PLOS ONE

RESEARCH ARTICLE

Nothing but the truth? Effects of faking on the validity of the crosswise model

Adrian Hoffmann*, Julia Meisters, Jochen Musch

Department of Experimental Psychology, University of Duesseldorf, Duesseldorf, Germany

* adrian.hoffmann@uni-duesseldorf.de

OPEN ACCESS

Citation: Hoffmann A, Meisters J, Musch J (2021) Nothing but the truth? Effects of faking on the validity of the crosswise model. PLoS ONE 16(10): e0258603. https://doi.org/10.1371/journal.pone.0258603

Editor: Paolo Roma, Sapienza, University of Rome, ITALY

Received: May 31, 2021

Accepted: September 30, 2021

Published: October 29, 2021

Data Availability Statement: The data are now publicly available via the Open Science Framework (see https://doi.org/10.17605/OSF.IO/KJZQR).

Funding: This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation, https://www.dfg.de/), grant numbers 393108549 and 439602023 (awarded to Adrian Hoffmann and Jochen Musch). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

In self-reports, socially desirable responding threatens the validity of prevalence estimates for sensitive personal attitudes and behaviors. Indirect questioning techniques such as the crosswise model attempt to control for the influence of social desirability bias. The crosswise model has repeatedly been found to provide more valid prevalence estimates than direct questions. We investigated whether crosswise model estimates are also less susceptible to deliberate faking than direct questions. To this end, we investigated the effect of "fake good" instructions on responses to direct and crosswise model questions. In a sample of 1,946 university students, 12-month prevalence estimates for a sensitive road traffic behavior were higher and thus presumably more valid in the crosswise model than in a direct question. Moreover, "fake good" instructions severely impaired the validity of the direct questioning estimates, whereas the crosswise model estimates were unaffected by deliberate faking. Participants also reported higher levels of perceived confidentiality and a lower perceived ease of faking in the crosswise model compared to direct questions. Our results corroborate previous studies finding the crosswise model to be an effective tool for counteracting the detrimental effects of positive self-presentation in surveys on sensitive issues.

Introduction

When questioned about sensitive personal attributes, some individuals tend to answer in line with social norms rather than truthfully. Socially desirable responding results in an underestimation of socially undesirable attributes and an overestimation of socially desirable attributes [1–4]. To overcome social desirability bias, indirect questioning formats such as randomized response techniques [RRT; 5] have been proposed. RRTs grant full confidentiality to respondents by adding random noise to their answers. In the original RRT format, questions present participants with a sensitive statement A (e.g., "I have done cocaine"), and its negation, statement B (e.g., "I have never done cocaine"). Based on the outcome of a randomization procedure (e.g., the roll of a die), participants are instructed to respond to either statement A with probability p (e.g., "Respond to statement A if you rolled 1 or 2"; $p = 2/6$), or to statement B with probability 1-p (e.g., "Respond to statement B if you rolled any other number"; 1-$p = 4/6$). As the randomization outcome remains unknown to the experimenter, neither a "true" nor a "false" response reveals the respondent to be a carrier of the sensitive attribute. This is expected

Competing interests: The authors have declared that no competing interests exist.

to increase the respondent's motivation to answer truthfully. Moreover, since the probability p of selecting statement A is known, an estimate for the prevalence of the sensitive attribute can be obtained on the sample level. This estimate is assumed to be more valid than estimates obtained in conventional surveys, as it may be less biased by socially desirable responding [5].

Numerous studies have shown that conventional direct questions (DQ) do indeed lead to an underestimation of the prevalence of socially undesirable attributes. In contrast, RRT questions result in higher estimates that are considered to be less biased and hence more valid [e.g., 6,7–11]. Furthermore, meta-analyses have shown that RRT estimates often exceed DQ estimates, but still underestimate the true prevalence in studies in which the prevalence of the sensitive attribute is known [12]. Some studies even report RRT estimates comparable to, or lower than, DQ estimates [e.g., 13,14]. It is possible that such findings are due to respondents not understanding or not trusting the rather complex RRT procedure [15–17].

Consequently, nonrandomized response techniques [NRRT; 18], a recent advancement on RRT, promise to improve participants' trust and understanding by providing simplified and more easily comprehensible instructions. The most frequently used NRRT is the crosswise model [CWM; 19]. Questions in the CWM format include two statements: a sensitive statement A (e.g., "I have done cocaine"), and a nonsensitive statement B with known prevalence (e.g., "I was born in November or December"; $p = .158$ according to official birth statistics provided by the German Federal Statistical Office). Participants are instructed to indicate whether "both statements are true, or both statements are false", or "exactly one statement is true, irrespective of which one". As in the original RRT, none of the answer options expose a respondent to be a carrier of the sensitive attribute. However, the prevalence of the sensitive attribute (π) can be estimated on the sample level using the formula:

$$\hat{\pi}_{\mathrm{CWM}} = \frac{\hat{\lambda}_{\mathrm{CWM}} + p - 1}{2 * p - 1} \tag{1}$$

where $\hat{\lambda}_{\mathrm{CWM}}$ is the observed proportion of respondents choosing the first answer option ("both statements are true, or both statements are false").

The CWM has repeatedly been found to provide higher and, therefore, potentially more valid estimates for the prevalence of sensitive attributes than direct questions [e.g., 9,20–28]. Building on 45 empirical studies following this "more is better" validation approach, a recent meta-analysis supported the superiority of the CWM over DQ in controlling for the influence of socially desirable responses [29]. In addition, the CWM successfully recovered the true value in a "strong" validation study involving a sensitive attribute with known prevalence that served as an external validation criterion. In contrast, a direct question led to a vast underestimation [30]. However, more critical evaluations of the CWM have also suggested that the model is sometimes incapable of controlling for the influence of social desirability bias [e.g., 31–33]. More importantly, the CWM has been demonstrated to sometimes produce substantial shares of false positives (that is, non-carriers of the sensitive attribute falsely being classified as carriers) and false negatives (that is, carriers of the sensitive attribute falsely being classified as non-carriers) [34,35]. Both false positives and false negatives are likely attributable to respondents failing to understand the CWM instructions and choosing their answer at random. Such answering behavior potentially distorts prevalence estimates towards 50%, and is therefore especially problematic when socially undesirable attributes with a zero, or very low, prevalence are investigated. In these cases, a substantial share of random responses can inflate CWM estimates and thereby lead to problematic overestimations [36,37]. A recent investigation of the CWM has, however, shown that for undesirable attributes with a prevalence well above 0%, the inflating effect of false positives was outweighed by a stronger deflating influence

of false negatives, resulting in an overall under- rather than an overestimation of the true prevalence [38]. Taken together, the current empirical and meta-analytic evidence suggests that CWM prevalence estimates for socially undesirable attributes are not necessarily perfect, but likely closer to the true value than DQ estimates, and therefore usually more valid.

The overall mostly positive results are presumably due to the comparatively high comprehensibility of CWM instructions. Unlike the original RRT, the CWM integrates the required randomization directly into the answer options, thereby eliminating the need for an external randomization device. This arguably makes the method easier to administer for both the interviewer and the interviewee [19]. Supporting this assumption, the CWM has been evaluated as the most comprehensible format among several indirect questioning techniques and exhibits a significant improvement on perceived privacy protection compared to direct questions [39].

Most importantly, however, the CWM offers symmetrical answer options. None of the available answers represents a "safe" alternative respondents can choose in order to explicitly deny being a carrier of the sensitive attribute. This property of the model can be demonstrated by computing the conditional probabilities [40,41] of being identified as a carrier of the sensitive attribute when choosing the first ("both statements are true, or both statements are false") versus the second answer option ("exactly one statement is true, irrespective of which one") using Bayes' formula [42]:

$$Pr_{\text{CWM}}(\text{carrier}|\text{"both/none true"}) = \frac{Pr_{\text{CWM}}(\text{carrier} \cap \text{"both/none true"})}{Pr_{\text{CWM}}(\text{"both/none true"})} \tag{2.1}$$

$$Pr_{\text{CWM}}(\text{carrier}|\text{"one true"}) = \frac{Pr_{\text{CWM}}(\text{carrier} \cap \text{"one true"})}{Pr_{\text{CWM}}(\text{"one true"})} \tag{2.2}$$

These equations can be reformulated using the parameters for prevalence estimation from Eq 1:

$$Pr_{\text{CWM}}(\text{carrier}|\text{"both/none true"}) = \frac{\hat{\pi}_{\text{CWM}} * \text{p}}{\hat{\lambda}_{\text{CWM}}} \tag{2.3}$$

$$Pr_{\text{CWM}}(\text{carrier}|\text{"one true"}) = \frac{\hat{\pi}_{\text{CWM}} * (1 - \text{p})}{(1 - \hat{\lambda}_{\text{CWM}})} \tag{2.4}$$

