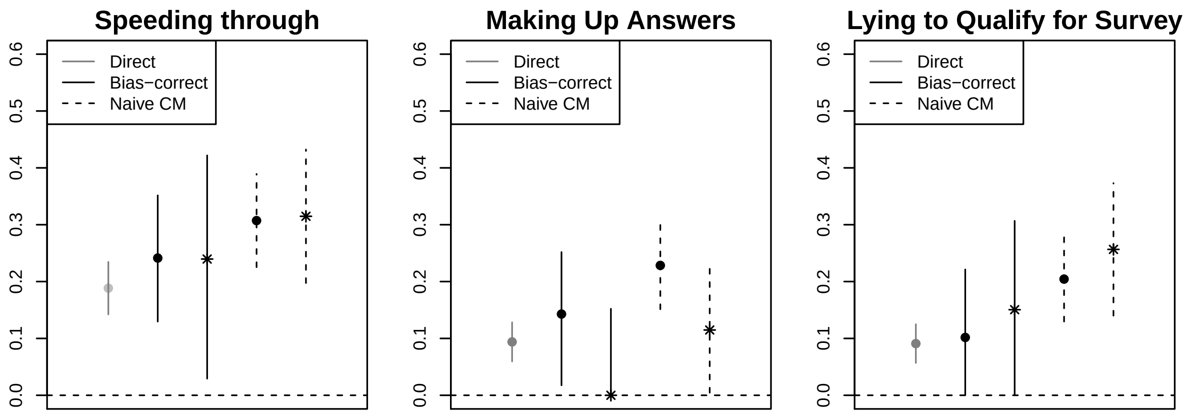

Figure D.1: **Comparison of Prevalence Estimates**
Note: This graph visualizes the estimated prevalence of sensitive attributes based on direct inquiry, bias-corrected estimator, and naïve crosswise estimator. For crosswise estimates, dots (second and fourth from the left) are based on low randomization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

Moreover, we use the same variable denoting false qualification as a predictor in a regression model. 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 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 a Constant Attentive Rate (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 Randomization Probabilities (Assumption 3)

The independent randomization assumption claims that randomization 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 randomization 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.