On the validity of non-randomized response techniques: an experimental comparison of the crosswise model and the triangular model

Adrian Hoffmann[1] · Julia Meisters[1] · Jochen Musch[1]

Abstract

Non-randomized response techniques (NRRTs) such as the crosswise model and the triangular model (CWM and TRM; Yu et al. *Metrika, 67*, 251–263, 2008) have been developed to control for socially desirable responding in surveys on sensitive personal attributes. We present the first study to directly compare the validity of the CWM and TRM and contrast their performance with a conventional direct questioning (DQ) approach. In a paper-pencil survey of 1382 students, we obtained prevalence estimates for two sensitive attributes (xenophobia and rejection of further refugee admissions) and one nonsensitive control attribute with a known prevalence (the first letter of respondents' surnames). Both NRRTs yielded descriptively higher prevalence estimates for the sensitive attributes than DQ; however, only the CWM estimates were significantly higher. We attribute the higher prevalence estimates for the CWM to its response symmetry, which is lacking in the TRM. Only the CWM provides symmetric answer options, meaning that there is no "safe" alternative respondents can choose to distance themselves from being carriers of the sensitive attribute. Prevalence estimates for the nonsensitive control attribute with known prevalence confirmed that neither method suffered from method-specific bias towards over- or underestimation. Exploratory moderator analyses further suggested that the sensitive attributes were perceived as more sensitive among politically left-oriented than among politically right-oriented respondents. Based on our results, we recommend using the CWM over the TRM in future studies on sensitive personal attributes.

Keywords Non-randomized response technique · Crosswise model · Triangular model · Validity · Xenophobia

Introduction

In surveys of sensitive attributes, social desirability bias threatens the validity of direct self-reports (Tourangeau & Yan, 2007). If some respondents choose to reply in line with social or legal norms rather than truthfully, prevalence estimates will be distorted in the direction of the socially desirable answer (Paulhus, 1991; Tourangeau & Yan, 2007). This response bias poses a serious threat to the interpretation of the results of survey studies on socially (un-)desirable or illegal

behaviors, such as drug use, prejudice, abortion, tax evasion, and plagiarism. Indirect questioning techniques try to control for socially desirable responding. In the present article, we focus on two non-randomized response techniques (NRRT) that have recently been proposed as an advancement upon traditional randomized response techniques (RRT; Warner, 1965). In particular, we investigate the crosswise model and the triangular model (CWM and TRM; Yu, Tian, & Tang, 2008) and present the first direct comparison of the two methods' validity.

Randomized response techniques (RRT) and non-randomized response techniques (NRRT)

Randomized response techniques (RRT; Warner, 1965) ensure that individual responses remain completely confidential in order to encourage respondents to provide more honest answers less distorted by social desirability bias. In the original RRT, respondents are presented with two statements: a sensitive statement (e.g., "I have taken cocaine") and its negation (e.g., "I have *never* taken cocaine"). Respondents are then

---

Adrian Hoffmann and Julia Meisters contributed equally to this work.

Electronic supplementary material The online version of this article (https://doi.org/10.3758/s13428-020-01349-9) contains supplementary material, which is available to authorized users.

✉ Adrian Hoffmann
adrian.hoffmann@uni-duesseldorf.de

[1] Department of Experimental Psychology, University of Duesseldorf, Universitaetsstrasse 1, Building 23.03, Floor 00, Room 27, 40225 Duesseldorf, Germany

asked to reply to only one of the statements with "true" or "false". The statement respondents are asked to react to is determined by the outcome of a randomization procedure (e.g., the roll of a die or the respondent's birth month). It is impossible for the experimenter to tell whether any individual respondent has admitted to consuming cocaine because the experimenter is not informed of the outcome of the randomization. The distribution of the potential randomization outcomes is known, however; therefore, prevalence estimates for the sensitive attribute can be obtained on the sample level via appropriate statistical procedures (Warner, 1965). In a meta-analysis of 32 "weak" validation studies, RRTs obtained higher prevalence estimates than direct questions (DQ) for various sensitive personal attributes (Lensvelt-Mulders, Hox, van der Heijden, & Maas, 2005). The superiority of RRTs over DQ was found to increase with the increasing sensitivity of the attribute under investigation.

According to the "more is better" criterion, higher prevalence estimates for sensitive attributes are assumed to be more valid because they are presumably less distorted by social desirability bias. However, finding such a pattern can only be considered "weak" validity evidence, since higher prevalence estimates may still underestimate—or overestimate—the true prevalence (Umesh & Peterson, 1991). Therefore, "strong" validation studies in which the prevalence of the sensitive attribute is known and can be used as an external validation criterion are considered the gold standard. However, such studies are costly and difficult to implement, making strong validation studies quite rare (Lensvelt-Mulders et al., 2005; Umesh & Peterson, 1991). A meta-analysis found only six strong validation studies comparing the prevalence estimates of RRTs to estimates obtained via DQ. Overall, RRTs were found to provide more valid estimates than DQ, but still substantially underestimated the true prevalence (Lensvelt-Mulders et al., 2005).

RRTs have also been criticized for their relative inefficiency, the complexity of their instructions, and the need to use an external randomization device (Landsheer, van der Heijden, & van Gils, 1999; Ulrich, Schröter, Striegel, & Simon, 2012; Yu et al., 2008). Consequently, non-randomized response techniques (NRRTs) such as the CWM and the TRM (Tian & Tang, 2014; Yu et al., 2008) have recently been proposed as an advancement upon RRTs. In contrast to RRTs, NRRTs directly integrate the randomization procedure into the answer options. They thus have simpler instructions and are easier for the experimenter to administer and for respondents to understand. Moreover, at least in contrast to the original RRT (Warner, 1965), the TRM is more efficient in most cases (Yu et al., 2008).

The crosswise model In the CWM, respondents are simultaneously presented with two statements. The first statement A refers to a sensitive attribute with unknown prevalence π (e.g., "I have taken cocaine"); the second statement B refers to a nonsensitive control attribute with known prevalence r that is

used for randomization (e.g., "My mother was born in November or December"). Respondents are requested to provide a joint answer to both statements by indicating whether "both statements are true or none of the statements is true" or "exactly one statement (irrespective of which one) is true". As in the original RRT, respondents can honestly choose either of the answers, while their individual status with respect to the sensitive statement A remains completely confidential. On the sample level, however, a maximum likelihood estimate for the prevalence π of the sensitive attribute can be obtained by using the formula (Yu et al., 2008):

$$\widehat{\pi}_{\text{CWM}} = \frac{\widehat{\lambda}_{\text{CWM}} + r - 1}{2 * r - 1} \tag{1}$$

where $\widehat{\lambda}_{\text{CWM}}$ is the observed proportion of respondents choosing the first answer option ("both statements are true or none of the statements is true"). An essential advantage of the CWM is its response symmetry: both answer options can and must be chosen by both carriers and noncarriers of the sensitive attribute, depending on their status with respect to the nonsensitive statement B (e.g., their mother's birth month).

To quantify the objective confidentiality protection of the two answer options in the CWM, conditional probabilities can be derived using Bayes' theorem (Bayes, 1763) according to the procedure described by Lanke (1976) and Fligner, Policello, and Singh (1977). The conditional probabilities of being identified as a carrier of the sensitive attribute given that one has selected the first ("both statements are true or none of the statements is true") versus the second answer option ["exactly one statement (irrespective of which one) is true"] are:

$$Pr_{\text{CWM}}(\text{carrier} \mid \text{"both/none true"})$$

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

$$Pr_{\text{CWM}}(\text{carrier} \mid \text{"one true"})$$

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

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

$$Pr_{\text{CWM}}(\text{carrier} \mid \text{"both/none true"}) = \frac{\widehat{\pi}_{\text{CWM}} * r}{\widehat{\lambda}_{\text{CWM}}} \tag{2.3}$$

$$Pr_{\text{CWM}}(\text{carrier} \mid \text{"one true"}) = \frac{\widehat{\pi}_{\text{CWM}} * (1 - r)}{\left(1 - \widehat{\lambda}_{\text{CWM}}\right)} \tag{2.4}$$

As can be seen when comparing the numerators of Eqs. 2.3 and 2.4, the probability of being identified as a carrier of the sensitive attribute is dependent on both the randomization

probability, r, and its complement, $1 - r$. As can also be seen from the equations, the conditional probability of being identified as a carrier is lower when choosing the first ("both/none true") compared with the second answer option ("one true") when the randomization probability is $0 < r < .5$, because $r < 1 - r$. For $.5 < r < 1$, this reverses to $r > 1 - r$; hence, in these cases, choosing the second answer option is associated with a lower risk. Importantly, however, the probability of being identified as a carrier of the sensitive attribute exceeds zero regardless of whether the respondent chooses the first or the second answer option for all cases of $0 < \widehat{\pi}_{CWM} < 1$, $0 < r < 1$, and $0 < \widehat{\lambda}_{CWM} < 1$. In practical applications of the CWM, these conditions are usually met 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%. In such cases, no CWM answer option provides a "safe" alternative respondents can choose to explicitly deny being a carrier of the sensitive attribute.

Even though a "safe" answer option is unavailable, respondents confronted with a CWM question might still be tempted to try to assess the relative risk of either answer option. To succeed, they would however have to (i) correctly estimate the randomization probability (r), and (ii) derive and understand the relationship between the randomization probability and the conditional probabilities from Eqs. 2.3 and 2.4. Soeken and Macready (1982) have already demonstrated that with the exception of extreme randomization probabilities, which eliminate confidentiality, respondents are rather poor at estimating the relationship between the randomization probability and the objective privacy protection afforded by the RRT. In light of this finding, as well as the time-consuming computations that would be necessary, we argue that it is quite unlikely that respondents will successfully assess the relative risk of the answer options. Instead, considering that response symmetry reduces the incentive to provide untruthful answers (Ostapczuk, Moshagen, Zhao, & Musch, 2009a), we propose that the high symmetry of the two CWM response options will lead to a higher proportion of honest responses compared with a direct question affording no confidentiality and offering a safe answer option that eliminates all risk of being associated with an undesirable behavior.

The validity of the CWM is supported by several weak validation studies in which higher, and therefore presumably more valid, prevalence estimates for sensitive attributes were obtained when using the CWM rather than DQ. The attributes investigated in these validation studies included crossing the street on a "No Walk" signal in plain view of children (Hoffmann, Meisters, & Musch, 2019), xenophobia (Hoffmann & Musch, 2016), the use of anabolic steroids among bodybuilders (Nakhaee, Pakravan, & Nakhaee, 2013), distrust in the Trust Game (Thielmann, Heck, & Hilbig, 2016), plagiarism (Jann, Jerke, & Krumpal, 2012),

tax evasion (Korndörfer, Krumpal, & Schmukle, 2014; Kundt, Misch, & Nerré, 2017), prejudice against female leaders (Hoffmann & Musch, 2019), and the intention to vote for the German right-wing party *Alternative for Germany* (Waubert de Puiseau, Hoffmann, & Musch, 2017). Additionally, in a first strong validation study, the CWM provided highly accurate prevalence estimates for the known prevalence of an experimentally induced sensitive attribute, while DQ provided a severe underestimation (Hoffmann, Diedenhofen, Verschuere, & Musch, 2015). The CWM has also been proven to be more comprehensible than other indirect questioning techniques, and to evoke a higher level of trust than conventional direct questions (Hoffmann, Waubert de Puiseau, Schmidt, & Musch, 2017). A recent study provided evidence that the CWM is quite robust even against deliberate faking, as "fake good" instructions impaired the validity of DQ but not of CWM prevalence estimates (Hoffmann et al., 2019). This robustness against deliberate faking is likely attributable to respondents' inability to identify a "safe" answer in the symmetric CWM.

Some recent studies have reported a problematic tendency for the CWM to produce false positives, that is, some noncarriers of the sensitive attribute were falsely classified as carriers. This can potentially result in an overestimation of the prevalence of sensitive attributes (Höglinger & Diekmann, 2017; Höglinger & Jann, 2018). In Höglinger and Diekmann (2017), the prevalence of two sensitive attributes with a known prevalence near zero was overestimated by the CWM (at 5% and 8%, respectively). Höglinger and Jann (2018) also observed false positives in a CWM survey. However, Meisters, Hoffmann, and Musch (2019) found no evidence for an overestimation of the known prevalence of an experimentally induced sensitive attribute, and suggested that the problem of false positive can be addressed by providing respondents with more comprehensive and detailed instructions and by ensuring that they actually comprehend the procedure.

The triangular model In the TRM, respondents are also presented with a sensitive statement A with unknown prevalence π (e.g., "I have taken cocaine") and a nonsensitive statement B with known prevalence r (e.g., "My mother was born in November or December") to which they must provide a joint response. The response options in the TRM are: "none of the statements is true" versus "at least one of the statements (irrespective of which one) is true". A maximum likelihood estimate for the prevalence π is given by (Yu et al., 2008):

$$\widehat{\pi}_{TRM} = 1 - \frac{\widehat{\lambda}_{TRM}}{1-r} \qquad (3)$$

where $\widehat{\lambda}_{TRM}$ is the observed proportion of respondents choosing the first answer option ("none of the statements is true"). As in the CWM, carriers of the sensitive attribute can choose

the second answer option ["at least one of the statements (irrespective of which one) is true"] without disclosing their true status with respect to the sensitive statement, since this response must also be given by noncarriers of the sensitive attribute who carry the nonsensitive attribute used for randomization (e.g., respondents whose mother was born in November or December). However, in contrast to the CWM, the TRM is an asymmetric model, because the answer option "none of the statements is true" provides a "safe" answer alternative that explicitly precludes being a carrier of the sensitive attribute. Respondents who are eager to distance themselves from the sensitive attribute may therefore likely be tempted to opt for this safe response option even when told otherwise by the randomization instructions.

The asymmetry of the TRM is reflected in the conditional probabilities of being identified as a carrier of the sensitive attribute given the first ("none of the statements is true") versus the second answer option ["at least one of the statements (irrespective of which one) is true"]:

$$Pr_{TRM}(\text{carrier} \mid \text{"none true"})$$

$$= \frac{Pr_{TRM}(\text{carrier} \cap \text{"none true"})}{Pr_{TRM}(\text{"none true"})} \quad (4.1)$$

$$Pr_{TRM}(\text{carrier} \mid \text{"at least one true"})$$

$$= \frac{Pr_{TRM}(\text{carrier} \cap \text{"at least one true"})}{Pr_{TRM}(\text{"at least one true"})} \quad (4.2)$$

The reformulation of Eq. 4.2 using the parameters for prevalence estimation from Eq. 3 is straightforward because in the TRM, the numerator of Eq. 4.1 refers to an impossible event. As per the TRM instructions, no carrier of the sensitive attribute may choose the first answer option ("none of the statements is true"), since for carriers, the sensitive statement A is true by definition. Therefore, according to the TRM, the probability of being a carrier when answering "none of the statements is true" is 0:

$$Pr_{TRM}(\text{carrier} \mid \text{"none true"}) = \frac{0}{\widehat{\lambda}_{TRM}} = 0 \quad (4.3)$$

$$Pr_{TRM}(\text{carrier} \mid \text{"at least one true"}) = \frac{\widehat{\pi}_{TRM}}{\left(1 - \widehat{\lambda}_{TRM}\right)} \quad (4.4)$$

As Eq. 4.3 shows, respondents can be sure that the first answer option ("none of the statements is true") is associated with a zero probability of being identified as a carrier of the sensitive attribute, irrespective of the randomization probability. Choosing this "safe" answer option is likely to attract respondents who are keen to make a positive or avoid a negative impression; this in turn is likely to result in underestimates due to dishonest responses (cf. Jerke & Krumpal, 2013).

Research on the validity of the TRM is relatively scarce. Two experimental validation studies have compared a TRM and a DQ control condition (Erdmann, 2019; Jerke & Krumpal, 2013). One study found prevalence estimates for plagiarism obtained via the TRM to descriptively exceed those obtained via DQ. However, this difference was not statistically significant (Jerke & Krumpal, 2013). In a second study, TRM estimates were comparable to and not significantly different from DQ estimates for three different sensitive attributes (Erdmann, 2019). Self-protective answering behavior facilitated by the asymmetric nature of the TRM is a possible explanation for these findings.

Comparison of the CWM and the TRM In terms of theoretical properties, a potential advantage of the CWM over the TRM is that only the CWM offers response symmetry. If respondents confronted with CWM questions are therefore less tempted to provide evasive answers, or are less successful in identifying a self-protective choice, this should result in prevalence estimates with higher validity. On the other hand, the TRM is usually more efficient than the CWM (with some exceptions for high values of r; cf. Theorem 3 in Yu et al., 2008). Accordingly, the TRM would be preferable to the CWM if both models were equally valid. However, we argue that the validity of the prevalence estimates is even more important than the efficiency of parameter estimation.

The only existing evidence regarding this question is based on a comparison across studies. The prevalence of plagiarism has been assessed with both the TRM (Jerke & Krumpal, 2013) and the CWM (Jann et al., 2012). Comparing the two results reveals that the CWM estimate significantly exceeded the DQ estimate, whereas the TRM and DQ estimates were not significantly different. Moreover, the CWM estimate (Jann et al., 2012) was descriptively higher than the TRM estimate (Jerke & Krumpal, 2013). This pattern of results tentatively suggests that the symmetric CWM might be superior to the asymmetric TRM in discouraging dishonest responses and thus in obtaining more valid estimates (cf. Jann et al., 2012; Jerke & Krumpal, 2013). However, a more conclusive comparison would involve directly comparing the CWM and the TRM in a single sample using an experimental design, and thus allowing alternative explanations for the observed differences in prevalence estimates to be ruled out. Taking up this challenge, the current study extends the existing body of research by providing the first experimental comparison of the validity of the CWM and the TRM and contrasting the performance of the two models to a DQ control condition. Xenophobia and opposition to further refugee admissions were chosen as sensitive attributes for the purpose of this validation study.

Xenophobia and opposition to reception of refugees in Germany

Xenophobia, a fear of—or negative attitude towards—foreigners, is quite prevalent in Germany (Heitmeyer, 2012; Krumpal, 2012; Wagner & van Dick, 2001). Since the "refugee crisis" of 2015, attitudes towards foreigners in general and refugees in particular have become more negative among the German population (Bertelsmann Stiftung, 2017). A representative survey revealed that 54% of the German population opposes the further intake of refugees, whereas most Germans still perceive a "welcoming culture" in Germany (Bertelsmann Stiftung, 2017). This discrepancy is likely to lead to social pressure to deny xenophobic attitudes and endorse further refugee admissions (Zick, Hövermann, & Krause, 2012). Direct self-reports on xenophobic attitudes and reluctance to grant asylum to refugees have indeed been found to be distorted by social desirability bias, leading to underestimates of their prevalence (D'Ancona, 2013; Krumpal, 2012; Moshagen & Musch, 2012; Ostapczuk, Musch, & Moshagen, 2009b). Indirect questioning techniques have been shown to lead to higher, and thus presumably more valid, estimates of both the prevalence of xenophobic attitudes (Hoffmann & Musch, 2016; Krumpal, 2012; Ostapczuk, Musch, et al., 2009b) and opposition to further refugee admissions (Moshagen & Musch, 2012).

We expected more xenophobic attitudes and greater opposition to further refugee admissions among respondents with a right-oriented versus a left-oriented political orientation, as a right-oriented political orientation has been shown to be positively associated with xenophobic attitudes (cf. Alba & Johnson, 2000; Zick, Küpper, & Hövermann, 2011). However, we also expected that the perceived sensitivity of these attitudes might vary as a function of political orientation. Right-oriented respondents may be more willing to openly express their disapproval of foreigners and refugees, whereas left-oriented respondents might feel hesitant to openly admit to a xenophobic attitude, and therefore choose to respond in a socially desirable rather than truthful manner. Therefore, the differences in prevalence estimates obtained via direct versus indirect questioning were used not only to assess the validity of the competing non-randomized response techniques, but also to investigate the influence of political orientation on question sensitivity.

The current study

This study is the first to experimentally compare the validity of two NRRTs (symmetric CWM and asymmetric TRM) and contrast their performance with that of conventional direct questioning (DQ). We expected that both NRRTs would outperform DQ in terms of delivering more valid prevalence estimates for two socially undesirable attributes. We also expected a beneficial effect of response symmetry (cf. Ostapczuk, Moshagen, et al., 2009a), and therefore predicted that the symmetric CWM would outperform the asymmetric TRM with respect to successful control of social desirability bias. Furthermore, we investigated a potential moderating influence of self-ascribed political orientation (from "left" to "right") on prevalence estimates for xenophobia and opposition to the further intake of refugees. Finally, a nonsensitive control attribute with known prevalence was included to test for method-specific biases in the form of a general tendency towards over- or underestimation. If the CWM and the TRM allow us to obtain prevalence estimates for the control attribute that correspond to DQ estimates and to the known prevalence, this provides strong evidence for the validity of these indirect questioning techniques.

Method

Sample

The initial sample consisted of 1544 students from the University of Düsseldorf recruited using a non-probability sampling plan (a convenience sample consisting of the attendants of large introductory lectures from various faculties and courses of study). Due to nonresponse to at least one of the experimental, demographic, or political orientation questions, 162 respondents (10.49%) were excluded from further analyses. Dropout rates were unaffected by experimental condition [condition 1: 8.27%, condition 2: 12.65%, condition 3: 9.77%, condition 4: 12.17%, condition 5: 10.00%, condition 6: 10.08%; $X^2(5) = 3.64$, $p = .603$, Cramer-V = .05], age [final sample: $M = 21.40$, dropouts: $M = 21.78$; $t(1515) = 0.71$, $p = .478$, $d = 0.06$], and political orientation [final sample: $M = -0.88$, dropouts: $M = -0.95$; $t(1445) = -0.32$, $p = .748$, $d = 0.04$]. Therefore, neither the assignment to a specific experimental condition nor the age or political orientation influenced whether respondents provided complete data and were thus included in the final sample. The final sample consisted of 1382 respondents (60.1% female) with a mean age of 21.40 years ($SD = 5.66$). Respondents were contacted on the university campus prior to the start of lectures and asked to complete a short one-page survey. This survey study was carried out in accordance with the revised Declaration of Helsinki (World Medical Association, 2013) and the ethical guidelines of the German Society for Psychology (Berufsverband Deutscher Psychologinnen und Psychologen & Deutsche Gesellschaft für Psychologie, 2016). All respondents were informed of the purpose of the study and the strict anonymization of all data prior to participation, and consented to participate on a voluntary basis without receiving any financial compensation.

Survey design

The one-page paper-pencil questionnaire contained three experimental questions, one question concerning respondents' self-reported political orientation, and demographic questions asking for the respondents' age and gender. The first two experimental questions asked about the two sensitive attributes (xenophobia and opposition to further refugee admissions). The third question referred to a nonsensitive control attribute and asked about the first letter of the respondents' surname. The prevalence of this attribute is known to be $\pi_{control} = 22\%$ in Germany according to official statistics (Reinders, 1996), and was also confirmed by the university's student registry. All respondents were presented with all three experimental questions. The order of the sensitive attributes the experimental questions referred to was fixed (question 1: xenophobia, question 2: opposition to refugee admission, question 3: first letter of surname K, L, M). The order of the questioning techniques (CWM, TRM, DQ) that were assigned to each experimental question was randomized. This resulted in six different experimental conditions to which respondents were randomly assigned. An overview of questions, questioning techniques, and number of respondents by experimental condition is given in Table 1. This design allowed us to manipulate and analyze the questioning technique as a between-subjects variable for each sensitive attribute. After data collection, responses were pooled across experimental conditions to obtain answer frequencies for all three questioning techniques and for each experimental question. Question 1 (xenophobia) was answered in CWM format by 233 + 221 = 454 respondents (32.85%), in TRM format by 231 + 231 = 462 respondents (33.43%), and in DQ format by 234 + 232 = 466 respondents (33.72%). Question 2 (opposition to refugee admission) was answered in CWM format by 231 + 234 = 465 respondents (33.65%), in TRM format by 233 + 232 = 465 respondents (33.65%), and in DQ format by 221 + 231 = 452 respondents (32.71%). Question 3 (first letter of surname: K, L, M) was answered in CWM format by 231 + 232 = 463 respondents (33.50%), in TRM format by 221 + 234 = 455 respondents (32.92%), and in DQ format by 233 + 231 = 464 respondents (33.57%).

Sensitive statements The sensitive statement used to measure xenophobia read as follows: "I would mind if my 20-year-old daughter had a relationship with a Turkish man." It was adapted from Bogardus' social distance scale (Bogardus, 1933) and had previously been used by Hoffmann and Musch (2016) in this form and by Jimenez (1999); Ostapczuk, Musch, et al. (2009b); and Silbermann and Hüsers (1995) with respect to other minority groups. The sensitive statement regarding opposition to further intake of refugees read as follows: "Germany has already received more than enough refugees."

CWM format In the CWM format, two statements were presented simultaneously: one of the two sensitive statements and a nonsensitive control statement with known prevalence r (father's month of birth: $r = .158$ based on official birth statistics; Pötzsch, 2012).

> Statement A: "I would mind if my 20-year-old daughter had a relationship with a Turkish man."
> Statement B: "My father was born in November or December."

Respondents could choose from the two answer options "both statements are true or both statements are false" and "exactly one statement is true (irrespective of which one)". The statements regarding the other two topics were adapted accordingly.

TRM format In the TRM format, two statements were presented simultaneously: the remaining sensitive statement and a nonsensitive control statement with known prevalence r (mother's month of birth: $r = .158$ based on official birth statistics; Pötzsch, 2012)

> Statement A: "Germany has already received more than enough refugees."
> Statement B: "My mother was born in November or December."

Respondents could choose from the two answer options "both statements are false" and "at least one statement is true

Table 1 Questions, questioning techniques, and number of respondents by experimental condition

	Experimental condition					
	1	2	3	4	5	6
Question 1 (xenophobia)	CWM	CWM	TRM	TRM	DQ	DQ
Question 2 (opposition to refugee admission)	TRM	DQ	CWM	DQ	CWM	TRM
Question 3 (first letter of surname: K, L, M)	DQ	TRM	DQ	CWM	TRM	CWM
Number of respondents	233 (16.86%)	221 (15.99%)	231 (16.71%)	231 (16.71%)	234 (16.93%)	232 (16.79%)

Note: CWM = crosswise model; TRM = triangular model; DQ = direct questioning

(irrespective of which one)". The statements regarding the other two topics were adapted accordingly.

DQ format In the DQ format, the nonsensitive control statement with known prevalence ($\pi = 22\%$, Reinders, 1996) read as follows: "My surname begins with one of the following letters: K, L, M." Respondents then had to indicate whether the statement was "true" or "false". The statements regarding the two sensitive attributes were presented in the same way.

Political orientation To assess respondents' political orientation, we presented the question: "Political beliefs are often labeled as rather 'left' or rather 'right'. Where on that scale would you place yourself?" Responses were recorded on an 11-point Likert scale from "left" (−5) to "right" (+5).

Statistical analyses

To obtain and compare prevalence estimates for the three attributes under investigation, we formulated multinomial processing tree models for all three questioning techniques (Batchelder, 1998; Batchelder & Riefer, 1999), following the procedure detailed in works such as Moshagen, Hilbig, and Musch (2011); Moshagen, Musch, and Erdfelder (2012); and Ostapczuk, Musch, and Moshagen (2011). The parameter π referred to the prevalence of the attribute to be estimated. In the CWM and the TRM question formats, the parameter r referred to the known prevalence of the nonsensitive attributes used for randomization, that is, the respondents' mother or father being born in November or December. Official statistics on more than 2.3 million births in Germany provided by the Federal Statistical Office show that of all births over the course of a year, about 15.8% take place in November or December (Pötzsch, 2012, p. 17). This number was therefore considered to be the best estimate of the prevalence of the nonsensitive attributes. Consequently, the parameter r was set constant to .158 in both the CWM and TRM formats. Processing tree diagrams for the CWM, TRM, and DQ formats are shown in Fig. 1.

Based on the empirically observed answer frequencies, prevalence estimates were obtained using the expectation–maximization algorithm (Dempster, Laird, & Rubin, 1977; Hu & Batchelder, 1994) as implemented in the multiTree software (Moshagen, 2010). Model fit was assessed via the asymptotically X^2-distributed log-likelihood statistic G^2 as detailed in, for example, Ostapczuk, Musch, et al. (2009b), Moshagen et al. (2011), and Hoffmann and Musch (2016). The multinomial processing tree models for all three questioning techniques per attribute were saturated with $df = 0$ and $G^2 = 0$, since the number of independent answer categories was just sufficient to estimate all parameters in the three questioning technique conditions. For each of the attributes under investigation (xenophobia, opposition to further refugee

admission, first letter of surname: K, L, M), three prevalence estimates were obtained (CWM, TRM, and DQ) based on the response frequencies from three independent, nonoverlapping groups (cf. Table 1). For example, one independent proportion of respondents answering "both/none true" to the xenophobia question in CWM format, another independent proportion of respondents answering "none true" to the xenophobia question in TRM format, and a third independent proportion of respondents answering "true" to the xenophobia question in DQ format allowed us to obtain completely independent estimates for the prevalence of xenophobia for each of these questioning technique formats (π_{CWM}, π_{TRM}, and π_{DQ}). To compare these prevalence estimates, we assessed the difference in model fit (ΔG^2) between an unrestricted baseline model and a restricted alternative model in which the respective parameters were equalized or set to be constant (e.g., $\pi_{CWM} = \pi_{DQ}$ or $\pi_{CWM} = .22$).

To analyze the influence of political orientation, we split the sample into two independent, nonoverlapping groups of left- versus right-oriented respondents via their answers to the Likert-scaled item on self-ascribed political orientation (from "left" = −5 to "right" = +5). For each of these groups, we established separate multinomial processing trees to obtain and compare prevalence estimates for the two sensitive attributes (xenophobia and opposition to a further admission of refugees) following the procedure detailed above. Both political orientation (left, right) and questioning technique format (CWM, TRM, DQ) varied between subjects. This allowed us to conduct pairwise comparisons of prevalence estimates between political orientation groups within a specific questioning technique format (e.g., $\pi_{CWM, left}$ versus $\pi_{CWM, right}$), and between questioning technique formats within a specific political orientation group (e.g., $\pi_{CWM, left}$ versus $\pi_{TRM, left}$) by assessing the difference in model fit (ΔG^2) between an unrestricted baseline model and a restricted alternative model in which the respective parameters were equalized or set to be constant (e.g., $\pi_{CWM, left} = \pi_{CWM, right}$ or $\pi_{CWM, left} = \pi_{TRM, left}$).

To assess interactions between questioning technique and political orientation on the prevalence estimates for the two sensitive attributes, we introduced parametric order constraints by re-parameterizing the original multinomial models established for estimating the prevalence of xenophobia and the prevalence of opposition to further refugee admission, respectively (Hoffmann & Musch, 2019; Knapp & Batchelder, 2004). Within each of these models, we replaced the parameter used for estimating the prevalence among left-oriented respondents in the DQ condition ($\pi_{DQ, left}$) with the corresponding parameter for right-oriented respondents ($\pi_{DQ, right}$), multiplied by a shrinkage factor ($\alpha_{DQ, left:right}$); the CWM and TRM conditions were re-parameterized in an analogous way ($\pi_{CWM, left} = \pi_{CWM, right} * \alpha_{CWM, left:right}$; $\pi_{TRM, left} = \pi_{TRM, right} * \alpha_{TRM, left:right}$). The shrinkage factors $\alpha_{DQ,}$
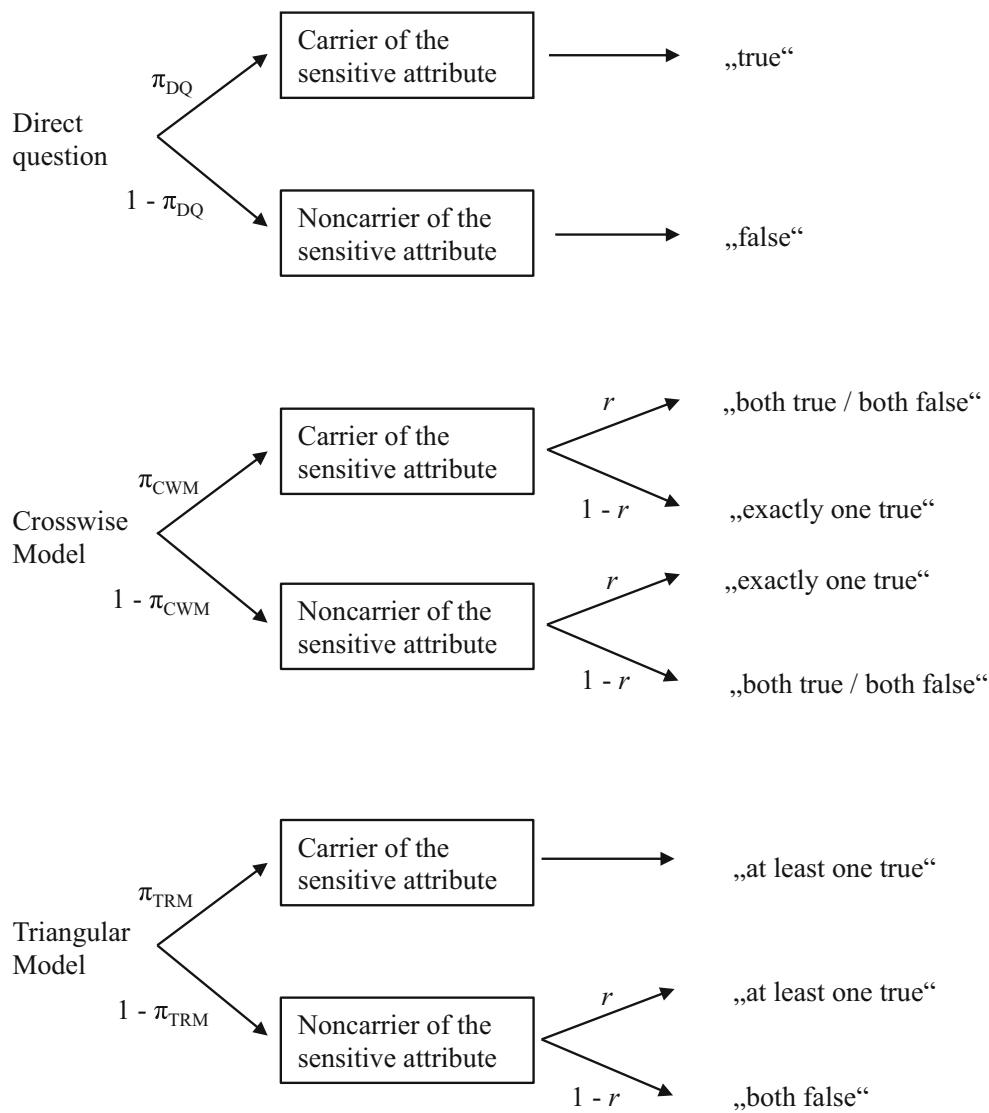
Fig. 1 Tree diagram for direct questions and questions posed according to the crosswise model and the triangular model. The parameter π represents the unknown prevalence of the sensitive attribute, and the parameter r represents the known randomization probability

left:right, $\alpha_{\text{CWM, left:right}}$, and $\alpha_{\text{TRM, left:right}}$ thus represent the ratio of the prevalence for politically left-oriented to the prevalence for politically right-oriented respondents in the DQ, CWM, and TRM conditions, respectively. For example, for the question on xenophobia, a shrinkage factor of $\alpha_{\text{DQ, left:right}} = 33\%$ means that in the DQ condition, respondents who label themselves as politically "left" are only .33 times as likely to admit to the xenophobic attitude as respondents who label themselves as politically "right". To test for interactions between questioning technique and political orientation, the shrinkage factors were compared using the ΔG^2 statistic, as described above (e.g., $\alpha_{\text{DQ, left:right}} = \alpha_{\text{CWM, left:right}}$). In this analysis, significant changes in model fit indicate significant interactions (Hoffmann & Musch, 2019). MultiTree model equations and empirically observed answer frequencies are available in appendices A and B in the electronic supplementary material.

Results

Table 2 contains the prevalence estimates and test statistics for parameter comparisons for both the two sensitive attributes and the nonsensitive control attribute obtained via the different questioning techniques.

Xenophobia (Sensitive Attribute 1)

Prevalence estimates for xenophobic responses were significantly higher when assessed via the CWM (31.65%) than when assessed via the TRM (20.05%) or DQ (15.45%). The TRM resulted in descriptively but not significantly higher prevalence estimates for xenophobia than DQ. This finding suggests that the prevalence of xenophobia is presumably underestimated in DQ and TRM, while the CWM seems to successfully control for socially desirable responding.

Table 2 Parameter estimates (standard errors in parentheses) and parameter comparisons for the sensitive and nonsensitive attributes by questioning technique

Parameter estimates	CWM	TRM	DQ
Xenophobia	31.65 (3.32)	20.05 (2.59)	15.45 (1.67)
Opposition to refugee admissions	43.56 (3.38)	37.43 (2.75)	36.73 (2.27)
First letter of surname: K, L, M	23.32 (3.16)	22.22 (2.65)	24.35 (1.99)

Parameter comparisons	$\lvert difference \rvert$	Model fit	
		$\Delta G^2\ (df = 1)$	p
Xenophobia			
$\hat{\pi}_{CWM}$ = $\hat{\pi}_{TRM}$	11.60	7.65	.006*
$\hat{\pi}_{CWM}$ = $\hat{\pi}_{DQ}$	16.20	19.61	< .001*
$\hat{\pi}_{TRM}$ = $\hat{\pi}_{DQ}$	4.60	2.24	.135
Opposition to refugee admissions			
$\hat{\pi}_{CWM}$ = $\hat{\pi}_{TRM}$	6.13	1.99	.159
$\hat{\pi}_{CWM}$ = $\hat{\pi}_{DQ}$	6.83	2.82	.093
$\hat{\pi}_{TRM}$ = $\hat{\pi}_{DQ}$	0.70	0.04	.844
First letter of surname: K, L, M			
$\hat{\pi}_{CWM}$ = $\hat{\pi}_{TRM}$	1.10	0.07	.789
$\hat{\pi}_{CWM}$ = $\hat{\pi}_{DQ}$	1.03	0.08	.782
$\hat{\pi}_{TRM}$ = $\hat{\pi}_{DQ}$	2.13	0.42	.519
$\hat{\pi}_{CWM}$ = π (22%)	1.32	0.18	.675
$\hat{\pi}_{TRM}$ = π (22%)	0.22	0.01	.935
$\hat{\pi}_{DQ}$ = π (22%)	2.35	1.46	.227

*Significant at $p < .05$

Opposition to refugee intake (Sensitive Attribute 2)

The prevalence estimates for opposition to further intake of refugees obtained via the CWM (43.56%) were descriptively but not significantly higher than those obtained via the TRM (37.43%) or DQ (36.73%). Prevalence estimates obtained via TRM and DQ also did not differ significantly.

Nonsensitive control attribute with known prevalence: First letter of surname

For the first letter of the respondents' surnames as a nonsensitive control attribute, all questioning techniques obtained similar prevalence estimates that did not differ from the known prevalence of 22% (CWM: 23.32%, TRM: 22.22%, DQ: 24.35%). These highly accurate prevalence estimates suggest that none of the questioning techniques under

investigation was subject to systematic bias in the form of a general tendency towards over- or underestimation.

## Exploratory moderator analyses

A median split of the sample by self-reported political orientation revealed a moderating influence on prevalence estimates for xenophobia and opposition to further refugee admissions (see Table 3).

As expected, the prevalence estimates for both sensitive attributes were higher among politically right-oriented than among politically left-oriented respondents for all questioning techniques. For xenophobia, the prevalence estimates obtained via TRM and DQ were significantly higher for right-oriented than for left-oriented respondents; in the CWM, however, prevalence estimates varied only slightly as a function of political orientation. For opposition to refugee intake, all three

Table 3 Parameter estimates (standard errors in parentheses) and parameter comparisons for the sensitive attributes by questioning technique and political orientation

Parameter estimates	CWM	TRM	DQ
Xenophobia			
$\hat{\pi}$ left political orientation	28.67 (4.84)	13.67 (3.51)	7.37 (1.77)
$\hat{\pi}$ right political orientation	34.19 (4.56)	26.21 (3.76)	22.49 (2.65)
$\hat{\alpha}$ shrinkage left : right	83.85 (18.03)	52.16 (15.34)	32.79 (8.78)
Opposition to refugee admissions			
$\hat{\pi}$ left political orientation	35.79 (4.88)	25.28 (3.93)	22.32 (2.78)
$\hat{\pi}$ right political orientation	50.29 (4.63)	47.69 (3.71)	50.88 (3.31)
$\hat{\alpha}$ shrinkage left : right	71.16 (11.71)	53.02 (9.22)	43.87 (6.17)

Parameter comparisons			|*difference*|	Model fit	
				ΔG^2 ($df = 1$)	p
Xenophobia					
	$\hat{\pi}_{CWM}$	$= \hat{\pi}_{TRM}$	15.00	6.41	.011*
"left"	$\hat{\pi}_{CWM}$	$= \hat{\pi}_{DQ}$	21.30	18.62	< .001*
	$\hat{\pi}_{TRM}$	$= \hat{\pi}_{DQ}$	6.30	2.63	.105
	$\hat{\pi}_{CWM}$	$= \hat{\pi}_{TRM}$	7.98	1.83	.176
"right"	$\hat{\pi}_{CWM}$	$= \hat{\pi}_{DQ}$	11.70	5.00	.025*
	$\hat{\pi}_{TRM}$	$= \hat{\pi}_{DQ}$	3.72	0.66	.417
CWM	$\hat{\pi}_{left}$	$= \hat{\pi}_{right}$	5.52	0.69	.407
TRM	$\hat{\pi}_{left}$	$= \hat{\pi}_{right}$	12.54	5.88	.015*
DQ	$\hat{\pi}_{left}$	$= \hat{\pi}_{right}$	15.12	21.50	< .001*
$\hat{\alpha}_{CWM, left:right}$	$=$	$\hat{\alpha}_{TRM, left:right}$	31.69	1.79	.181
$\hat{\alpha}_{CWM, left:right}$	$=$	$\hat{\alpha}_{DQ, left:right}$	51.06	7.54	.006*
$\hat{\alpha}_{TRM, left:right}$	$=$	$\hat{\alpha}_{DQ, left:right}$	19.37	1.29	.257
Opposition to refugee admissions					
	$\hat{\pi}_{CWM}$	$= \hat{\pi}_{TRM}$	10.51	2.83	.093
"left"	$\hat{\pi}_{CWM}$	$= \hat{\pi}_{DQ}$	13.47	5.84	.016*
	$\hat{\pi}_{TRM}$	$= \hat{\pi}_{DQ}$	2.96	0.38	.538
	$\hat{\pi}_{CWM}$	$= \hat{\pi}_{TRM}$	2.60	0.19	.660
"right"	$\hat{\pi}_{CWM}$	$= \hat{\pi}_{DQ}$	0.59	0.01	.918
	$\hat{\pi}_{TRM}$	$= \hat{\pi}_{DQ}$	3.19	0.41	.521
CWM	$\hat{\pi}_{left}$	$= \hat{\pi}_{right}$	14.50	4.60	.032*
TRM	$\hat{\pi}_{left}$	$= \hat{\pi}_{right}$	22.41	16.60	< .001*
DQ	$\hat{\pi}_{left}$	$= \hat{\pi}_{right}$	28.56	40.49	< .001*
$\hat{\alpha}_{CWM, left:right}$	$=$	$\hat{\alpha}_{TRM, left:right}$	18.14	1.52	.217
$\hat{\alpha}_{CWM, left:right}$	$=$	$\hat{\alpha}_{DQ, left:right}$	27.29	4.75	.029*
$\hat{\alpha}_{TRM, left:right}$	$=$	$\hat{\alpha}_{DQ, left:right}$	9.15	0.70	.404

*Significant at $p < .05$
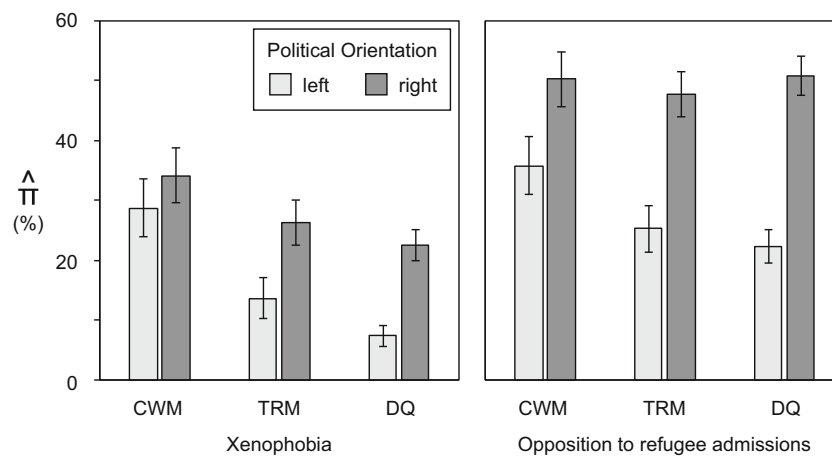
Fig. 2 Prevalence estimates for xenophobia (left panel) and opposition to refugee admissions in Germany (right panel) by political orientation (median split). CWM = crosswise model, TRM = triangular model, DQ = direct questioning

questioning techniques provided significantly higher prevalence estimates among politically right-oriented than politically left-oriented respondents. In the CWM condition, however, this difference was descriptively smaller than in the two other conditions. Interaction analyses revealed a significant interaction between questioning technique and political orientation for both sensitive attributes. Shrinkage factors indicated that the difference in prevalence estimates between politically right-oriented and politically left-oriented respondents was significantly higher in the DQ than in the CWM condition (see Fig. 2).

In the subsample of left-oriented respondents, prevalence estimates in the CWM condition were higher and thus presumably more valid than prevalence estimates in the DQ condition. In the subsample of right-oriented respondents, however, the CWM estimates only slightly and insignificantly exceeded those obtained via DQ. Thus, social desirability bias seems to have exerted a substantially stronger influence on politically left-oriented than politically right-oriented respondents. This pattern of results suggests higher perceived sensitivity of the questions measuring xenophobia and opposition to refugee admissions among politically left-oriented compared with politically right-oriented respondents.

Discussion

In this study, we conducted the first experimental comparison of the validity of a symmetric (crosswise model; CWM) and an asymmetric NRRT (triangular model; TRM), and contrasted the performance of these two models with a conventional direct questioning approach (DQ). To this end, we assessed respondents' attitudes towards xenophobia and opposition to the further intake of refugees in Germany with two sensitive statements of unknown prevalence. Additionally, following a "strong" validation approach, we included a

nonsensitive control statement (first letter of respondents' surnames) with known prevalence to test for potential method-specific biases leading to a general tendency towards over- or underestimation.

As expected, both NRRTs yielded higher estimates for xenophobia than DQ. However, only the CWM provided estimates that were significantly higher than the estimates obtained via DQ, thus sufficing the "more is better" criterion. Moreover, the CWM estimates were significantly higher than the TRM estimates, indicating the superiority of the symmetric CWM over the asymmetric TRM. For opposition to further refugee admissions, the CWM yielded descriptively higher prevalence estimates than TRM and DQ; however, none of the pairwise comparisons of parameter estimates was significant. All three questioning techniques accurately recovered the known prevalence of the nonsensitive control attribute. Thus, both the CWM and the TRM met the criteria of a successful "strong" validation. As expected, exploratory analyses of the influence of political orientation revealed that prevalence estimates for the sensitive attributes were higher among right-oriented respondents than among left-oriented respondents. Interestingly, interaction analyses showed that this discrepancy was less pronounced when prevalence estimates were obtained via the CWM than via the TRM or DQ.

Our results indicate that while both NRRTs outperformed DQ, only the CWM satisfied the "more is better" criterion for xenophobia and was thus better able to control for social desirability with respect to this question than DQ. The estimates obtained via the TRM were descriptively but not significantly higher than estimates obtained via DQ. None of the questioning techniques under investigation exhibited a method-specific tendency towards over- or underestimation; instead, the prevalence estimates for the nonsensitive control attribute with known prevalence were highly accurate for all questioning techniques. In light of these findings, we recommend that the CWM be used to control for social desirability

bias and maximize the validity of prevalence estimates in surveys of sensitive personal attributes.

We also found that for xenophobia, the symmetric CWM outperformed the asymmetric TRM in terms of validity. Response symmetry, or the absence of an objectively "safe" answer option, has been shown to increase the confidentiality of individual answers and thus also compliance with RRT instructions (Ostapczuk, Moshagen, et al., 2009a). Response symmetry seems to prevent respondents from faking their answers (cf. Hoffmann et al., 2019), either because they understand that their privacy is perfectly protected and they cannot make a negative impression or because they are simply unable to identify a self-protective response. In contrast, the asymmetric TRM offers a "safe" answer option and therefore seems to be more prone to deliberate faking than the CWM. The TRM offers the theoretical advantage that, under many conditions, it provides the more efficient estimates (Yu et al., 2008), as also confirmed by smaller standard errors for TRM than for CWM estimates in the current study. However, the CWM was found to be clearly preferable to the TRM in terms of the superordinate criterion of measurement validity. Interestingly, our results are in line with the results of two previous studies examining plagiarism, one via the CWM, the other via the TRM (Jann et al., 2012; Jerke & Krumpal, 2013). The study applying the CWM obtained descriptively higher prevalence estimates than the study applying the TRM. However, this observation is based on a comparison across studies and samples, and therefore does not provide unequivocal evidence for the superiority of one of the two models. The present study conducted a first experimental comparison of the two models, and found direct evidence for the assumption first formulated by Jann et al. (2012) and Jerke and Krumpal (2013) that the CWM outperforms the TRM in terms of validity.

In an exploratory analysis, we found a moderating influence of self-reported political orientation on prevalence estimates. For left-oriented respondents, the CWM provided substantially higher prevalence estimates than DQ, indicating that within this subgroup, the prevalence estimates obtained via DQ were strongly distorted by social desirability bias. For right-oriented respondents, the CWM obtained only slightly higher prevalence estimates than DQ, indicating that within this subgroup, social desirability bias had a somewhat weaker impact, as the topics under investigation were likely perceived as less sensitive by respondents less reluctant to openly express negative attitudes towards foreigners. Hence, the CWM proved particularly effective among left-oriented respondents, affirming the assumption that the usefulness of indirect questioning techniques increases with topic sensitivity (cf. Lensvelt-Mulders et al., 2005). Consequently, we particularly recommend that indirect questioning techniques such as the CWM be applied when investigating issues that are highly sensitive for a particular group of respondents, as a strong social desirability influence might result in strongly biased results for such groups if only direct self-reports are used.

Limitations and future research directions

It is necessary to acknowledge that the student population investigated in the present study is not representative of the population at large. Therefore, our pattern of findings and the generalizability of the prevalence estimates obtained in the present study are limited to the sample we investigated, and would need to be replicated in other populations. Estimates for the prevalence of xenophobic attitudes might turn out to be even higher in a more representative sample also including lower-educated respondents, as lower education has repeatedly been shown to be associated with a higher incidence of xenophobic attitudes (D'Ancona, 2013; Hjerm, 2001; Ostapczuk, Musch, et al., 2009b; Zick et al., 2011). Student-only samples are also presumably more homogeneous, thereby increasing the statistical power to detect differences between questioning techniques. Lower-educated respondents generally exhibit greater problems understanding indirect questioning techniques (Hoffmann et al., 2017) and therefore tend to produce more false positives (Meisters et al., 2019). Developing better instructions that are easily comprehensible even for lower-educated respondents is thus of considerable importance for future research using randomized and nonrandomized response techniques in more heterogeneous samples.

The results of the current study revealed no method-specific tendency for over- or underestimation. This result is in line with several other studies that also found no deviation between CWM estimates and the known prevalence of a nonsensitive control attribute (Hoffmann & Musch, 2016) and a sensitive attribute (Hoffmann et al., 2015). To check for potential bias in the form of a general preference for one of the two answer options, future studies should try to replicate the present results using the extended crosswise model (ECWM; Heck, Hoffmann, & Moshagen, 2018), a recent modification of the CWM that allows for detecting instruction nonadherence without negatively affecting statistical efficiency.

As a final remark, it should be noted that the difference between the CWM and DQ condition was larger for self-reported ethnic discrimination than for self-reported opposition to further refugee admissions. A potential explanation for this finding is that opposition to further refugee admissions was perceived by respondents as less sensitive than the expression of xenophobic attitudes. This reasoning is supported by the higher percentage of respondents admitting that they opposed further intake of refugees (36.73%) compared with the much lower percentage of respondents admitting to

xenophobia (15.45%) in the DQ condition. The lower perceived sensitivity of opposing further refugee intake might potentially be fueled by the increasing popularity of right-wing populist parties such as the *Alternative for Germany,* which cites social and economic concerns as a reason for limiting further refugee admissions.

Conclusions

The present research showed that non-randomized response techniques provide more valid prevalence estimates for socially undesirable attributes than conventional direct questions (DQ). The crosswise model (CWM) in particular was able to successfully control for the influence of social desirability bias, and outperformed the triangular model (TRM), presumably due to the favorable influence of the response symmetry found in the CWM but not the TRM. We also found that the sensitivity of two questions was contingent on respondents' political orientation, and that the CWM provided the most valid estimates for respondents for whom these questions were most sensitive. Based on these results, we recommend the use of the CWM over the TRM or DQ for topics that are highly sensitive in a survey's target population.

Open practice statement All data and equation files necessary to reproduce the parameter estimates reported in this manuscript are provided in appendices A and B in the electronic supplementary material.

References

Alba, R., & Johnson, M. (2000). Zur Messung aktueller Einstellungsmuster gegenüber Ausländern in Deutschland [On the measurment of current attitudes towards foreigners in Germany]. In R. Alba, P. Schmidt, & M. Wasmer (Eds.), *Blickpunkt Gesellschaft 5. Deutsche und Ausländer: Freunde, Fremde oder Feinde?* (pp. 229-253). Wiesbaden: Westdeutscher Verlag.

Batchelder, W. H. (1998). Multinomial processing tree models and psychological assessment. *Psychological Assessment, 10*, 331-344. doi: https://doi.org/10.1037/1040-3590.10.4.331

Batchelder, W. H., & Riefer, D. M. (1999). Theoretical and empirical review of multinomial process tree modeling. *Psychonomic Bulletin & Review, 6*, 57-86. doi:https://doi.org/10.3758/Bf03210812

Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances. *Philosophical transactions of the Royal Society of London, 53*, 370-418.

Bertelsmann Stiftung. (2017). *Willkommenskultur im „Stresstest" - Einstellungen in der Bevölkerung 2017 und Entwicklungen und Trends seit 2011/2012 - Ergebnisse einer repräsentativen Bevölkerungsumfrage [Welcoming culture under 'stress test' - attitudes in the population in 2017 and developments and trends since 2011/2012 - findings of a representative survey of the population].* Retrieved from http://www.bertelsmann-stiftung.de/fileadmin/files/Projekte/28_Einwanderung_und_Vielfalt/IB_Umfrage_Willkommenskultur_2017.pdf

Berufsverband Deutscher Psychologinnen und Psychologen & Deutsche Gesellschaft für Psychologie. (2016). Berufsethische Richtlinien des Berufsverbandes Deutscher Psychologinnen und Psychologen e.V. und der Deutschen Gesellschaft für Psychologie e.V. [Professional ethical guidelines of the Berufsverband Deutscher Psychologinnen und Psychologen e.V. and the Deutsche Gesellschaft für Psychologie e.V.]. Retrieved Sep 10th, 2018 https://www.dgps.de/fileadmin/documents/Empfehlungen/berufsethische_richtlinien_dgps.pdf

Bogardus, E. S. (1933). A Social Distance Scale. *Sociology & Social Research, 17*, 265-271.

D'Ancona, M. Á. C. (2013). Measuring xenophobia: Social desirability and survey mode effects. *Migration Studies*, 1-26.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum Likelihood from Incomplete Data Via Em Algorithm. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 39*, 1-38.

Erdmann, A. (2019). Non-Randomized Response Models: An Experimental Application of the Triangular Model as an Indirect Questioning Method for Sensitive Topics. *Methods, data, analyses, 13*, 139-167. doi:https://doi.org/10.12758/mda.2018.07

Fligner, M. A., Policello, G. E., & Singh, J. (1977). A Comparison of 2 Randomized Response Survey Methods with Consideration for Level of Respondent Protection. *Communications in Statistics Part a-Theory and Methods, 6*(15), 1511-1524. doi:https://doi.org/10.1080/03610927708827593

Heck, D. W., Hoffmann, A., & Moshagen, M. (2018). Detecting nonadherence without loss in efficiency: A simple extension of the crosswise model. *Behavior Research Methods, 50*, 1895-1905. doi:https://doi.org/10.3758/s13428-017-0957-8

Heitmeyer, W. (2012). Gruppenbezogene Menschenfeindlichkeit (GMF) in einem entsicherten Jahrzehnt [Group-focused enmity in a disengaged decade]. In W. Heitmeyer (Ed.), *Deutsche Zustände: Folge 10* (pp. 15-41). Berlin: Suhrkamp.

Hjerm, M. (2001). Education, xenophobia and nationalism: A comparative analysis. *Journal of Ethnic and Migration Studies, 27*, 37-60. doi:https://doi.org/10.1080/13691830124482

Hoffmann, A., Diedenhofen, B., Verschuere, B. J., & Musch, J. (2015). A strong validation of the Crosswise Model using experimentally induced cheating behavior. *Experimental Psychology, 62*, 403-414. doi:https://doi.org/10.1027/1618-3169/a000304

Hoffmann, A., Meisters, J., & Musch, J. (2019). Nothing but the truth? Effects of faking on the validity of the Crosswise Model. [Manuscript submitted for publication].

Hoffmann, A., & Musch, J. (2016). Assessing the validity of two indirect questioning techniques: a Stochastic Lie Detector versus the Crosswise Model. *Behavior Research Methods, 48*, 1032-1046. doi:https://doi.org/10.3758/s13428-015-0628-6

Hoffmann, A., & Musch, J. (2019). Prejudice against women leaders: Insights from an indirect questioning approach. *Sex Roles, 80*, 681-692. doi:https://doi.org/10.1007/s11199-018-0969-6

Hoffmann, A., Waubert de Puiseau, B., Schmidt, A. F., & Musch, J. (2017). On the comprehensibility and perceived privacy protection of indirect questioning techniques. *Behavior Research Methods, 49*, 1470-1483. doi:https://doi.org/10.3758/s13428-016-0804-3

Höglinger, M., & Diekmann, A. (2017). Uncovering a Blind Spot in Sensitive Question Research: False Positives Undermine the Crosswise-Model RRT. *Political Analysis, 25*, 131-137. doi:https://doi.org/10.1017/pan.2016.5

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

Hu, X., & Batchelder, W. H. (1994). The Statistical-Analysis of General Processing Tree Models with the Em Algorithm. *Psychometrika, 59*, 21-47. doi:https://doi.org/10.1007/Bf02294263

Jann, B., Jerke, J., & Krumpal, I. (2012). Asking Sensitive Questions Using the Crosswise Model. *Public Opinion Quarterly, 76*, 32-49. doi:https://doi.org/10.1093/Poq/Nfr036

Jerke, J., & Krumpal, I. (2013). Plagiate in studentischen Arbeiten [Plagiarism in Student Papers]. *methoden, daten, analysen, 7*, 347-368. doi:https://doi.org/10.12758/mda.2013.017

Jimenez, P. (1999). Weder Opfer noch Täter - die alltäglichen Einstellungen 'unbeteiligter' Personen gegenüber Ausländern [Neither victim nor offender—the common attitudes of 'non-involved' persons towards foreigners]. In R. Dollase, T. Kliche, & H. Moser (Eds.), *Politische Psychologie der Fremdenfeindlichkeit. Opfer - Täter - Mittäter* (pp. 293–306). Weinheim: Juventa.

Knapp, B. R., & Batchelder, W. H. (2004). Representing parametric order constraints in multi-trial applications of multinomial processing tree models. *Journal of Mathematical Psychology, 48*, 215-229. doi: https://doi.org/10.1016/j.jmp.2004.03.002

Korndörfer, M., Krumpal, I., & Schmukle, S. C. (2014). Measuring and explaining tax evasion: Improving self-reports using the crosswise model. *Journal of Economic Psychology, 45*, 18-32. doi:https://doi.org/10.1016/j.joep.2014.08.001

Krumpal, I. (2012). Estimating the prevalence of xenophobia and anti-Semitism in Germany: A comparison of randomized response and direct questioning. *Social Science Research, 41*, 1387-1403. doi: https://doi.org/10.1016/j.ssresearch.2012.05.015

Kundt, T. C., Misch, F., & Nerré, B. (2017). Re-assessing the merits of measuring tax evasion through business surveys: an application of the crosswise model. *International Tax and Public Finance, 24*, 112-133. doi:https://doi.org/10.1007/s10797-015-9373-0

Landsheer, J. A., van der Heijden, P. G. M., & van Gils, G. (1999). Trust and understanding, two psychological aspects of randomized response - A study of a method for improving the estimate of social security fraud. *Quality & Quantity, 33*, 1-12. doi:https://doi.org/10.1023/A:1004361819974

Lanke, J. (1976). Degree of Protection in Randomized Interviews. *International Statistical Review, 44*, 197-203. doi:https://doi.org/10.2307/1403277

Lensvelt-Mulders, G. J. L. M., Hox, J. J., van der Heijden, P. G. M., & Maas, C. J. M. (2005). Meta-analysis of randomized response research: thirty-five years of validation. *Sociological Methods & Research, 33*, 319-348. doi:https://doi.org/10.1177/0049124104268664

Meisters, J., Hoffmann, A., & Musch, J. (2019). Can detailed instructions and comprehension checks increase the validity of crosswise model estimates? [Manuscript under preparation].

Moshagen, M. (2010). multiTree: A computer program for the analysis of multinomial processing tree models. *Behavior Research Methods, 42*, 42-54. doi:https://doi.org/10.3758/BRM.42.1.42

Moshagen, M., Hilbig, B. E., & Musch, J. (2011). Defection in the dark? A randomized-response investigation of cooperativeness in social dilemma games. *European Journal of Social Psychology, 41*, 638-644. doi:https://doi.org/10.1002/Ejsp.793

Moshagen, M., & Musch, J. (2012). Surveying Multiple Sensitive Attributes using an Extension of the Randomized-Response Technique. *International Journal of Public Opinion Research, 24*, 508-523. doi:https://doi.org/10.1093/ijpor/edr034

Moshagen, M., Musch, J., & Erdfelder, E. (2012). A stochastic lie detector. *Behavior Research Methods, 44*, 222-231. doi:https://doi.org/10.3758/s13428-011-0144-221858604

Nakhaee, M. R., Pakravan, F., & Nakhaee, N. (2013). Prevalence of Use of Anabolic Steroids by Bodybuilders Using Three Methods in a City of Iran. *Addiction & Health, 5*, 77–82.

Ostapczuk, M., Moshagen, M., Zhao, Z., & Musch, J. (2009a). Assessing sensitive attributes using the randomized response technique: Evidence for the importance of response symmetry. *Journal of Educational and Behavioral Statistics, 34*, 267-287. doi:https://doi.org/10.3102/1076998609332747

Ostapczuk, M., Musch, J., & Moshagen, M. (2009b). A randomized-response investigation of the education effect in attitudes towards foreigners. *European Journal of Social Psychology, 39*, 920-931. doi:https://doi.org/10.1002/ejsp.588

Ostapczuk, M., Musch, J., & Moshagen, M. (2011). Improving self-report measures of medication non-adherence using a cheating detection extension of the randomised-response-technique. *Statistical Methods in Medical Research, 20*, 489-503. doi:https://doi.org/10.1177/0962280210372843

Paulhus, D. L. (1991). Measurement and Control of Response Bias. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes, Vol. 1* (pp. 17-59). San Diego, CA: Academic Press.

Pötzsch, O. (2012). Geburten in Deutschland [Births in Germany]. Retrieved Jun 6, 2012, from German Federal Statistical Office, https://www.destatis.de/DE/Publikationen/Thematisch/Bevoelkerung/Bevoelkerungsbewegung/BroschuereGeburtenDeutschland0120007129004.pdf

Reinders, M. (1996). Häufigkeit von Namensanfängen [Frequency of first letters of surnames]. *Statistische Rundschau Nordrhein-Westfalen, 11*, 651-660.

Silbermann, A., & Hüsers, F. (1995). *Der 'normale' Haß auf die Fremden. Eine sozialwissenschaftliche Studie zu Ausmaß und Hintergründen von Fremdenfeindlichkeit in Deutschland [The 'normal' xenophobia. A socio-scientific study on the extent and determinants of xenophobia in Germany]*. München: Quintessenz.

Soeken, K. L., & Macready, G. B. (1982). Respondents Perceived Protection When Using Randomized-Response. *Psychological Bulletin, 92*, 487-489. doi:https://doi.org/10.1037/0033-2909.92.2.487

Thielmann, I., Heck, D. W., & