Bias-Corrected Crosswise Estimators for Sensitive Inquiries*

Yuki Atsusaka† Ahra Wu‡ Randolph Stevenson§

March 23, 2020

Abstract

The crosswise model is an increasingly popular survey technique to elicit candid answers from respondents in sensitive questions. We demonstrate, however, that conventional crosswise estimators for the true prevalence of sensitive attributes are biased toward limit values (0 or 1) due to the presence of inattentive responses, which could make researchers mistakenly believe that they elicited more candid answers when using crosswise models. 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 – crosswise estimates are higher than direct questioning estimates – might be mostly artifacts of inattentive responses. We illustrate the proposed methodology by applying it to an online survey where Qualtrics paid-survey takers are asked about their survey-taking behavior. Finally, we provide a practical guide for designing surveys to enable our proposed bias-correction.

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

†Graduate student, Department of Political Science, Rice University. atsusaka@rice.edu.

‡Affiliation. contact.

§Professor of Political Science, 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, then 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 question that 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 limit values (0 or 1), 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 proposed methodology. 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. Third, we present a weighting method for bias-corrected crosswise estimates. Our proposed methodology including the extensions will be soon available as an R package, `bcw:Bias-Corrected Crosswise Estimators for Sensitive Inquiries` at the Comprehensive R Archive Network.

We further illustrate the proposed methodology by applying it to a data set about unethical behaviors of paid survey takers. Finally, we provide a practical guide for researchers about design-stage considerations when applying crosswise models with our proposed methodology. The final section concludes with future research directions.

2 Crosswise Models for Sensitive Inquiries

To understand why bias-correction is critical in crosswise models, let us first introduce crosswise models as differential privacy survey techniques and elaborate the potential problems caused by careless and inattentive responses. Due to somewhat complicated instructions, crosswise models are 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 Models

Yu, Tian and Tang (2008) proposed crosswise models building on a class of randomized response techniques (RRT) in sensitive inquiries (Warner, 1965; Blair, Imai and Zhou, 2015; Fox, 2015). The basic idea is that by adding tractable random noise to survey questions such that interviewers cannot know individual responses, but know aggregated information researchers can protect respondents' privacy and thus extract their candid answers – free from a social desirability bias – at the expense of losing statistical efficiency (i.e., increasing uncertainty). Crosswise models then take the following form of inquiries:

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
- Only one of the statements is TRUE

Generally, Statement A is a sensitive question of interest, while Statement B is a randomized question. Our quantity of interest is the true prevalence or proportion of individuals with an attribute described in Statement A (often called as *prevalence rate*). For Statement B, we *ex ante* know the true population prevalence of individuals who fit the description. Importantly, the responses that researchers can observe are about the proportion of respondents who choose the first choice (i.e., TRUE-TRUE or FALSE-FALSE); and we call this as a *crosswise proportion* in this article. As such, individual-level true answers are unknown to anyone, including interviewers, and this property of crosswise models is expected to reduce the room for a social desirability bias.

To demonstrate how the crosswise model works, let us work through a simple numerical example. Suppose we know that the true probability for Statement B is 0.25 (we know this from prior knowledge) and the observed crosswise proportion (i.e., TRUE-TRUE/FALSE-FALSE) is 0.65. As the next section elaborates, the crosswise model uses the fact that the crosswise proportion is a "combination" of the probabilities for Statement A and Statement B. In this example, it *must* be true that Pr(TRUE-TRUE $\cup$ FALSE-FALSE) $=$

Pr(Statement A=True) $* 0.25 + (1 -$ Pr(Statement A=True)) $* (1 - 0.25) = 0.65$. Thus, if we know the crosswise proportion and the probability for Statement B, we can easily reverse engineer the probability for Statement A. Here, such probability is calculated as $\frac{0.65+0.25-1}{2*0.25-1} = 0.2$.[2] The next section introduces the mechanics more formally.

Among other RRTs, crosswise models are 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 to directly answer the sensitive question or the randomized question; therefore, respondents are expected to follow the instruction and provide their candid answers.

It is worth noting, however, that this theoretical feature of crosswise models is only attained when all respondents carefully read, fully understand, and follow the instruction. Indeed, as shown in the next subsection, "random picks" by respondents in crosswise models not only increase standard errors, but also lead to biased point estimates.

2.2 Careless and Inattentive Responses

In surveys and survey experiments, careless and 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 found that inattentive responses are also common in crosswise models (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 in without following the crosswise instruction).

If such random answers will just increase the uncertainty of point estimates, the problem is less consequential. However, crosswise estimates are highly sensitive to such random answers such that the more random answers crosswise questions have the closer to 0.5 point estimates become. This is highly problematic in the context of sensitive inquiries because such random answers might make researchers falsely conclude that *they successfully induced more candid (and thus true) answers from respondents about sensitive attributes even when the sensitive attribute is less pervasive (i.e., closer to 0 or 1) than the estimates*.

[2] To "re-reverse" engineer this, we can show that Pr(TRUE-TRUE $\cup$ FALSE-FALSE) $= 0.2*0.25+(1-0.2)*(1-0.25) = 0.65$.

To demonstrate how problematic this can get, Figure 1 plots the estimated prevalence rate of the sensitive attribute against the percentage of inattentive responses based on a hypothetical (and typical) crosswise data. It portrays that there is always a positive bias (i.e., divergence between the true proportion of the sensitive attribute and its estimate) in the crosswise estimate and the bias increases as the percentage of inattentive responses grows.

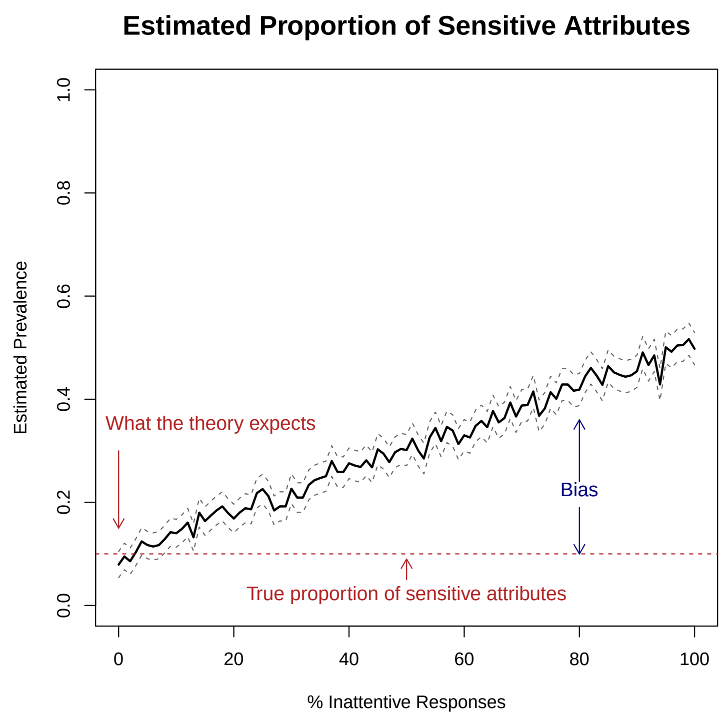

Figure 1: **Consequences of Inattentive Responses on Crosswise Estimates**
Note: This plot illustrates that the bias (i.e., divergence between the true proportion of the sensitive attribute and its estimate) increases as the percentage of inattentive responses grows. Gray dashed lines show confidence intervals. The results are based on a hypothetical data where the true prevalence rate is 0.1, the probability for Statement B is 0.2, and the crosswise proportion is 0.6.

So far, several solutions to the presence of inattentive respondents are proposed (even though previous research has not illustrated what the bias actually looks like as in Figure 1). The first approach is to remove such 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 an biased estimate of the true prevalence rate unless researchers assume that inattentive respondents are a simple random (sub)sample from the original sample. This is an unreasonable strong assumption in most situations and needs to be empirically tested. As many researchers use a crosswise question as the treatment with direct questioning as the control in survey experiments, removing and thus subsetting the data by attentiveness yields post-treatment bias in the average treatment effect (Montgomery, Nyhan and Torres, 2018). Even if the random inattentiveness assumption holds, this approach necessarily affects inferences by losing efficiency.

The second solution is to detect whether respondents answered crosswise questions randomly via direct questioning and then adjust the prevalence estimates (Schnapp, 2019). This approach is valid as long as researchers make assumptions that respondents fully comply with the direct question and their responses are not susceptible to a social desirability bias. However, these assumptions are questionable because, as we expect, inattentive respondents may continue to be inattentive respondents after crosswise questions, and there may exist another social desirability bias to such direct question (i.e., Did you lie about your answer in the last question?)

In the next section, we present an alternative solution to the problem of careless and inattentive responses as a bias-correct crosswise estimator. Although our bias-corrected estimators are less efficient than uncorrected or naïve estimators, it yields unbiased estimates of the true prevalence rates with a remarkably weaker set of assumptions. We also suggest how to minimize the efficiency loss and satisfy the assumption, both of which are achieved in the design stage of crosswise models.

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 former quantities are known and the latter is unknown, it is still possible to "estimate" such proportion (i.e., how many responses are generated not according to the crosswise instruction). One of our contributions is to point out that the exact property of the crosswise model enables us to obtain such estimate when used with few reasonable assumptions.

3.1 Setting and Notation

For simplicity, assume that we have N respondents drawn from the finite population of interest via a simple random sampling and one sensitive question of interest. Also assume that a crosswise model was applied to the sensitive inquiry and there are no missing data in the responses. Let π be the population level proportion of people who have the sensitive attribute described in Statement A (our quantity of interest), and let λ denote the proportion of respondents who choose the first item (i.e., TRUE-TRUE or FALSE-FALSE) in a crosswise question. Define p as the randomization probability, which is the known proportion of individuals who fit the description in Statement B.

Let Y_i be a binary random variable denoting whether a respondent i chooses the first crosswise item (i.e., TRUE-TRUE or FALSE-FALSE) and its realization $y_i \in \{0, 1\}$. Let the number of respondents choosing the first crosswise item be $k = \sum_{i=1}^{N} y_i$, where $k < N$. Then, the likelihood function given any observed k becomes $L(\lambda|N, k) = \binom{N}{k}\lambda^k(1-\lambda)^{N-k}$. Applying the first-order condition yields a maximum likelihood estimate (MLE) of λ, $\widehat{\lambda} = \frac{k}{N}$, where $\mathbb{E}[\widehat{\lambda}] = \lambda$.[3]

Yu, Tian and Tang (2008) show that the proportion of respondents choosing the first crosswise item is expressed as a function of the true prevalence rate and the randomization probability,

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

When inattentive responses are present, however, the same quantity becomes a function of the true prevalence rate, the randomization probability, and the proportion of attentive responses. Thus,

$$\lambda = \Big\{\pi p + (1-\pi)(1-p)\Big\}\gamma + \kappa(1-\gamma), \tag{2}$$

where γ is the proportion of attentive respondents and κ is the probability that inattentive respondents randomly pick the first item.

Assuming $\gamma = 1$, conventional crosswise models express the population level prevalence of sensitive items as $\pi = \frac{\lambda+p-1}{2p-1}, p \neq 0.5$. According to the parameterization invariance property of the MLE, the naïve crosswise estimator becomes $\widehat{\pi}_{CM} = \frac{\widehat{\lambda}+p-1}{2p-1}, p \neq 0.5$. Again, assuming $\gamma = 1$, Yu, Tian and Tang (2008) show that $\widehat{\pi}_{CM}$ is an unbiased estimator of π_{CM}.

However, whenever $\gamma < 1$, $\widehat{\pi}_{CM}$ will not be unbiased because it does not take into account the presence of inattentive respondents. Rearranging Equation (3) and taking expectation, we can show (details in

[3]The unbiasedness follows from the fact that $\mathbb{E}[\widehat{\lambda}] = \mathbb{E}\big[\frac{k}{N}\big] = \frac{1}{N}\mathbb{E}[k] = \frac{1}{N}N\lambda = \lambda$.

Appendix A):

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

Here, we define B_{CM} as a bias with respect to the true prevalence that is caused by inattentive responses. Importantly, whenever the true prevalence is less than 0.5, the bias term becomes positive, which suggests that conventional crosswise estimators overestimate the true prevalence rate in the presence of inattentive responses. This property of the bias yields a highly problematic consequence as discussed above.[4]

3.2 Proposed Bias Correction

We propose the following bias-correction to conventional crosswise estimators.[5]

$$\widehat{\pi}_{BC} = \max(0, \min(1, \widehat{\pi}_{CM} - \widehat{B}_{CM})) \tag{3}$$

with the bias estimate,

$$\widehat{B}_{CM} = \frac{\widehat{\gamma} - 1}{2\widehat{\gamma}}\left(\frac{\widehat{\lambda} - \kappa}{p - \frac{1}{2}}\right), \tag{4}$$

where $\mathbb{E}[\widehat{\gamma}] = \gamma$ and thus $\mathbb{E}[\widehat{B}_{CM}] = B_{CM}$. In practice, however, we rely on the following bias estimator,

$$\widehat{B}_{CM} = \frac{\widehat{\gamma} - 1}{2\widehat{\gamma}}\left(\frac{\widehat{\lambda} - \frac{1}{2}}{p - \frac{1}{2}}\right) \tag{5}$$

Notice that here we assume $\kappa = 0.5$, in other words, inattentive respondents randomly pick the first crosswise item by the probability of 0.5. Let us state this assumption as follows:

Assumption 1 (Random Pick). *Inattentive respondents randomly choose the first crosswise item (i.e., TRUE-TRUE or FALSE-FALSE) by the probability of 0.5.*

[4] When all respondents are attentive, the bias term becomes 0 ($B_{CM} \to 0$) as $\gamma \to 1$. In contrast, when all respondents are inattentive, (1) the bias term becomes ∞ ($B_{CM} \to \infty$) as $\gamma \to 0$ when $\lambda > 0.5$ and $p < 0.5$ (which is often the case) or (2) the bias term becomes $-\infty$ when $\lambda < 0.5$ and $p < 0.5$ or $\lambda > 0.5$ and $p > 0.5$.

[5] We apply the logical constrain to the bias-corrected estimate.

The survey literature easily tells that this assumption does not hold in most situations because most careless respondents tend to choose the first item than the second or lower items (Krosnick, 1991; Galesic et al., 2008). Nevertheless, it is still possible to *design* a survey so that careless respondents can pick the first crosswise item, which indicates TRUE-TRUE/FALSE-FALSE and is not necessarily the "top" item any more, via randomization of choice orders. It should be emphasized that this assumption is much easier to satisfy via survey design compared to the assumptions in other adjustment strategies discussed in Section 2.

When estimating the bias term, the last challenge is to obtain an estimated proportion of attentive respondents ($\widehat{\gamma}$). We solve this problem by employing what we call an "anchor" sensitive question with known prevalence under the following assumption.

Assumption 2 (Constant Attentive Rate). *The population proportion of attentive "responses" is constant across the crosswise sensitive question of interest and the anchor-sensitive question(s).*

It should be emphasized that Assumption 2 holds whenever we have the same number of attentive responses and not attentive respondents across the two questions.

Given that, when the expected prevalence of sensitive item is less than 0.5, we propose including one or more anchor crosswise questions in the survey with different randomization probabilities and with known prevalence of 0. For this purpose, we can ask about a phenomena that we are sure nobody experiences. For example, we can consider the following anchor item:

Instruction: Please read the two statements below

Statement C: I was born in the U.S. but do not have a permanent residency

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
- Only one of the statements is TRUE

Here, we are assuming that the population level prevalence of people who fit the description in Statement C is zero (or very near zero). If Assumption 1 holds, we can rearrange Equation (3) to yield,

$$\gamma = \frac{\lambda' - \frac{1}{2}}{\frac{1}{2} - p'}, \tag{6}$$

Via the MLE's parameterization invariance property, we can estimate this quantity by,

$$\widehat{\gamma} = \frac{\widehat{\lambda}' - \frac{1}{2}}{\frac{1}{2} - p'}, \tag{7}$$

where $\widehat{\lambda}'$ is an observed proportion of respondents choosing the first crosswise item with the quasi-sensitive item, and p' is the population prevalence of people who fit the description of Statement D. 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 (TRUE-TRUE or FALSE-FALSE) in the quasi-sensitive question.[6]

To simply the variance derivation, let us make one additional assumption.

Assumption 3 (Independent Randomization). *The randomization probability of the sensitive question of interest is statistically independent from the randomization probability of the quasi-sensitive question such that the two observed crosswise probabilities are also independent. Formally,* $p \perp\!\!\!\perp p' \Rightarrow \widehat{\lambda} \perp\!\!\!\perp \widehat{\lambda}'$.

The advantage of Assumption 3 is that it can simplify the derivation of the variance of the crosswise estimate, whereas it can be easily satisfied by researchers at the design stage of survey. In Section 7, we discuss how to design surveys in order to satisfy the three assumptions.

Finally, Equation (5) implies that the absolute size of the bias is a function of the randomization probability is the sensitive question (p). Specifically, the bias becomes infinity as p approaches 0.5.

3.3 Large-Sample Inference

We now derive the variance of the bias-corrected crosswise estimators and its sample estimator. To begin with, Yu, Tian and Tang (2008, 257) show that the population variance of a regular crosswise estimator and

[6]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 $\Pr(\widehat{\lambda}' = \frac{n}{N}) = \binom{N}{n}(k')^n(1-k')^{N-n}$.

its sample analog are as follows:[7]

$$\mathbb{V}(\widehat{\pi}_{CM}) = \mathbb{V}\left[\frac{\widehat{\lambda}}{2p-1}\right] = \frac{\lambda(1-\lambda)}{n(2p-1)^2} = \frac{\pi(1-\pi)}{n} + \frac{p(1-p)}{n(2p-1)^2} \tag{8}$$

$$\widehat{\mathbb{V}(\widehat{\pi}_{CM})} = \widehat{\mathbb{V}}\left[\frac{\widehat{\lambda}}{2p-1}\right] = \frac{\widehat{\lambda}(1-\widehat{\lambda})}{n(2p-1)^2} = \frac{\widehat{\pi}_{CM}(1-\widehat{\pi}_{CM})}{n-1} + \frac{p(1-p)}{(n-1)(2p-1)^2} \tag{9}$$

Based on the similar derivation, we obtain the population variance and its sample analog for the bias-corrected crosswise estimator (details in Appendix B):

$$\mathbb{V}(\widehat{\pi}_{BC}) = \mathbb{V}\left[\frac{\widehat{\lambda}}{\widehat{\lambda}' - \frac{1}{2}}\right] \tag{10}$$

$$\widehat{\mathbb{V}(\widehat{\pi}_{BC})} = \widehat{\mathbb{V}}\left[\frac{\widehat{\lambda}}{\widehat{\lambda}' - \frac{1}{2}}\right] \tag{11}$$

Now, let us step back and compare Equation 10 to Equation 8 (the second term). An important observation is that while Equation 8 only has one random variable inside the bracket ($\widehat{\lambda}$), Equation 10 has two random variables ($\widehat{\lambda}$ and $\widehat{\lambda}'$). Thus, Equation 10 is a variance of the ratio of the two random variables and inevitably has a larger value than Equation 8. Substantively, this means that the bias-corrected estimator inevitably has a larger 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 (10) and (11), in practice, we employ bootstrapping to construct confidence intervals and perform hypothesis testing (Efron, 1982; Efron and Tibshirani, 1994).

4 Simulation Studies

To illustrate our bias-correction procedure, we present several simulated results in this subsection. Under several regularity conditions, the bias-corrected estimator accounts for a potential bias caused by the presence of inattentive responses. Figure 2 (left plot) presents the results of the bias-correction to simulated survey responses, where the sample size is 2000, the true prevalence (π) is 0.1, the main randomization probability (p) is 0.15, the secondary randomization probability (p') is 0.15, and the population attentiveness (γ) is set to 0.8. In this simulation, the estimated bias is 0.060, which pulls the naïve estimate toward

[7] To supplement another step in Equation (8), between the first two quantities is $\mathbb{V}\left(\frac{p+\widehat{\lambda}-1}{2p-1}\right)$. Since p and -1 are constants, they can be excluded from the variance computation without changing anything.

one. In addition to the bias, the 95% confidence intervals of the naïve estimate do not capture the population prevalence rate. On the contrary, the bias-corrected estimator produces an unbiased point estimate, although its confidence intervals are wider than the naïve estimate's intervals due to additional variance introduced in the bias-correction procedure.

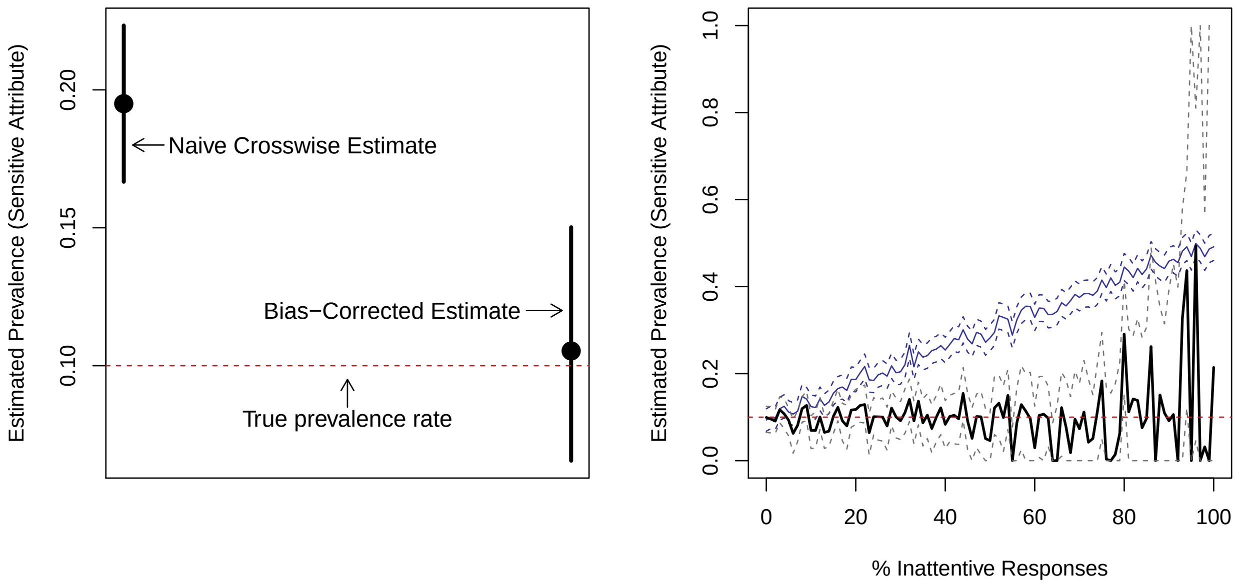

Figure 2: **Illustration of Bias-Correction with Bootstrap Confidence Intervals**
Note: This figure illustrates the bias-corrected estimate along with the native crosswise estimate. To simulate hypothetical responses, we used $N = 2000, \pi = 0.1, p = 0.15, p' = 0.15$ and $\gamma = 0.8$ (γ is only fixed in the left plot). The confidence intervals for the bias-corrected estimate is obtained via bootstrapping. This demonstrates that the naïve crosswise estimate is biased toward 0.5 and the proposed bias-correction accounts for such bias to correctly capture the true prevalence rate. The right plot shows that bias-corrected estimates (black line) are stable in most range of the inattentiveness, while naïve estimates (blue line) are not.

The right plot in Figure 2 displays the stability of bias-corrected estimates. Remarkably, the bias-corrected estimator (with its uncertainty estimates) covers the true prevalence rate (0.1) even when data contains a large number of inattentive responses. To highlight the contribution of the bias-correction, we also plot the behavior of naïve estimator which appeared in Figure 1 in the same graph. It is clear that our strategy properly solves the problem caused by inattentive responses unless 90% and more responses in data are inattentive. Under such pathological situation, even though the uncertainty estimates over the true prevalence rate as expected they are no longer very informative. However, in such surveys (where only 10%

of respondents are paying attention), *any statistic* is uninformative and researchers must not use the data without precautions.

Next, we replicate the above simulation 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 a realistic context, we choose a set of values from a reasonable parameter space. Specifically, in each simulation, we draw the true prevalence rate from a continuous uniform distribution (0.1, 0.45), the main and secondary 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 size of 200, 500, 1000, 2000, and 5000.

We report our results in Table 1. It demonstrates that the bias-corrected estimator has a 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 by 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 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.

Finally, we perform another simulation studies 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 (due to some psychological effects) 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 3 resolve such concern. In each panel, we plot the bias-corrected point and interval estimates over the true prevalence rate against varying levels of inattentiveness in a 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.[8] Nevertheless, Figure 3 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.

5 Several Extensions

In this section, we consider three extensions of the proposed bias-corrected estimators. First, we consider multivariate regression models in which crosswise estimates are incorporated either as the outcome or the predictor variable of interest. Second, we offer a sensitivity analysis for research where the estimated proportion of inattentive responses is not available. Third, 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 Multivariate Regression Model

In multivariate regression analysis, researchers seek to find which background characteristics of respondents are associated with sensitive attributes under investigation and generate predicted probabilities possessing sensitive attributes among unsampled data points. The application of logistic regression to crosswise response variables has been considered in Jann, Jerke and Krumpal (2011), Vakilian, Mousavi and Keramat (2014), and Korndörfer, Krumpal and Schmukle (2014). Our contribution is to extend previously proposed regression approaches to the bias-corrected estimate and provide an appropriate estimation strategy.

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

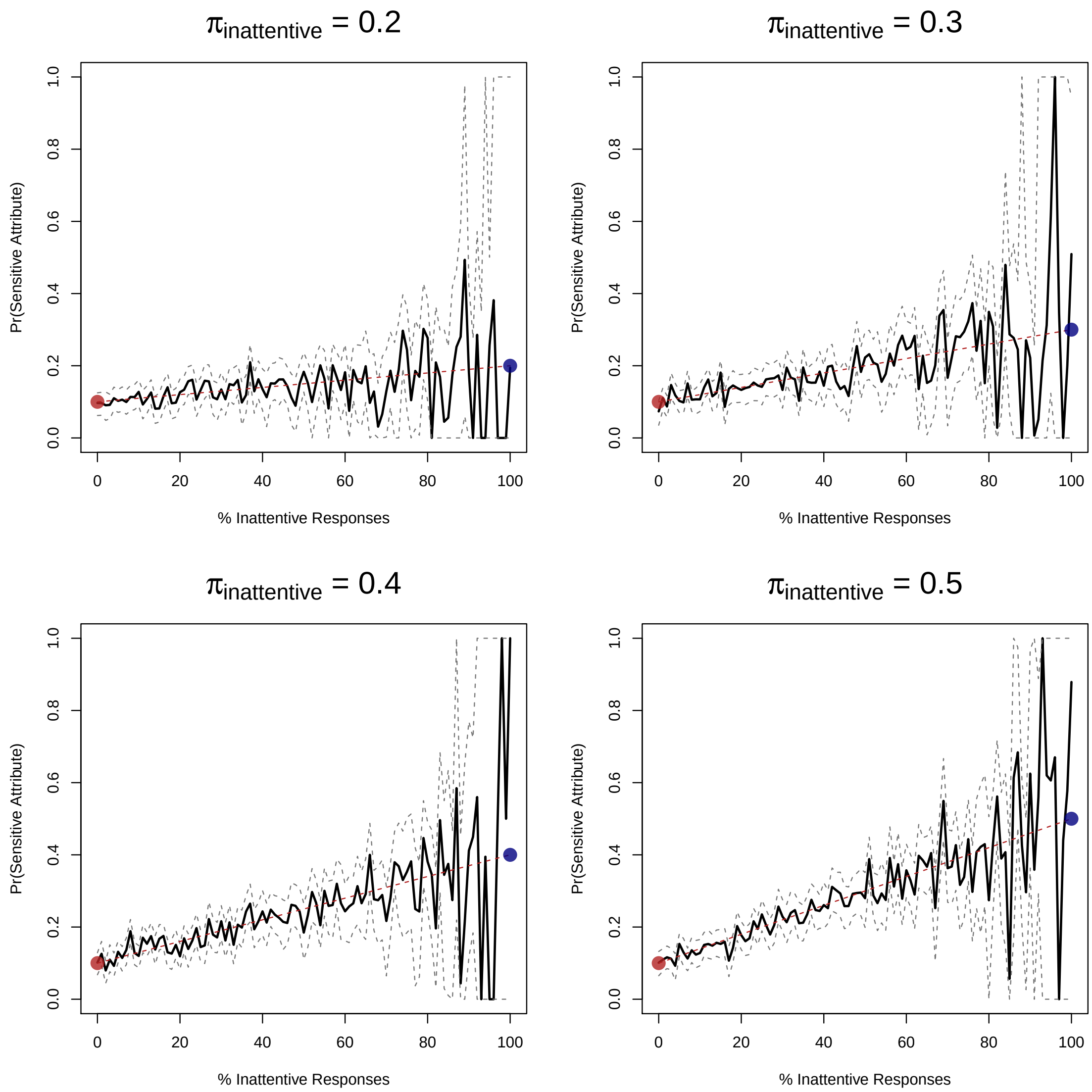

Figure 3: **Correlation between Attentiveness and Possession of Sensitive Attributes**
Note: This graph illustrates the bias-corrected estimates (black lines with gray dashed lines) with the true prevalence rates (red dashed lines) when the true prevalence rates are convex combinations of $\pi_{\text{attentive}}$ 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 $\pi_{\text{attentive}}$.

5.1.1 Using Bias-Corrected Estimates as Outcomes

We define the regression model of interest as

$$\mathbb{E}[Z_i|\mathbf{X_i} = \mathbf{x}] = \Pr(Z_i = 1|\mathbf{X_i} = \mathbf{x}) = \pi_{\boldsymbol{\beta}}(\mathbf{x}), \tag{12}$$

where Z_i is an 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 response:

$$\Pr(T_i = 1|\mathbf{X_i} = \mathbf{x}) = \gamma_{\boldsymbol{\theta}}(\mathbf{x}), \tag{13}$$

where T_i is an indicator for attentive respondent and $\boldsymbol{\theta}$ is a vector of unknown parameters that associate the same background characteristics to the probability of giving careless and inattentive responses. 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:

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\theta}|\{\mathbf{X_i}, Y_i, A_i\}_{i=1}^N, p, p') &= \prod_{i=1}^N \Big\{\lambda_{\boldsymbol{\beta}}(\mathbf{X_i})\Big\}^{Y_i}\Big\{1-\lambda_{\boldsymbol{\beta}}(\mathbf{X_i})\Big\}^{1-Y_i}\Big\{\lambda'_{\boldsymbol{\theta}}(\mathbf{X_i})\Big\}^{A_i}\Big\{1-\lambda'_{\boldsymbol{\theta}}(\mathbf{X_i})\Big\}^{1-A_i} \\
&= \prod_{i=1}^N \left\{\left((2p-1)\pi_{\boldsymbol{\beta}}(\mathbf{X_i}) + \left(\frac{1}{2}-p\right)\right)\gamma_{\boldsymbol{\theta}}(\mathbf{X_i}) + \frac{1}{2}\right\}^{Y_i} \\
&\quad \times \left\{1 - \left[\left((2p-1)\pi_{\boldsymbol{\beta}}(\mathbf{X_i}) + \left(\frac{1}{2}-p\right)\right)\gamma_{\boldsymbol{\theta}}(\mathbf{X_i}) + \frac{1}{2}\right]\right\}^{1-Y_i} \\
&\quad \times \left\{\left(\frac{1}{2}-p'\right)\gamma_{\boldsymbol{\theta}}(\mathbf{X_i}) + \frac{1}{2}\right\}^{A_i}\left\{1 - \left[\left(\frac{1}{2}-p'\right)\gamma_{\boldsymbol{\theta}}(\mathbf{X_i}) + \frac{1}{2}\right]\right\}^{1-A_i}
\end{aligned} \tag{14}$$

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

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

Substituting these into Equation (17), 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).

5.1.2 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.

$$g_{\boldsymbol{\theta}}(V_i|\mathbf{X_i}, Z_i), \tag{15}$$

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

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\theta}|\{V_i, \mathbf{X_i}, Y_i\}_{i=1}^N, p) &= \prod_{i=1}^N g_{\boldsymbol{\theta}}(V_i, \mathbf{X_i}, Z_i) \\
&= \prod_{i=1}^N \Bigg\{ g_{\boldsymbol{\theta}}(V_i, X_i, 1)\pi_{\boldsymbol{\beta}}(\mathbf{X_i})p^{Y_i}(1-p)^{1-Y_i} \\
&\qquad + g_{\boldsymbol{\theta}}(V_i, X_i, 0)(1-\pi_{\boldsymbol{\beta}}(\mathbf{X_i}))(1-p)^{Y_i}p^{Y_i} \Bigg\}
\end{aligned} \tag{16}$$

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:

$$
\begin{aligned}
\mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\theta} | \{V_i, \mathbf{X_i}, Y_i, A_i\}_{i=1}^N, p, p') &= \prod_{i=1}^N g_{\boldsymbol{\theta}}(V_i, \mathbf{X_i}, Z_i, T_i) \\
&= \prod_{i=1}^N \Bigg\{ g_{\boldsymbol{\theta}}(V_i, \mathbf{X_i}, 1, 1)\pi_{\boldsymbol{\beta}}(\mathbf{X_i}) p^{Y_i}(1-p)^{1-Y_i}(1-p')^{A_i} p'^{1-A_i} \\
&\quad + g_{\boldsymbol{\theta}}(V_i, \mathbf{X_i}, 0, 1)(1-\pi_{\boldsymbol{\beta}}(\mathbf{X_i}))(1-p)^{Y_i} p^{1-Y_i}(1-p')^{A_i} p'^{1-A_i} \\
&\quad + g_{\boldsymbol{\theta}}(V_i, \mathbf{X_i}, 1, 0)\pi_{\boldsymbol{\beta}}(\mathbf{X_i}) \frac{1}{2} \times \frac{1}{2} \\
&\quad + g_{\boldsymbol{\theta}}(V_i, \mathbf{X_i}, 0, 0)(1-\pi_{\boldsymbol{\beta}}(\mathbf{X_i})) \frac{1}{2} \times \frac{1}{2} \Bigg\}
\end{aligned} \tag{17}
$$

Here, each part is $g_{\boldsymbol{\theta}}(V_i, \mathbf{X_i}, z, t)\text{Pr}(Z_i = z|\mathbf{X_i})\text{Pr}(Y_i = 1|Z_i = z, T_i = t)\text{Pr}(A_i = 1|T_i = t)$, 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).

5.2 Sensitivity Analysis

Although our proposed bias correction requires researchers to obtain the estimated proportion of inattentive responses ($\widehat{\gamma}$) and satisfy Assumption 1, it may be the case that only one or none of the two conditions is met. Here, we propose a sensitivity analysis that, even in such pathological conditions, enables analysts to clarify how sensitive their crosswise estimates are and what assumption they must make in order to preserve their original substantive conclusions. Specifically, we offer a set of sensitivity bounds for a crosswise estimate of the population prevalence of sensitive attributes by applying the bias correction under varying levels of the attentive rate (γ) and the "first-choice pick" bias (κ).

With the sensitive bounds, thus, 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 and first-choice pickers in order to keep the initial claim?

To illustrate the sensitivity analysis, we apply our method 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), sexual behavior (Vakilian, Mousavi and Keramat, 2014), and academic misconduct (Höglinger, Jann and Diekmann, 2016). Figure 4 visualizes the sensitivity analysis. For each study, we plot the estimated prevalence rate against the potential attentive rate (which was not observed during the study) under Assumption 1. We also compare the corrected estimates with point prevalence estimate based on direct questions (if available) since many studies attempt to show that crosswise estimators *perform better* than direct question in terms of drawing honest responses about sensitive topics.

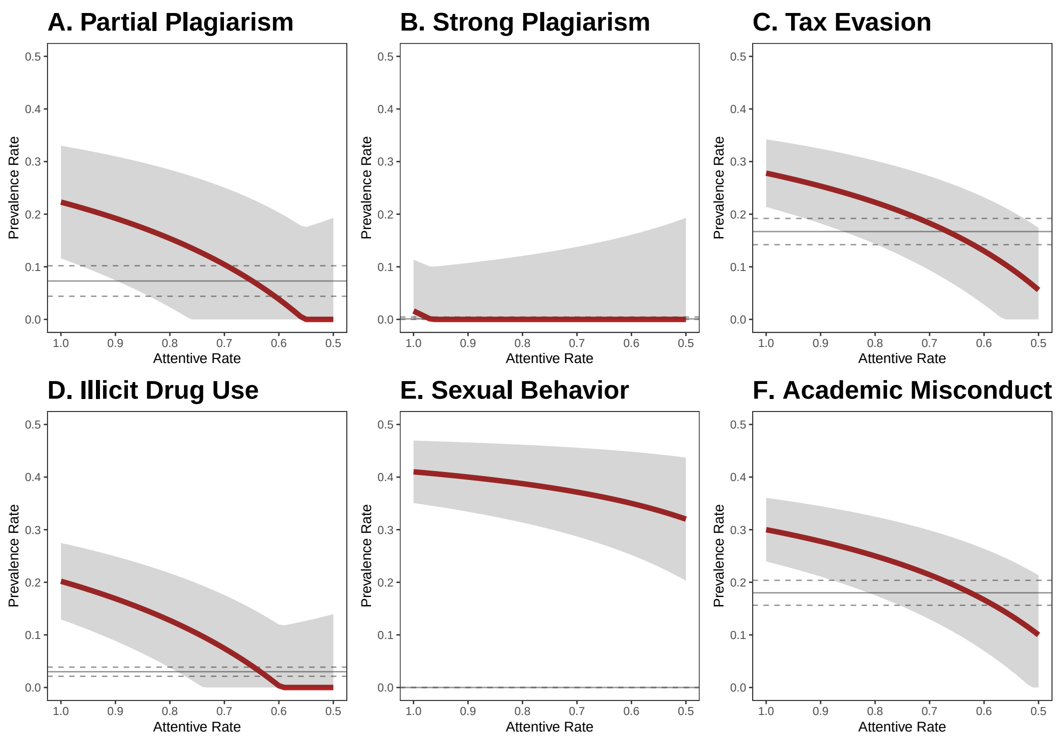

Figure 4: **Sensitivity Analysis of Previous Estimates**
Note: This figure shows the results of sensitivity analysis for six crosswise estimates. For each estimate, the bias correction is applied with varying levels of attentive rate under Assumption 1 (i.e., $\kappa = 0.5$). The confidence intervals are obtained via a normal approximation assuming that the population level attentive rate is known. Data come from surveys about partial and severe plagiarism (Jann, Jerke and Krumpal, 2011), tax evasion (Korndörfer, Krumpal and Schmukle, 2014), illicit drug use (Shamsipour et al., 2014), sexual behavior (Vakilian, Mousavi and Keramat, 2014), and academic misconduct (Höglinger, Jann and Diekmann, 2016).

The results are rather remarkable. They demonstrate that, in many cases, higher prevalence estimates may have been artifacts of the presence of inattentive responses rather than the property of crosswise models that mitigate social deniability bias. Our sensitivity analysis implies that most studies do not find any statistically significant difference between direct questioning and crosswise models unless they make an assumption that the population attentive rate is greater than 0.8. In other words, in order to claim that crosswise estimators do any better than direct questioning, researchers must assume that more than 85-90%

of the respondents were attentive and followed the instruction and met Assumption 1.

Our program for 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 (TRUE-TRUE/FALSE-FALSE) as indicated by Equation (4). 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.

5.3 Weighting in Crosswise Models

Weighting is a technique to estimate population level quantities based on sample statistics obtained through unrepresentative samples. While the methodological literature of sensitive inquiries usually assumes that samples are generated by a simple random sampling, a growing share of surveys is administered with unrepresentative platforms 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 inference into the population of interest. The benefit of using weighting is even larger for sensitive inquiries since sensitive questions are not usually asked on nationally representative samples and researchers need to conduct their own surveys. To date, however, no research has provided a practical guide about how to include population weights into randomized response estimates.

We argue that our bias-corrected crosswise estimator can incorporate population weights in a straightforward way. Recall that what we actually observe in crosswise models are λ and λ'. The idea is that we calculate weighed averages of these quantities, instead of simple arithmetic means, where weights reflect population 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 population weights in the following way:

$$\widehat{\lambda}_w = \sum_{i=1}^{n} w_i Y_i \tag{18}$$

$$\widehat{\gamma}_w = \frac{\sum_{i=1}^{n} w_i A_i - \frac{1}{2}}{\frac{1}{2} - p'}, \tag{19}$$

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 by their favorite weighting techniques such as raking (or iterative proportional fitting), matching, propensity score weighting, or sequential applications of them. 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 be also 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 must explore the optimal strategy to use MRP in sensitive inquiries.

6 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 crosswise models, 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 5. 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 crosswise models 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

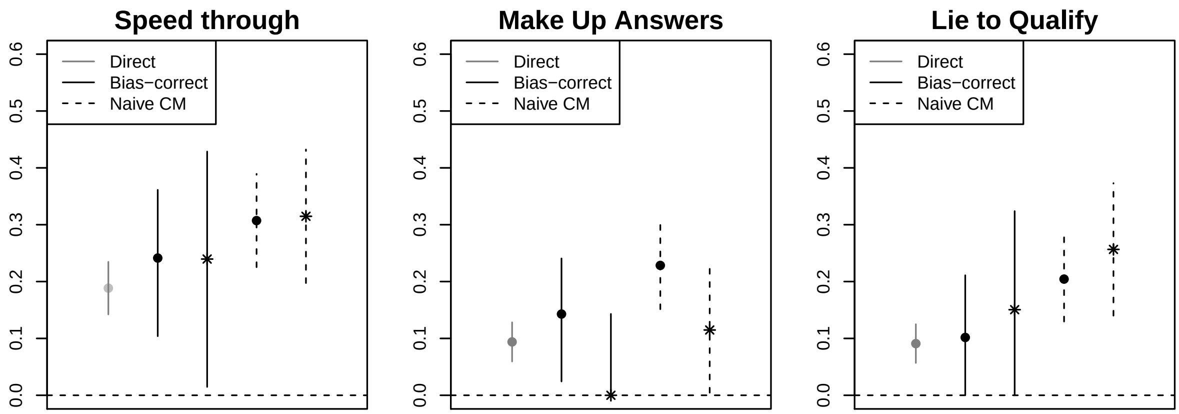

Figure 5: **Comparison of Prevalence Estimates**
Note: This graph visualizes the estimated prevalence of sensitive attributes based on direct inquiry, bias-corrected estimator, and naïve crosswise estimator. For crosswise estimates, dots (second and fourth from the left) are based on low randomization probabilities ($p = 0.086$) and asterisks (third and fifth from the left) are based on moderate randomization probabilities ($p = 0.25$).

logistic-type regression using crosswise responses with the randomization probability of 0.25. Column 1 of Table 2 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.

	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 2: **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 2 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.

7 Practical Guide: What Should We Care at the Survey Design Stage?

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 the assumptions.

7.1 How to Ensure Random Pick (Assumption 1)

The random pick assumption states that inattentive responses 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.

7.2 How to Achieve Constant Attentive Rate (Assumption 2)

The constant attentive rate assumption is satisfied when the sensitive question and its anchor question have the 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 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 is somehow different from another question in any way, the assumption could be violated. To satisfy the constant attentive rate, we recommend that researchers design both the sensitive and anchor questions look very similarly. 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 relative to sensitive questions will be helpful to guarantee that there is no carryover effect from one type of question to another.

7.3 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 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.[9] This strategy also helps researchers by separating the respondents' tasks of coming up with topics and thinking about questions.

7.4 Power Analysis

Power analysis is essential in choosing the adequate sample size to detect the population prevalence of sensitive attributes with acceptable margin of errors.

8 Concluding Remarks

In this article, 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

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

conventional crosswise estimator. In particular, the bias caused by inattentive responses make researchers draw incorrect conclusions that the usage of crosswise models 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 multivariate regression models, where bias-corrected estimates are used either outcomes or predictors, sensitivity analysis, and weighting methods were also provided. We also offered several suggestions on how to design crosswise surveys in order to enable our statistical adjustment.

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.

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.

Efron, Bradley. 1982. *The jackknife, the bootstrap, and other resampling plans*. Vol. 38 Siam.

Efron, Bradley and Robert J Tibshirani. 1994. *An introduction to the bootstrap*. CRC press.

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* .

Montgomery, Jacob M, Brendan Nyhan and Michelle Torres. 2018. "How conditioning on posttreatment variables can ruin your experiment and what to do about it." *American Journal of Political Science* 62(3):760–775.

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

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

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

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

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

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

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

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.

Appendix

A Derivation of the Bias

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

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

B Derivation of the Variance

Here, we derive the population and sample variance of the bias-corrected crosswise estimator discussed in Section 3. Rearranging Equation (6) and Equation (8), we obtain

$$\begin{aligned}
\mathbb{V}(\widehat{\pi}_{BC}) &= \mathbb{V}\left[\frac{\widehat{\lambda} - (1-p)\widehat{\gamma} - \frac{1}{2}(1-\widehat{\gamma})}{(2p-1)\widehat{\gamma}}\right] \\
&= \frac{1}{(2p-1)^2}\mathbb{V}\left[\frac{\widehat{\lambda} - (1-p)\widehat{\gamma} - \frac{1}{2}(1-\widehat{\gamma})}{\widehat{\gamma}}\right] \\
&= \frac{1}{(2p-1)^2}\mathbb{V}\left[\frac{\widehat{\lambda}}{\widehat{\gamma}} - (1-p) - \frac{1}{2\widehat{\gamma}} + \frac{1}{2}\right] \\
&= \frac{1}{(2p-1)^2}\mathbb{V}\left[\frac{2\widehat{\lambda}-1}{2\widehat{\gamma}}\right] \\
&= \frac{1}{(2p-1)^2}\mathbb{V}\left[(2\widehat{\lambda}-1)\left(\frac{\frac{1}{2}-p}{2\widehat{\lambda}'-1}\right)\right] \quad \text{(By Equation (10))} \\
&= \frac{(\frac{1}{2}-p)^2}{(2p-1)^2}\mathbb{V}\left[\frac{2\widehat{\lambda}-1}{2\widehat{\lambda}'-1}\right] \\
&= \frac{4(\frac{1}{2}-p)^2}{(2p-1)^2}\mathbb{V}\left[\frac{\widehat{\lambda}}{\widehat{\lambda}'-\frac{1}{2}}\right] \\
&= \mathbb{V}\left[\frac{\widehat{\lambda}}{\widehat{\lambda}'-\frac{1}{2}}\right] \\
&= \mathbb{E}\left[\frac{\widehat{\lambda}^2}{(\widehat{\lambda}'-\frac{1}{2})^2}\right] - \left(\mathbb{E}\left[\frac{\widehat{\lambda}}{\widehat{\lambda}'-\frac{1}{2}}\right]\right)^2
\end{aligned}$$

Now, by Assumption 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,

$$
\begin{aligned}
\mathbb{V}(\widehat{\pi}_{BC}) = &\left(\frac{\lambda(1-\lambda)}{N}\lambda\right)(1-k')^N \sum_{n=0}^{N} \binom{N}{n} \frac{1}{\left(n - \frac{1}{2}\right)^2} \left(\frac{k'}{1-k'}\right)^n \\
&+ \left(\lambda (1-k')^N \sum_{n=0}^{N} \binom{N}{n} \frac{1}{\left(n - \frac{1}{2}\right)} \left(\frac{k'}{1-k'}\right)^n\right)^2
\end{aligned} \tag{20}
$$

and its sample analog,

$$
\begin{aligned}
\widehat{\mathbb{V}(\widehat{\pi}_{BC})} = &\left(\frac{\widehat{\lambda}(1-\widehat{\lambda})}{N}\lambda\right)(1-k')^N \sum_{n=0}^{N} \binom{N}{n} \frac{1}{\left(n - \frac{1}{2}\right)^2} \left(\frac{k'}{1-k'}\right)^n \\
&+ \left(\widehat{\lambda} (1-k')^N \sum_{n=0}^{N} \binom{N}{n} \frac{1}{\left(n - \frac{1}{2}\right)} \left(\frac{k'}{1-k'}\right)^n\right)^2
\end{aligned} \tag{21}
$$