As can be seen in Eqs 2.3 and 2.4, the probability of being identified as a carrier of the sensitive attribute exceeds zero regardless of whether respondents choose the first or the second answer option for all cases of $0 < \hat{\pi}_{\text{CWM}} < 1$, $0 < p < 1$, and $0 < \hat{\lambda}_{\text{CWM}} < 1$. These conditions are usually met in applications of the CWM, because researchers typically ensure that the expected prevalence of the sensitive attribute, the randomization probability, and the proportion of respondents choosing the first answer option are different from 0% and 100%. Under such conditions, no "safe" answer option is available for respondents to choose in order to explicitly deny being a carrier of the sensitive attribute. Despite the absence of an objectively safe answer, respondents confronted with a CWM question might nevertheless try to assess the risk of being classified as a carrier of the sensitive attribute as a function of the answer option they decide to select. As neither the prevalence of the sensitive attribute ($\hat{\pi}_{\text{CWM}}$) nor the proportion of respondents choosing the first answer option ($\hat{\lambda}_{\text{CWM}}$) is known until data collection is complete, it is impossible for respondents to calculate the exact conditional probabilities depicted in Eqs 2.3 and 2.4, respectively. As can, however, also be seen when comparing the numerators of Eqs 2.3 and 2.4, the relative risk of being identified as a carrier is lower when

choosing the first ("both/none true") rather than the second answer option ("one true") in all cases of $0 < p < .5$, because $p < 1 - p$. For $.5 < p < 1$, the second answer option is associated with a lower risk because $p > 1 - p$. Consequently, participants following a self-protective response strategy could try to compare the relative risk of the two answer options based on the randomization probability, and then choose the less risky option. To do so, they would have to (a) correctly estimate the randomization probability p, and (b) derive and understand the relationship between the randomization probability and the conditional probabilities of being identified as a carrier when choosing either of the two answer options. Previous studies have shown that most respondents are poor at estimating the relationship between the randomization probability and the objective privacy protection afforded by indirect questioning techniques, except for cases in which extreme randomization probabilities eliminate confidentiality [43]. In addition, we believe that the necessary logical deductions and calculations are too time-consuming for common short survey situations. We therefore argue that it is improbable that a substantial share of respondents confronted with a CWM question will succeed in identifying a self-protective response. Because response symmetry also reduces the incentive to provide untruthful answers [44] and because the symmetrical CWM has been shown to provide potentially more valid estimates than a related model with asymmetric response options [26], we propose that the symmetry of the CWM will lead to a higher proportion of honest responses compared to a direct question, even or especially if respondents are incentivized to provide self-protective responses by asking them to make a positive impression. Our expectation was that respondents trying to follow the instruction to provide self-protective responses would be able to successfully change their answers to a direct question in order to make a positive impression, but would be much less successful when answering indirect CWM questions. We therefore surmised that "fake good" instructions would severely impair the validity of direct questioning estimates, whereas crosswise model estimates would be much less vulnerable to a deliberate faking attempt.

As a sensitive attribute, we chose to ask participants whether they had crossed the street on a "Don't Walk" sign in plain sight of children within the past twelve months. Crossing the street on a "Don't Walk" sign is a common behavior among German and Austrian adults, but is exhibited less frequently when children are present [45,46]. This is presumably because adults are aware of their function as a role model in street-crossing behavior for their own as well as other people's children [46]. Nevertheless, a considerable portion of the population still crosses the street on a "Don't Walk" sign even when in plain sight of children [46–48]. We therefore expected that a substantial share of the participants in our sample had exhibited this behavior in the past twelve months, but would feel reluctant to admit it due to self-presentation concerns.

Previous studies on the CWM have tended to investigate the validity of its prevalence estimates or variables that can elicit truthful responses, such as trust or understanding of the method. The present study extends these findings by conducting the first experimental examination of the influence of deliberate positive self-presentation. To this end, we obtained DQ and CWM prevalence estimates for a sensitive attribute in an "honest" control condition, in which participants were instructed to respond truthfully, and compared these estimates to those obtained in an experimental "fake good" condition, in which participants were instructed to respond in a way that would leave a positive impression. Similar "fake good" manipulations have successfully been employed to investigate the influence of positive self-presentation on, for example, the validity of personality tests [49–52], social desirability scales [53], gender differences in self-presentation [54], and prejudice [55]. Our decision to use an experimental "fake good" manipulation pursued the goal of maximizing experimental control. We had to accept that our decision to prioritize internal validity came at the cost of decreasing

ecological validity compared to real-world situations differing in their incentive to fake. However, any non-experimental approach would have provided less internal validity. We expected that direct self-reports in the DQ condition would be prone to the influence of social desirability, resulting in a substantial underestimation of the prevalence of crossing the street on a "Don't Walk" sign in plain sight of children. We expected CWM estimates to be higher than DQ estimates to the extent to which indirect questioning is capable of controlling for socially desirable responding. Applying the "more is better"-criterion, these higher estimates can be considered more valid and presumably less biased [56]. We also expected that deliberate attempts to fake good would severely impair the validity of DQ estimates, leading to considerably lower and therefore presumably less valid estimates in the "fake good" condition. Finally, if the CWM is indeed robust against deliberate faking due to respondents' inability to identify a self-protective answer, a much smaller (if any) difference between the "honest" and "fake good" conditions can be expected for CWM estimates.

In summary, the present study is the first to investigate the influence of a "fake good" manipulation on the validity of prevalence estimates obtained via indirect questioning techniques. As expected, we could show that deliberate positive self-presentation severely impaired the validity of prevalence estimates obtained via a conventional direct question, while estimates based on the CWM were largely unaffected.

Methods

Participants

A total of 2,024 subjects participated in our survey. Due to incomplete information on the questionnaire, 78 participants (3.9%) had to be excluded from further analyses. The final sample thus consisted of $N = 1{,}946$ respondents, of which 52.4% were female. The mean age was 20.9 years ($SD = 4.58$). Participants were recruited and assessed in lecture halls and public spaces at the Universities of Duesseldorf (71.2%), Aachen (21.2%) and Cologne (7.7%), Germany. The survey was carried out in accordance with the revised Declaration of Helsinki [57] and the ethical guidelines of the German Association of Psychologists and the German Psychological Society [58]. Written information on the survey content, the strict anonymization of all personal data, and the exclusive use of the collected data for research purposes was provided directly on the questionnaire. Potential participants were instructed to document their consent by filling out and returning the questionnaire to the experimenter; they were explicitly asked not to return the questionnaire should they not consent to participate. Participation was voluntary and not associated with any risk of physical or mental harm or discomfort beyond participants' everyday experiences. Therefore, ethics committee approval was not required according to the "Ethical Research Principles and Test Methods in the Social and Economic Sciences" formulated by the Ethics Research Working Group of the German Data Forum [59] and the "Ethical recommendations of the German Psychological Society for researchers and ethics committees" [60]. A small proportion of participants (1.7%) were underage because in Germany, some pupils finish high school and start attending university prior to their 18th birthday. For these students, parents had provided written consent for them to partake in all study-related activities, including participation in the present study.

The prevalence of the sensitive attribute at the population level, as well as the prevalence estimates in each subgroup, were unknown before the study was conducted. Therefore, a priori power considerations regarding the required sample size were initially made so that an acceptable power ($1\text{-}\beta > .80$) for detecting any proportion of carriers of the sensitive attribute in the direct and indirect questioning conditions ($\hat{\pi} > 0\%$) would likely be achieved. These considerations revealed that a total sample size of $N \geq 1{,}500$, and an allocation of twice as many

participants to the indirect compared to the direct questioning groups would ensure sufficient statistical power for the planned prevalence estimations [cf. 61].

For differences between experimental groups concerning the three additional variables–perceived sensitivity of the topic, perceived confidentiality, and subjective ease of faking–, we performed sensitivity analyses, as well as post-hoc power analyses, using the software G*Power 3 [62]. Based on our final sample size and a desired Type-I error probability of $\alpha = .05$, sensitivity analyses revealed that for all main and interaction effects in the ANOVAs, sufficient statistical power ($1\text{-}\beta > .80$) was achieved for a minimum effect size of $f \geq .06$, and thus even for small effects; post-hoc power analyses further confirmed that the achieved power was very high ($1\text{-}\beta > .99$) for large ($f \geq .40$), medium ($f \geq .25$), and small effects ($f \geq .10$). Sensitivity analyses, and post-hoc power analyses, for pairwise t-tests were performed applying Bonferroni corrections for multiple testing. As for each of the three additional variables, four pairwise comparisons were conducted, Bonferroni correction resulted in an adjusted Type-I error level of $\alpha_{adj} = \alpha / 4 = .0125$. Given the distribution of participants across experimental groups, sensitivity analyses revealed that the minimum effect size for which sufficient statistical power ($1\text{-}\beta > .80$) was achieved was small, and ranged from $d \geq 0.19$ for the comparison of "CWM, honest" to "CWM, fake good" conditions, to $d \geq 0.26$ for the comparison of "DQ, honest" to "DQ, fake good" conditions. Post-hoc power analyses further showed that the achieved power was very high ($1\text{-}\beta > .99$) for all pairwise comparisons when assuming large ($d \geq .80$) or medium effects ($d \geq .50$). For small effects ($d \geq .20$), sufficient power was achieved for the comparison of "CWM, honest" to "CWM, fake good" conditions ($1\text{-}\beta = .86$); however, power was less optimal for the comparisons of "DQ, honest" to "CWM, honest" conditions ($1\text{-}\beta = .68$), of "DQ, fake good" to "CWM, fake good" conditions ($1\text{-}\beta = .67$), and of "DQ, honest" to "DQ, fake good" conditions ($1\text{-}\beta = .53$).

Survey design

All instructions and questions were placed on a single-page paper-pencil questionnaire. The introduction section informed participants that they would be asked questions about serving as a role model for children in everyday traffic situations. In the experimental part of the questionnaire, a black-rimmed box provided "honest" versus "fake good" instructions and posed the experimental question in either a direct (DQ) or crosswise model (CWM) format. At the bottom of the questionnaire, five additional questions assessed participants' demographics and experience of the survey. The 2x2 between-subjects design resulted in four versions of the questionnaire, which differed only with respect to the content of the central black-rimmed box. Depending on condition, this box contained (a) the instruction to respond honestly to the question in DQ format; (b) the instruction to respond honestly to the question in CWM format; (c) the instruction to fake good when answering the question in DQ format, or (d) the instruction to fake good when answering the question in CWM format. To compensate for the lower efficiency of indirect questioning techniques resulting from the required randomization, twice as many participants were assigned to the CWM conditions (b) and (d) than to the DQ conditions (a) and (c) [cf. 61,63]. The distribution of participants across experimental conditions is shown in Table 1. The original German questionnaire used for data collection along with an English translation is available on the Open Science Framework (https://doi.org/10.17605/OSF.IO/KJZQR).

Instructions to respond honestly versus fake good. In the honest conditions (a) and (b), the following text was printed in bold red letters to attract participants' attention and placed right before the experimental question: "We are interested in the incidence of certain behaviors in traffic situations. Hence, please respond honestly to the question in this black-skimmed

Table 1. Distribution of participants across experimental conditions.

		Questioning technique		Total
		DQ	CWM	
Instruction	"honest"	$n_{(a)}$ = 334 (17.16%)	$n_{(b)}$ = 637 (32.73%)	971 (49.90%)
	"fake good"	$n_{(c)}$ = 326 (16.75%)	$n_{(d)}$ = 649 (33.35%)	975 (50.10%)
Total		660 (33.92%)	1286 (66.08%)	

DQ = direct questioning, CWM = crosswise model.

https://doi.org/10.1371/journal.pone.0258603.t001

box, and report your actual previous behavior." In the fake good conditions (c) and (d), this passage read: "We are interested in how dishonest responding affects survey results. Hence, please do not respond honestly to the question in this black-skimmed box, but in a way that will leave as positive an impression of yourself as possible."

Questioning technique. In the DQ conditions (a) and (c), we presented a single sensitive statement for which the prevalence π_{DQ} was to be estimated: "Within the past 12 months, I have crossed the street on a 'Don't Walk' sign even though I was in plain sight of a child." Participants were asked to indicate whether this statement was "true" or "false". In the CWM conditions (b) and (d), we presented two statements simultaneously – a sensitive statement A for which the prevalence π_{CWM} was to be estimated, and a nonsensitive statement B with known prevalence p (Statement A: "Within the past 12 months, I have crossed the street on a 'Don't Walk' sign even though I was in plain sight of a child."; Statement B: "I was born in November or December."). The prevalence for statement B was known to be $p = .158$ according to official birth statistics. Participants had to choose between the two answer options "I agree with both statements or with none of the statements", and "I agree with exactly one statement (irrespective of which one)".

Additional variables. To further explore the participants' experience, we included three additional questions asking about the perceived sensitivity of the question topic, the perceived confidentiality offered by the questioning technique used in the respective condition, and the subjective ease of faking on the questionnaire.

To assess the perceived sensitivity of the topic, we asked: "How bad do you think it is when an adult crosses the street on a 'Don't Walk' sign in plain sight of children?" Responses were recorded on a 7-point Likert scale ranging from "not bad at all" (1) to "very bad" (7). To measure the perceived confidentiality offered by the questioning technique, we presented the question: "How well do you think the confidentiality of your answer is protected in the above question?" Subjects answered on a 7-point Likert scale ranging from "confidentiality is not granted at all" (1) to "confidentiality is granted in an optimal way" (7). To assess the subjective ease of faking on the questionnaire, we asked: "How easy do you think it is to answer the above question in such a way that you give the impression that you have never crossed the street on a 'Don't Walk' sign in plain sight of children?" Subjects were required to indicate their response on a 7-point Likert scale ranging from "very easy" (1) to "very hard" (7).

Statistical analyses

To estimate the prevalence of the sensitive attribute, we used multinomial processing tree models [64,65] following the procedure detailed in, for example, [7,63,66]. To evaluate the influence of the independent variables *instruction* ("honest" versus "fake good") and

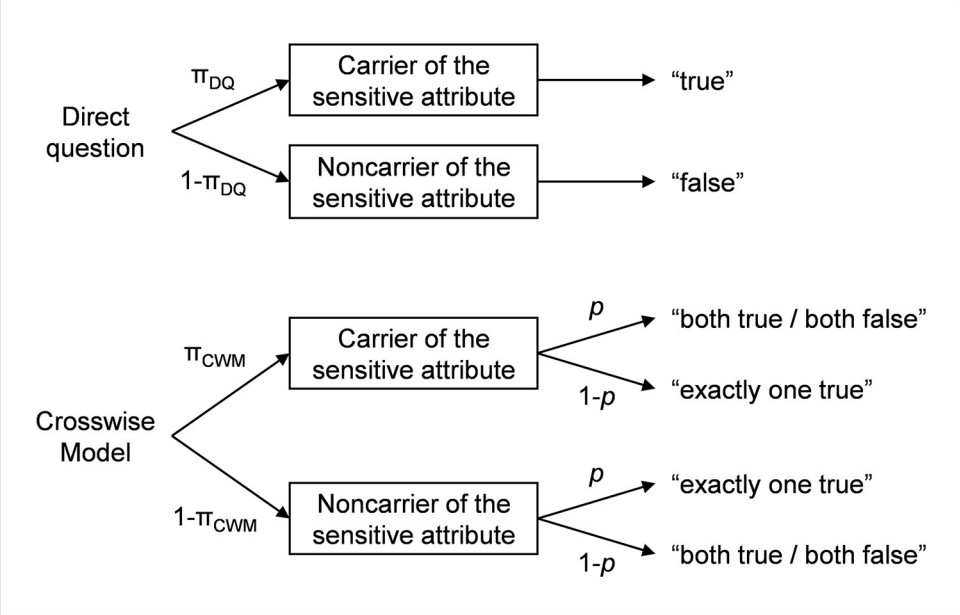

Fig 1. Tree diagram of the multinomial processing trees established for the direct questioning and the crosswise model conditions, respectively. π = unknown prevalence of the sensitive attribute, *p* = known randomization probability.

https://doi.org/10.1371/journal.pone.0258603.g001

questioning technique ("DQ" versus "CWM"), separate multinomial processing trees were formulated for each experimental condition (a) to (d). Within each processing tree, a parameter π represented the prevalence estimate of the sensitive attribute. In the CWM conditions, an additional parameter *p* reflected the probability of being born in November or December, which was used for randomization. This probability was known to be 15.8% from official birth statistics provided by the German Federal Statistical Office [67], and was thus set constant to *p* = .158. As an example, the processing trees established for the direct questioning versus crosswise model conditions are shown in Fig 1.

We computed maximum likelihood estimates for π on the basis of the empirically observed answer frequencies using the expectation maximization algorithm [68,69] as implemented in the software multiTree [70]. To assess differences between parameter estimates (for example, $\hat{\pi}_{DQ,\ \text{"honest"}}$ versus $\hat{\pi}_{CWM,\ \text{"honest"}}$), we compared unrestricted baseline models in which both parameters could be estimated freely to restricted alternative models in which both parameters were set to be equal (for example, $\pi_{DQ,\ \text{"honest"}} = \pi_{CWM,\ \text{"honest"}}$). In these model comparisons, significant changes in the asymptotically X^2-distributed log-likelihood statistic G^2 indicate that the restricted alternative model fits the data worse than the baseline model. If this is the case, the parameter restriction in the alternative model is shown to be inadmissible, and the two parameter estimates are shown to differ significantly (for example, $\hat{\pi}_{DQ,\ \text{"honest"}} \neq \hat{\pi}_{CWM,\ \text{"honest"}}$).

To investigate a potential interaction between instruction and questioning technique, we introduced parametric order constraints [71] by reparameterizing the original model as detailed in Hoffmann and Musch (28). In the reparameterized model, the shrinkage parameter α_{DQ} represented the ratio of the estimated prevalence in the "DQ, 'fake good'" condition ($\pi_{DQ,\ \text{"fake good"}}$) to the estimated prevalence in the "DQ, 'honest'" condition ($\pi_{DQ,\ \text{"honest"}}$); likewise, parameter α_{CWM} represented the ratio of the estimated prevalence in the "CWM, 'fake good'" condition ($\pi_{CWM,\ \text{"fake good"}}$) to the estimated prevalence in the "CWM, 'honest'" condition ($\pi_{CWM,\ \text{"honest"}}$). A significant difference between the estimated shrinkage ratios $\hat{\alpha}_{DQ}$ and $\hat{\alpha}_{CWM}$,

assessed via a G^2 test as described above, indicated a significant interaction between instruction and questioning technique.

The effects of instruction ("honest" versus "fake good") and questioning technique (DQ versus CWM) on the three additional variables perceived sensitivity of the topic, perceived confidentiality, and subjective ease of faking were assessed via three 2x2 between-subjects ANOVAs. Pairwise comparisons between specific experimental groups (for example, "DQ, 'honest'" versus "DQ, 'fake good'") were conducted via t-tests for independent samples; *p*-values for these pairwise comparisons were Bonferroni-corrected to account for multiple testing.

A raw dataset containing respondents' answers to the experimental and additional questions, as well as all multinomial model equations and empirically observed answer frequencies necessary to reproduce the parameter estimates reported in this manuscript, are available on the Open Science Framework (https://doi.org/10.17605/OSF.IO/KJZQR).

Results

Main results: Prevalence of the sensitive attribute

Parameter estimates and parameter comparisons for the sensitive attribute are shown in Table 2. As expected, the prevalence estimate for the sensitive attribute was substantially higher, and thus presumably more valid, in the conditions with a crosswise model question (CWM) rather than direct questioning (DQ). This was true for both "honest" and "fake good" instructions. Furthermore, "fake good" instructions resulted in substantially lower prevalence estimates than "honest" instructions in the DQ condition, but not in the CWM condition.

A significant difference between the shrinkage ratios $\hat{\alpha}_{DQ}$ and $\hat{\alpha}_{CWM}$ indicated a significant interaction between questioning technique and instruction (ΔG^2 [$df = 1$] = 18.12, $p < .001$). A shrinkage ratio of $\hat{\alpha}_{DQ}$ = 45% revealed that in the DQ condition, the estimated prevalence $\hat{\pi}_{DQ,\ \text{"fake good"}}$ was only .45 times the size of the estimated prevalence $\hat{\pi}_{DQ,\ \text{"honest"}}$; thus,

Table 2. Parameter estimates (standard errors in parentheses) and parameter comparisons for the prevalence of the sensitive attribute ("Within the past 12 months, I have crossed the street on a 'Don't Walk' sign even though I was in plain sight of a child.").

Parameter estimates		Questioning technique		
		DQ	CWM	
Estimated prevalence $\hat{\pi}$				
Instruction	"honest"	35.03% (2.61)	45.53% (2.89)	
	"fake good"	15.64% (2.01)	42.00% (2.85)	
Estimated shrinkage $\hat{\alpha}$				
"Fake good" relative to "honest" condition		44.66% (6.64)	92.26% (8.58)	
Parameter comparisons			Model fit	
		|*difference*|	ΔG^2 ($df = 1$)	*p*
Instruction				
	$\hat{\pi}_{DQ,\ \text{"honest"}} = \hat{\pi}_{DQ,\ \text{"fake good"}}$	19.39%	33.40	< .001 *
	$\hat{\pi}_{CWM,\ \text{"honest"}} = \hat{\pi}_{CWM,\ \text{"fake good"}}$	3.53%	0.75	.386
Questioning technique				
	$\hat{\pi}_{DQ,\ \text{"honest"}} = \hat{\pi}_{CWM,\ \text{"honest"}}$	10.50%	7.18	.007 *
	$\hat{\pi}_{DQ,\ \text{"fake good"}} = \hat{\pi}_{CWM,\ \text{"fake good"}}$	26.36%	52.97	< .001 *
*Instruction * Questioning technique*				
	$\hat{\alpha}_{DQ} = \hat{\alpha}_{CWM}$	47.60%	18.12	< .001 *

* significant at $p < .05$.

https://doi.org/10.1371/journal.pone.0258603.t002

participants were less than half as likely to admit to the sensitive attribute when instructed to fake good compared to when instructed to respond honestly. In the CWM condition, a shrinkage ratio of $\hat{\alpha}_{CWM} = 92\%$ suggested that the estimated prevalence $\hat{\pi}_{CWM,\ \text{"fake good"}}$ was .92 times the size of, and therefore roughly comparable to, the estimated prevalence $\hat{\pi}_{CWM,\ \text{"honest"}}$. Hence, fake good instructions severely impaired the validity of prevalence estimates obtained via DQ, but hardly affected the validity of estimates obtained via the CWM.

Exploratory analyses: Additional variables on subjective experience

Results for the additional variables perceived sensitivity of the topic, perceived confidentiality, and subjective ease of faking are shown in Table 3; plots of the observed means are shown in Fig 2.

Analyses of the perceived sensitivity question revealed that participants considered crossing the street on a 'Don't Walk' sign in plain sight of children to be rather poor behavior, as reflected in mean values of 5.07 to 5.28 on a scale from 1 ("not bad at all") to 7 ("very bad"). Bonferroni-corrected pairwise comparisons between groups did not find any effect of instruction or questioning technique on perceived sensitivity of the question; participants in all experimental groups considered the topic under investigation to be equally sensitive.

For perceived confidentiality, both instruction and questioning technique exhibited significant main effects. Overall, participants in the "honest" condition reported higher levels of perceived confidentiality than participants in the "fake good" condition. Furthermore, perceived confidentiality was higher in the CWM than in the DQ condition. A significant interaction between instruction and questioning technique revealed that the advantage of the CWM over DQ was only observed among participants instructed to "fake good"; for participants instructed to respond "honestly", the questioning technique did not affect perceived confidentiality.

With respect to the subjective ease of faking on the questionnaire, a significant main effect was found for questioning technique. Compared to participants in the DQ condition, participants in the CWM condition found it more difficult to respond in a way that did not make them appear to be a carrier of the sensitive attribute. This effect was found both in the "honest" and in the "fake good" conditions. Neither instruction nor the interaction between instruction and questioning technique affected the subjective ease of faking.

Discussion

Dishonest responses due to deliberate faking and social desirability bias threaten the validity of survey results on sensitive attitudes and behaviors. Indirect questioning techniques such as the crosswise model [CWM; 19] promise to control social desirability bias by granting respondents higher confidentiality with respect to their individual answers. In a direct test of the CWM's presumed ability to better control inflated self-presentations, we asked participants about a potentially sensitive behavior, whether they had crossed the street on a 'Don't Walk' sign in plain sight of children within the past twelve months. Prevalence estimates for this socially undesirable behavior were generally higher and therefore presumably more valid in the CWM than in the DQ condition. Moreover, to directly evaluate the influence of deliberate faking, this study was the first to experimentally investigate the influence of "fake good" instructions on the validity of results from DQ versus CWM questions. In the DQ condition, less than half as many participants admitted to the sensitive behavior when instructed to "fake good" than when instructed to respond "honestly", indicating that deliberate faking led to severe underestimation. In the CWM condition, however, prevalence estimates were unaffected by "honest" versus "fake good" instructions. Analyses of three additional measures of

Table 3. Descriptive statistics, ANOVA results, and pairwise t-test results for perceived sensitivity, perceived confidentiality, and subjective ease of faking.

Descriptive statistics		Questioning technique	
		DQ	CWM
		M (SE)	*M (SE)*
Perceived sensitivity of the topic (higher values = higher perceived sensitivity)			
Instruction	"honest"	5.27 (0.08)	5.28 (0.07)
	"fake good"	5.07 (0.09)	5.14 (0.07)
Perceived confidentiality (higher values = higher perceived confidentiality)			
Instruction	"honest"	4.66 (0.10)	4.78 (0.07)
	"fake good"	3.82 (0.12)	4.37 (0.08)
Subjective ease of faking (higher values = subjectively harder to fake)			
Instruction	"honest"	3.10 (0.11)	3.71 (0.07)
	"fake good"	2.83 (0.11)	3.73 (0.07)

ANOVAs	*F* (1, 1942)	*p*	*f*
Perceived sensitivity of the topic			
Instruction ("honest" vs. "fake good")	4.73	.030 *	.049
Questioning technique (DQ vs. CWM)	0.20	.653	.010
Instruction * Questioning technique	0.13	.719	.008
Perceived confidentiality			
Instruction ("honest" vs. "fake good")	47.81	< .001 *	.157
Questioning technique (DQ vs. CWM)	14.32	< .001 *	.086
Instruction * Questioning technique	5.75	.017 *	.054
Subjective ease of faking			
Instruction ("honest" vs. "fake good")	1.94	.164	.032
Questioning technique (DQ vs. CWM)	73.34	< .001 *	.194
Instruction * Questioning technique	2.70	.101	.037

Pairwise t-tests		*t*	*df*	$p^{\dagger}$	*d*
Perceived sensitivity of the topic					
DQ	"honest" vs. "fake good"	1.61	658	.430	0.126
CWM	"honest" vs. "fake good"	1.53	1284	.502	0.084
"Honest"	DQ vs. CWM	0.06	969	> .999	0.006
"Fake good"	DQ vs. CWM	0.57	973	> .999	0.042
Perceived confidentiality					
DQ	"honest" vs. "fake good"	5.42	658	< .001 *	0.424
CWM	"honest" vs. "fake good"	4.00	1284	< .001 *	0.226
"Honest"	DQ vs. CWM	1.03	969	> .999	0.066
"Fake good"	DQ vs. CWM	4.17	973	< .001 *	0.276
Subjective ease of faking					
DQ	"honest" vs. "fake good"	1.75	658	.323	0.136
CWM	"honest" vs. "fake good"	-0.22	1284	> .999	0.011
"Honest"	DQ vs. CWM	5.02	969	< .001 *	0.329
"Fake good"	DQ vs. CWM	7.04	973	<.001 *	0.466

* significant at $p < .05$.

† Bonferroni-corrected for multiple testing (original *p* values multiplied by 4).

https://doi.org/10.1371/journal.pone.0258603.t003

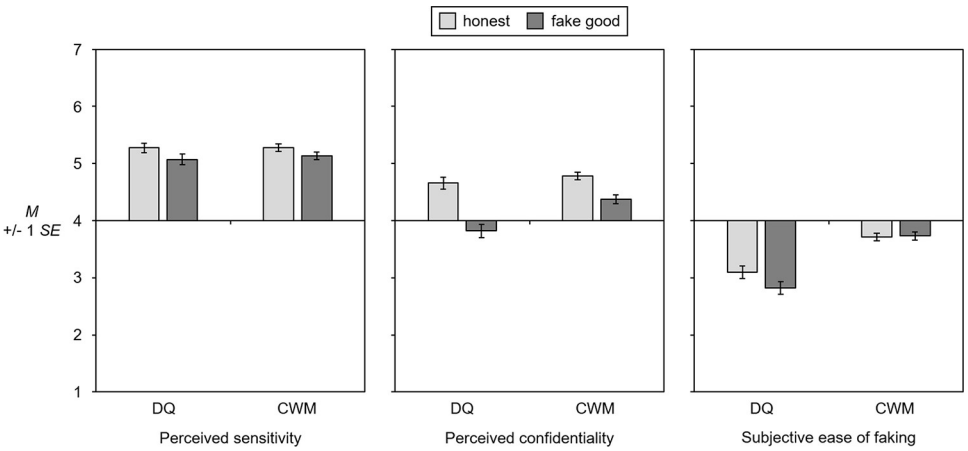

Fig 2. Mean plots for perceived sensitivity (higher values = higher sensitivity), perceived confidentiality (higher values = higher confidentiality), and subjective ease of faking (higher values = harder to fake). DQ = direct questioning, CWM = crosswise model.

https://doi.org/10.1371/journal.pone.0258603.g002

subjects' experience of the survey showed that crossing the street on a 'Don't Walk' sign in plain sight of children was indeed perceived as a sensitive behavior in all experimental groups; that CWM questions resulted in higher perceived confidentiality, though only under "fake good" instructions; and that participants in the CWM condition generally found it more difficult to fake their response in a way that would leave a positive impression.

Taken together, our results strongly suggest that direct self-reports of crossing the street on a 'Don't Walk' sign in plain sight of children are influenced by social desirability bias, and that indirect questioning can help to obtain more valid estimates of the prevalence of this behavior. Estimates in the CWM condition were presumably less biased, as they exceeded those in the DQ condition, sufficing the "more is better"-criterion [56]. This finding is consistent with several other positive evaluations of the technique [e.g., 9,21,23,26,27,30].

More importantly, however, and as an important extension to the existing literature, CWM questions were shown to be much more robust against deliberate faking than DQ questions. In the DQ condition, "fake good" instructions heavily impaired the validity of estimates; participants were quite successful in faking their response in a way that made a more positive impression. In contrast, CWM estimates did not differ between the "honest" and "fake good" conditions. This finding is consistent with reports of higher perceived confidentiality and reduced ease of faking in CWM compared to DQ.

With respect to higher perceived confidentiality under "fake good" conditions, we argue that the CWM's simple instructions are successful in helping respondents comprehend the rationale of the randomization process and how it protects the confidentiality of their answer. This interpretation is in line with a previous positive evaluation of the CWM's comprehensibility and perceived privacy protection [39]. Especially when instructed to respond in a way that will leave a positive impression (that is, to "fake good"), respondents seem to understand that in the CWM, they can provide a truthful response to an embarrassing question without making a negative impression. Unexpectedly, however, participants instructed to respond honestly did not report higher perceived confidentiality when confronted with a CWM rather than with a conventional direct question. As a potential explanation for this finding, participants might have taken general components of the survey situation into account when evaluating perceived confidentiality, such as the anonymity of their participation or the minimal amount of personal information they had to report. These extraneous factors might have masked a

potential effect of questioning techniques on perceived confidentiality. Conversely, the "fake good" instructions may have emphasized the sensitive nature of the behavior under investigation. They may thus have reminded participants that a direct question does not protect the confidentiality of their answers at all. Another potential explanation is that questioning technique was varied as a between-subjects factor in the current study. Participants in the DQ condition never saw a question in CWM format (and vice versa); hence, they were not able to establish a common frame of reference for these two conditions [cf. 72]. A within-subjects design would have presented participants with both question formats and would thus have allowed assessing the level of confidentiality they afford using the same frame of reference; this may have resulted in larger effect sizes. However, it was impossible to employ a within-subjects design in the current study, as directly answering the sensitive question first would have made a second presentation of the same question in an indirect format appear absurd. Future studies could combine a "fake good" manipulation with the scenario-based approach used in studies such as [39] to assess whether the two question formats differ in the perceived confidentiality they afford respondents.

As to the lower reported ease of faking, it seems that even if participants try to fake their answer to a CWM question, they cannot identify a self-protective response. We attribute this robustness towards deliberate faking to the symmetrical nature of the model [cf. 19]. As long as the prevalence of the sensitive attribute, the randomization probability, and the proportion of respondents choosing the first answer option are different from 0% and from 100% (all of which are true in the current study), there is no "safe" answer option respondents can choose to unambiguously deny being a carrier of the sensitive attribute. Moreover, it is impossible for respondents to assess the exact conditional probabilities of being identified as a carrier of the sensitive attribute when choosing either of the two answer options, as this would require knowledge of the prevalence of the sensitive attribute and the proportion of respondents choosing the first answer option. These values remain unknown to both the respondents and the experimenters until data collection is complete. Even if respondents tried to roughly assess the relative risk associated with either answer option, they would have to estimate the randomization probability p and derive and understand its effect on the conditional probability of being identified as a carrier when choosing either answer option. However, prior research has found that respondents are relatively poor at understanding the relationship between the randomization probability and the objective privacy protection in the context of indirect questioning techniques [43]. Against this background, we consider it highly unlikely that a substantial proportion of respondents will successfully make time-consuming inferences about the relative risk of the available answer options in short surveys. Hence, it seems that participants are doomed to fail when trying to "fake good" in the CWM.

The potential attribution of the robustness of the CWM to its response symmetry also relates to a first important limitation of our study. Due to limited resources, we decided to focus on the CWM because this model is associated with a comparatively high comprehensibility and perceived privacy protection, and an acceptable estimation efficiency [19,39]. These properties made it appear suitable for an initial investigation of the influence of "fake good" instructions on indirect questioning techniques. However, our comparison of CWM to DQ only allowed for the conclusion that the CWM is less susceptible to deliberate faking than direct questions. While the attribution of this robustness to the symmetry of the CWM seems highly plausible, this assumption needs to be tested explicitly in future studies that should also include asymmetric models such as, for example, the Triangular Model (TRM) [19]. In a recent comparison based on the "more is better"-criterion, the symmetric CWM has been shown to outperform the asymmetric TRM in terms of estimation validity [26]. For the proportion of respondents holding socially undesirable, Xenophobic attitudes, a higher and thus

potentially more valid estimate was achieved in a CWM compared to a TRM condition. As CWM and TRM questions are very similar and only differ concerning the symmetry of their answer options, the superiority of the CWM was explicitly attributed to its response symmetry. In contrast, the availability of a "safe" answer option in the asymmetric TRM potentially allowed respondents to follow a self-protective answering strategy, thereby leading to an underestimation of the prevalence of Xenophobia. In light of these findings, we expect that when compared in terms of susceptibility to deliberate faking, the symmetric CWM will also yield more favorable results than the asymmetric TRM.

As a second limitation to our study, it should be noted that the "more is better" validation criterion allowed us to evaluate the validity of the prevalence estimates obtained only on a relative as opposed to an absolute level. Because the true prevalence of crossing the street on a 'Don't Walk' sign in plain sight of children remained unknown, we cannot know whether any of the prevalence estimates we obtained in our sample closely reflected the true value, or whether they were still under-, or even overestimates. Even more importantly, the "more is better"-approach also did not allow testing for the influence of false positives or false negatives, as this would have required knowing the status of individual respondents concerning the sensitive attribute. False positives and false negatives in the CWM are likely a consequence of some respondents not understanding the comparatively complex instructions and therefore choosing one of the answer options at random. Especially in cases in which the true prevalence is zero or very low, false positives have been shown to lead to somewhat inflated CWM prevalence estimates [34,35]. In the current study, we cannot rule out that some participants in the CWM conditions disregarded the CWM instructions and chose their answers at random, potentially leading to some false positives. However, a recent study has shown that if the true prevalence of the sensitive attribute is well above 0% (as is likely the case in the current study), the influence of false negatives on CWM estimates outweighs the influence of false positives [38]. Therefore, even if random responding affected CWM estimates in the current study, overall, it most likely deflated rather than inflated CWM estimates and can therefore not explain the difference between DQ and CWM estimates. Most importantly, random responding cannot account for the interaction between instruction and questioning technique, reflected in the reduction of the large effect of faking instructions on respondents in the DQ condition to a close-to-zero effect on respondents confronted with a CWM question. It would be necessary to assume that virtually all respondents in the CWM conditions completely disregarded the "honest" versus "fake good" instructions and the instructions on how to operate the CWM question to explain this interaction solely based on random responding. This explanation appears highly unlikely given the observed effect of faking instructions in the DQ conditions, the random allocation to experimental conditions, and previous empirical studies suggesting false positive rates in the CWM of 5% to 14% rather than 100% [34,35,38]. Nevertheless, to counter the apparent limitations associated with the "more is better"-criterion, future studies should investigate the effect of deliberate faking on questions on sensitive attributes for which the true prevalence and the status of individual respondents are known or can be determined directly from the sample [cf. 30,34,38]. Only such strong validation studies can provide conclusive evidence for the absolute validity of the prevalence estimates obtained, on the absolute influence of false positives and false negatives due to random responding, and consequently on the absolute robustness of indirect questioning techniques such as the CWM against deliberate faking attempts.

A third limitation of our study is due to the method we used to elicit "honest" versus "fake good" responses. In the present first investigation of the influence of positive self-presentation on estimates obtained via indirect questioning, we opted for an experimental manipulation of respondent honesty via "fake good" instructions. We preferred this approach over a non-

experimental comparison of real-world situations providing more serious and potentially differing incentives to fake, for example, in a study of job applicants. This decision was taken to maximize internal validity even though this choice was associated with limitations in terms of the ecological validity of our study. Our design allowed us to attribute observed group differences to our experimental manipulation of respondent honesty. In contrast, a non-experimental design would have been open to potential alternative explanations, including, for example, confounding variables or self-selection effects. Presumably, because of the high degree of experimental control they afford, similar experimental "fake good" manipulations have been - and still are - successfully applied to investigate the influence of positive self-presentation on various measures of self-report [49–55]. While the CWM has proven to be robust against an experimenter-instructed "faking good" in the current study, the generalizability of this finding should be tested further due to the peculiar nature of this manipulation. Future studies need to evaluate whether the robustness of the model is also maintained in real-world situations presenting high incentives towards deliberate, positive self-presentation.

A fourth limitation of the current study is the composition of our sample, which exclusively comprised people with a high level of education, that is, university students. As educational attainment and academic performance have repeatedly been shown to be positively, and strongly, associated with cognitive ability [73–75], the generalizability of our results is also potentially restricted to people with comparatively high cognitive abilities. In less educated samples, indirect questioning techniques have been found to be associated with lower acceptance rates [76], lower comprehensibility [39], and a higher share of participants disregarding the instructions [77]. Hence, future studies should include respondents' level of education as a quasi-experimental factor to investigate its effect on the ability to deliberately fake on direct versus indirect questions. In such a design, a recently proposed extension of the crosswise model, the ECWM, might additionally be employed to quantify the share of participants who do not follow the instructions [78,79].

Finally, we would like to encourage researchers to contrast indirect questioning techniques such as the CWM with alternative approaches to measuring and controlling the influence of deliberate faking on self-reports such as, for example, social desirability scales [3,80,81], behavioral indicators [82–84], lie detection and the bogus pipeline [85–87], and the overclaiming technique [88–90]. Such extended studies could help identify those methods–or possibly even a combination of methods–that optimally counteract the detrimental influence of deliberate positive self-presentation on the validity of self-reports.

Conclusion

In summary, we have demonstrated that the crosswise model is capable of controlling for the influence of deliberate faking in surveys on sensitive issues. This robustness of the CWM is presumably attributable to the higher level of perceived confidentiality it affords compared to conventional direct questions, and to respondents' inability to fake their answers in a self-protective manner in the CWM question format. We therefore recommend using the CWM in surveys on sensitive personal attitudes and behaviors to reduce bias due to self-presentational concerns and minimize the influence of deliberate faking.

Author Contributions

Conceptualization: Adrian Hoffmann, Julia Meisters, Jochen Musch.

Funding acquisition: Adrian Hoffmann, Jochen Musch.

Investigation: Adrian Hoffmann, Julia Meisters, Jochen Musch.

Methodology: Adrian Hoffmann, Julia Meisters, Jochen Musch.

Writing – original draft: Adrian Hoffmann, Julia Meisters, Jochen Musch.

Writing – review & editing: Adrian Hoffmann, Julia Meisters, Jochen Musch.

References

1. Gnambs T, Kaspar K. Disclosure of sensitive behaviors across self-administered survey modes: a meta-analysis. Behavior Research Methods. 2015; 47:1237–59. https://doi.org/10.3758/s13428-014-0533-4 WOS:000364511400025. PMID: 25410404
2. Krumpal I. Determinants of social desirability bias in sensitive surveys: a literature review. Qual Quant. 2013; 47:2025–47. https://doi.org/10.1007/s11135-011-9640-9 ISI:000316267500014.
3. Paulhus DL. Measurement and Control of Response Bias. In: Robinson JP, Shaver PR, Wrightsman LS, editors. Measures of personality and social psychological attitudes, Vol 1. San Diego, CA: Academic Press; 1991. p. 17–59.
4. Phillips DL, Clancy KJ. Some Effects of Social Desirability in Survey Studies. Am J Sociol. 1972; 77:921–40. https://doi.org/10.1086/225231 ISI:A1972M445200028.
5. Warner SL. Randomized-Response - a Survey Technique for Eliminating Evasive Answer Bias. J Am Stat Assoc. 1965; 60:63–9. ISI:A1965CKX1300005. PMID: 12261830
6. Franke AG, Bagusat C, Dietz P, Hoffmann I, Simon P, Ulrich R, et al. Use of illicit and prescription drugs for cognitive or mood enhancement among surgeons. Bmc Med. 2013; 11(102):1–9. https://doi.org/10.1186/1741-7015-11-102 PubMed ISI:235702562000001.
7. Ostapczuk M, Musch J, Moshagen M. Improving self-report measures of medication non-adherence using a cheating detection extension of the randomised-response-technique. Stat Methods Med Res. 2011; 20:489–503. https://doi.org/10.1177/0962280210372843 ISI:206392695700003.
8. Moshagen M, Musch J, Ostapczuk M, Zhao Z. Reducing Socially Desirable Responses in Epidemiologic Surveys. An Extension of the Randomized-response Technique. Epidemiology. 2010; 21:379–82. https://doi.org/10.1097/EDE.0b013e3181d61dbc ISI:203861721500015.
9. Hoffmann A, Musch J. Assessing the validity of two indirect questioning techniques: a Stochastic Lie Detector versus the Crosswise Model. Behavior Research Methods. 2016; 48:1032–46. https://doi.org/10.3758/s13428-015-0628-6 PMID: 26182857
10. Striegel H, Ulrich R, Simon P. Randomized response estimates for doping and illicit drug use in elite athletes. Drug Alcohol Depen. 2010; 106:230–2. https://doi.org/10.1016/j.drugalcdep.2009.07.026 ISI:197406120600019.
11. Moshagen M, Hilbig BE, Erdfelder E, Moritz A. An Experimental Validation Method for Questioning Techniques That Assess Sensitive Issues. Experimental Psychology. 2014; 61:48–54. https://doi.org/10.1027/1618-3169/a000226 PMID: 23948389
12. Lensvelt-Mulders GJLM, Hox JJ, van der Heijden PGM, Maas CJM. Meta-analysis of randomized response research: thirty-five years of validation. Sociol Method Res. 2005; 33:319–48. https://doi.org/10.1177/0049124104268664 ISI:000226871800001.
13. Holbrook AL, Krosnick JA. Measuring Voter Turnout by Using the Randomized Response Technique: Evidence Calling into Question the Method's Validity. Public Opinion Quarterly. 2010; 74:328–43. https://doi.org/10.1093/Poq/Nfq012 ISI:000278731600007.
14. Wolter F, Preisendörfer P. Asking Sensitive Questions: An Evaluation of the Randomized Response Technique Versus Direct Questioning Using Individual Validation Data. Sociol Method Res. 2013; 42:321–53. https://doi.org/10.1177/0049124113500474 ISI:000324743900004.
15. James RA, Nepusz T, Naughton DP, Petroczi A. A potential inflating effect in estimation models: Cautionary evidence from comparing performance enhancing drug and herbal hormonal supplement use estimates. Psychol Sport Exerc. 2013; 14:84–96. https://doi.org/10.1016/j.psychsport.2012.08.003 ISI:000312426000010.
16. Landsheer JA, van der Heijden PGM, van Gils G. Trust and understanding, two psychological aspects of randomized response - A study of a method for improving the estimate of social security fraud. Qual Quant. 1999; 33:1–12. https://doi.org/10.1023/A:1004361819974 ISI:000079006700001.
17. Coutts E, Jann B. Sensitive Questions in Online Surveys: Experimental Results for the Randomized Response Technique (RRT) and the Unmatched Count Technique (UCT). Sociol Method Res. 2011; 40:169–93. https://doi.org/10.1177/0049124110390768 ISI:000286103900008.
18. Tian G-L, Tang M-L. Incomplete Categorical Data Design: Non-Randomized Response Techniques for Sensitive Questions in Surveys. Boca Raton, FL: CRC Press, Taylor & Francis Group; 2014.

19. Yu J-W, Tian G-L, Tang M-L. Two new models for survey sampling with sensitive characteristic: design and analysis. Metrika. 2008; 67:251–63. https://doi.org/10.1007/s00184-007-0131-x ISI:000254204100001.

20. Waubert de Puiseau B, Hoffmann A, Musch J. How indirect questioning techniques may promote democracy: A pre-election polling experiment. Basic And Applied Social Psychology. 2017; 39:209–17. https://doi.org/10.1080/01973533.2017.1331351

21. Thielmann I, Heck DW, Hilbig BE. Anonymity and incentives: An investigation of techniques to reduce socially desirable responding in the Trust Game. Judgment and Decision Making. 2016; 11:527–36.

22. Kundt TC, Misch F, Nerré B. Re-assessing the merits of measuring tax evasion through business surveys: an application of the crosswise model. International Tax and Public Finance. 2017; 24:112–33. https://doi.org/10.1007/s10797-015-9373-0

23. Korndörfer M, Krumpal I, Schmukle SC. Measuring and explaining tax evasion: Improving self-reports using the crosswise model. Journal of Economic Psychology. 2014; 45:18–32. https://doi.org/10.1016/j.joep.2014.08.001

24. Nakhaee MR, Pakravan F, Nakhaee N. Prevalence of Use of Anabolic Steroids by Bodybuilders Using Three Methods in a City of Iran. Addiction & Health. 2013; 5:77–82.

25. Jann B, Jerke J, Krumpal I. Asking Sensitive Questions Using the Crosswise Model. Public Opinion Quarterly. 2012; 76:32–49. https://doi.org/10.1093/Poq/Nfr036 ISI:000301068300002.

26. Hoffmann A, Meisters J, Musch J. On the validity of Nonrandomized Response Techniques: An Experimental Comparison of the Crosswise Model and the Triangular Model. Behavior Research Methods. 2020; 52:1768–82. https://doi.org/10.3758/s13428-020-01349-9 PMID: 32185638

27. Mieth L, Mayer MM, Hoffmann A, Buchner A, Bell R. Do they really wash their hands? Prevalence estimates for personal hygiene behaviour during the COVID-19 pandemic based on indirect questions. BMC Public Health. 2021; 21:1–8. https://doi.org/10.1186/s12889-020-10013-y PMID: 33388037

28. Hoffmann A, Musch J. Prejudice against women leaders: Insights from an indirect questioning approach. Sex Roles. 2019; 80:681–92. https://doi.org/10.1007/s11199-018-0969-6

29. Sagoe D, Cruyff M, Spendiff O, Chegeni R, de Hon O, van der Heijden P, et al. Functionality of the Crosswise Model for Assessing Sensitive or Transgressive Behavior: A Systematic Review and Meta-Analysis. Frontiers in Psychology. 2021; 12. https://doi.org/10.3389/fpsyg.2021.655592 PMID: 34248750

30. Hoffmann A, Diedenhofen B, Verschuere BJ, Musch J. A strong validation of the Crosswise Model using experimentally induced cheating behavior. Experimental Psychology. 2015; 62:403–14. https://doi.org/10.1027/1618-3169/a000304 PMID: 27120562

31. Walzenbach S, Hinz T. Pouring water into wine: Revisiting the advantages of the crosswise model for asking sensitive questions. Survey Methods: Insights from the Field. 2019. https://doi.org/10.13094/SMIF-2019-00002

32. When and Why Randomized Response Techniques (Fail to) Elicit the Truth. Harvard Business School Working Paper, No 16-125 [Internet]. http://www.hbs.edu/faculty/Pages/item.aspx?num=51059. 2017 [cited Mar 3rd, 2017].

33. Götze A, Wahl A. Psychische Gesundheit: Eine heikle Thematik in der empirischen Umfrageforschung? Zur Validierung des Crosswise Modells [Mental health: a sensitive topic in empirical survey research? On the validation of the Crosswise Model]. SISS: Schriftenreihe des Instituts für Sozialwissenschaften der Universität Stuttgart. 2020; 49:1–36.

34. Höglinger M, Jann B. More is not always better: An experimental individual-level validation of the randomized response technique and the crosswise model. Plos One. 2018; 13:e0201770. https://doi.org/10.1371/journal.pone.0201770 PMID: 30106973

35. Höglinger M, Diekmann A. Uncovering a Blind Spot in Sensitive Question Research: False Positives Undermine the Crosswise-Model RRT. Political Analysis. 2017; 25:131–7. https://doi.org/10.1017/pan.2016.5

36. Jerke J, Johann D, Rauhut H, Thomas K. Too sophisticated even for highly educated survey respondents? A qualitative assessment of indirect question formats for sensitive questions. Survey Research Methods. 2019; 13(3):319–51. https://doi.org/10.18148/srm/2019.v13i3.7453 WOS:000502096700006.

37. Schnapp P. Sensitive Question Techniques and Careless Responding: Adjusting the Crosswise Model for Random Answers. methods, data, analyses. 2019; 13(2):307–20. https://doi.org/10.12758/mda.2019.03

38. Meisters J, Hoffmann A, Musch J. Can detailed instructions and comprehension checks increase the validity of crosswise model estimates? Plos One. 2020; 15(6):e0235403. https://doi.org/10.1371/journal.pone.0235403 PMID: 32603352

39. Hoffmann A, Waubert de Puiseau B, Schmidt AF, Musch J. On the comprehensibility and perceived privacy protection of indirect questioning techniques. Behavior Research Methods. 2017; 49:1470–83. https://doi.org/10.3758/s13428-016-0804-3 PMID: 27631988

40. Lanke J. Degree of Protection in Randomized Interviews. Int Stat Rev. 1976; 44:197–203. https://doi.org/10.2307/1403277 ISI:A1976CD40700002.

41. Fligner MA, Policello GE, Singh J. A Comparison of 2 Randomized Response Survey Methods with Consideration for Level of Respondent Protection. Commun Stat a-Theor. 1977; 6(15):1511–24. https://doi.org/10.1080/03610927708827593 ISI:A1977EP80200007.

42. Bayes T. An essay towards solving a problem in the doctrine of chances. Philosophical transactions of the Royal Society of London. 1763; 53:370–418.

43. Soeken KL, Macready GB. Respondents Perceived Protection When Using Randomized-Response. Psychol Bull. 1982; 92:487–9. https://doi.org/10.1037/0033-2909.92.2.487 ISI:A1982PK39700012.

44. Ostapczuk M, Moshagen M, Zhao Z, Musch J. Assessing sensitive attributes using the randomized response technique: Evidence for the importance of response symmetry. Journal of Educational and Behavioral Statistics. 2009; 34:267–87. https://doi.org/10.3102/1076998609332747 2010-07441-007.

45. Kroher M. Should I stay or should I go? Abweichendes Verhalten im Straßenverkehr [Should I stay or should I go? Deviant behavior at traffic lights]. Soziale Welt. 2014; 65:201–20. https://doi.org/10.5771/0038-6073-2014-2-201 WOS:000340216400005.

46. Knowles D, Aigner-Breuss E, Braun E, Donabauer M, Körmer C, Witzik A, et al. Role Model - Verkehrsunfallreduktion durch Vorbildwirkung und rücksichtsvolles Fahrverhalten [Role model - Reduction of road traffic accidents through role model function and responsible driving behavior] 2016 Report No. 051.

47. Muir C, Devlin A, Oxley J, Kopinathan C, Charlton J, Koppel S. Parents as role models in road safety. Monash University Accident Research Centre Report, 2010.

48. Wölfl J, Weber K. Unterwegs zur Schule - Das Verkehrsverhalten von Kindern und Erwachsenen im Schulumfeld [On the way to school - Road traffic behavior of children and adults in the school environment]. Wien2007.

49. Roma P, Mazza C, Mammarella S, Mantovani B, Mandarelli G, Ferracuti S. Faking-Good Behavior in Self-Favorable Scales of the MMPI-2 - A Study With Time Pressure. European Journal of Psychological Assessment. 2019; 36(2):250–8. https://doi.org/10.1027/1015-5759/a000511

50. Stoeber J, Hotham S. Perfectionism and social desirability: Students report increased perfectionism to create a positive impression. Pers Indiv Differ. 2013; 55(5):626–9. https://doi.org/10.1016/j.paid.2013.04.023 WOS:000322410500032.

51. Alliger GM, Dwight SA. A meta-analytic investigation of the susceptibility of integrity tests to faking and coaching. Educ Psychol Meas. 2000; 60(1):59–72. https://doi.org/10.1177/00131640021970367 WOS:000084961300004.

52. Zickar MJ, Robie C. Modeling faking good on personality items: An item-level analysis. J Appl Psychol. 1999; 84(4):551–63. https://doi.org/10.1037/0021-9010.84.4.551 WOS:000082737800006.

53. Cole DM, Rung JM, Madden GJ. Assessing susceptibility of a temporal discounting task to faking. J Clin Psychol. 2019; 75(10):1959–74. https://doi.org/10.1002/jclp.22831 WOS:312787513500001.

54. Sutton RM, Robinson B, Farrall SD. Gender, fear of crime, and self-presentation: an experimental investigation. Psychol Crime Law. 2011; 17(5):421–33. https://doi.org/10.1080/10683160903292261 WOS:000290409300003.

55. Eyssel F, Ribas X. How to be good (or bad): On the fakeability of dehumanization and prejudice against outgroups. Group Process Interg. 2012; 15(6):804–12. https://doi.org/10.1177/1368430212447204 WOS:000310329500008.

56. Umesh UN, Peterson RA. A Critical Evaluation of the Randomized-Response Method - Applications, Validation, and Research Agenda. Sociol Method Res. 1991; 20:104–38. https://doi.org/10.1177/0049124191020001004 ISI:A1991GA53200004.

57. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. JAMA. 2013; 310:2191–4. https://doi.org/10.1001/jama.2013.281053 PMID: 24141714.

58. Berufsethische Richtlinien des Berufsverbandes Deutscher Psychologinnen und Psychologen e.V. und der Deutschen Gesellschaft für Psychologie e.V. [Professional ethical guidelines of the German Association of Psychologists and the German Psychological Society] [Internet]. 2016 [cited Sep 10th, 2018]. Available from: https://www.dgps.de/fileadmin/documents/Empfehlungen/berufsethische_richtlinien_dgps.pdf.

59. RatSWD. Forschungsethische Grundsätze und Prüfverfahren in den Sozial- und Wirtschaftswissenschaften [Ethical research principles and test methods in the social and economic sciences]. Berlin: Rat für Sozial- und Wirtschaftsdaten (RatSWD); 2017.

60. DPGs. In welchen Fällen auf einen Ethikantrag verzichtet werden kann [In which cases an ethics application is not needed]. Ethisches Handeln in der psychologischen Forschung - Empfehlungen der Deutschen Gesellschaft für Psychologie für Forschende und Ethikkommissionen. Göttingen: Hogrefe; 2018.

61. Ulrich R, Schröter H, Striegel H, Simon P. Asking Sensitive Questions: A Statistical Power Analysis of Randomized Response Models. Psychol Methods. 2012; 17:623–41. https://doi.org/10.1037/a0029314 ISI:000312113600010. PMID: 22924599

62. Faul F, Erdfelder E, Lang A-G, Buchner A. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behavior Research Methods. 2007; 39:175–91. https://doi.org/10.3758/bf03193146 PMID: 17695343

63. Moshagen M, Musch J, Erdfelder E. A stochastic lie detector. Behavior Research Methods. 2012; 44:222–31. https://doi.org/10.3758/s13428-011-0144-2 21858604. 2012-04194-018. PMID: 21858604

64. Batchelder WH, Riefer DM. Theoretical and empirical review of multinomial process tree modeling. Psychon B Rev. 1999; 6:57–86. https://doi.org/10.3758/bf03210812 ISI:000079575700004. PMID: 12199315

65. Batchelder WH. Multinomial processing tree models and psychological assessment. Psychol Assessment. 1998; 10:331–44. https://doi.org/10.1037/1040-3590.10.4.331 ISI:000077959900003.

66. Moshagen M, Hilbig BE, Musch J. Defection in the dark? A randomized-response investigation of cooperativeness in social dilemma games. European Journal of Social Psychology. 2011; 41:638–44. https://doi.org/10.1002/Ejsp.793 ISI:000293687800012.

67. Geburten in Deutschland [Births in Germany] [Internet]. German Federal Statistical Office, . 2012 [cited Jun 6, 2012]. Available from: https://www.destatis.de/DE/Publikationen/Thematisch/Bevoelkerung/BevoelkerungsbewegungBroschuereGeburtenDeutschland0120007129004.pdf.

68. Dempster AP, Laird NM, Rubin DB. Maximum Likelihood from Incomplete Data Via Em Algorithm. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 1977; 39:1–38. ISI: A1977DM46400001.

69. Hu X, Batchelder WH. The Statistical-Analysis of General Processing Tree Models with the Em Algorithm. Psychometrika. 1994; 59:21–47. https://doi.org/10.1007/Bf02294263 ISI:A1994NA79500002.

70. Moshagen M. multiTree: A computer program for the analysis of multinomial processing tree models. Behavior Research Methods. 2010; 42:42–54. https://doi.org/10.3758/BRM.42.1.42 PMID: 20160285

71. Knapp BR, Batchelder WH. Representing parametric order constraints in multi-trial applications of multinomial processing tree models. J Math Psychol. 2004; 48:215–29. https://doi.org/10.1016/j.jmp.2004.03.002 ISI:000222831000001.

72. Birnbaum MH. How to show that 9 > 221: Collect judgments in a between-subjects design. Psychol Methods. 1999; 4:243–9. https://doi.org/10.1016/0006-2952(75)90080-5 ISI:000082696900001. PMID: 9

73. Weinberg D, Stevens GWJW, Finkenauer C, Brunekreef B, Smit HA, Wijga AH. The pathways from parental and neighbourhood socioeconomic status to adolescent educational attainment: An examination of the role of cognitive ability, teacher assessment, and educational expectations. Plos One. 2019; 14(5):e0216803. https://doi.org/10.1371/journal.pone.0216803 PMID: 31116770

74. Leeson P, Ciarrochi J, Heaven PCL. Cognitive ability, personality, and academic performance in adolescence. Pers Indiv Differ. 2008; 45(7):630–5. https://doi.org/10.1016/j.paid.2008.07.006 2008-13169-013.

75. Berry CM, Gruys ML, Sackett PR. Educational attainment as a proxy for cognitive ability in selection: Effects on levels of cognitive ability and adverse impact. J Appl Psychol. 2006; 91(3):696–705. https://doi.org/10.1037/0021-9010.91.3.696 16737365. 2006-07101-015. PMID: 16737365

76. Krumpal I. Estimating the prevalence of xenophobia and anti-Semitism in Germany: A comparison of randomized response and direct questioning. Soc Sci Res. 2012; 41:1387–403. https://doi.org/10.1016/j.ssresearch.2012.05.015 ISI:000310045600010. PMID: 23017963

77. Ostapczuk M, Musch J, Moshagen M. A randomized-response investigation of the education effect in attitudes towards foreigners. European Journal of Social Psychology. 2009; 39:920–31. https://doi.org/10.1002/ejsp.588 2009-17720-004.

78. Heck DW, Hoffmann A, Moshagen M. Detecting nonadherence without loss in efficiency: A simple extension of the crosswise model. Behavior Research Methods. 2018; 50:1895–905. https://doi.org/10.3758/s13428-017-0957-8 PMID: 28916924

79. Meisters J, Hoffmann A, Musch J. Controlling social desirability bias: An experimental investigation of the extended crosswise model. Plos One. 2020; 15(12):e0243384. https://doi.org/10.1371/journal.pone.0243384 PMID: 33284820

80. Ventimiglia M, MacDonald DA. An examination of the factorial dimensionality of the Marlowe Crowne Social Desirability Scale. Pers Indiv Differ. 2012; 52:487–91. https://doi.org/10.1016/j.paid.2011.11.016 ISI:000301218100004.

81. Stöber J, Dette DE, Musch J. Comparing continuous and dichotomous scoring of the balanced inventory of desirable responding. J Pers Assess. 2002; 78:370–89. https://doi.org/10.1207/S15327752JPA7802_10 ISI:120671995500010.

82. Monaro M, Mazza C, Colasanti M, Ferracuti S, Orrù G, di Domenico A, et al. Detecting faking-good response style in personality questionnaires with four choice alternatives. Psychological Research. 2021:Advance online publication. https://doi.org/10.1007/s00426-020-01473-3 PMID: 33452928

83. Mazza C, Monaro M, Burla F, Colasanti M, Orrù G, Ferracuti S, et al. Use of mouse-tracking software to detect faking-good behavior on personality questionnaires: an explorative study. Scientific Reports. 2020; 10(1):1–13. https://doi.org/10.1038/s41598-019-56847-4 PMID: 31913322

84. Mazza C, Monaro M, Orrù G, Burla F, Colasanti M, Ferracuti S, et al. Introducing Machine Learning to Detect Personality Faking-Good in a Male Sample: A New Model Based on Minnesota Multiphasic Personality Inventory-2 Restructured Form Scales and Reaction Times. Front Psychiatry. 2019; 10:389. https://doi.org/10.3389/fpsyt.2019.00389 PMID: 31275176; PubMed Central PMCID: PMC6593269.

85. Roese NJ, Jamieson DW. 20 Years of Bogus Pipeline Research - a Critical-Review and Metaanalysis. Psychol Bull. 1993; 114:363–75. https://doi.org/10.1037/0033-2909.114.2.363 ISI:A1993LX35200009.

86. Elliott E, Egan V, Grubin D. Not So Bogus Pipeline: A Study of the Bogus Pipeline Effect And Its Implications for Polygraph Testing. Polygraph. 2017; 46(1):1–9.

87. Jones EE, Sigall H. The bogus pipeline: A new paradigm for measuring affect and attitude. Psychol Bull. 1971; 76(5):349–64. https://doi.org/10.1037/h0031617

88. Goecke B, Weiss S, Steger D, Schroeders U, Wilhelm O. Testing competing claims about overclaiming. Intelligence. 2020; 81. https://doi.org/10.1016/j.intell.2020.101470 2020-56775-001.

89. Müller S, Moshagen M. Controlling for response bias in self-ratings of personality: A comparison of impression management scales and the overclaiming technique. Journal of Personality Assessment. 2019; 101(3):229–36. https://doi.org/10.1080/00223891.2018.1451870 2018-21301-001. PMID: 29648894

90. Paulhus DL, Harms PD. Measuring cognitive ability with the overclaiming technique. Intelligence. 2004; 32(3):297–314. https://doi.org/10.1016/j.intell.2004.02.001 2004-16343-006.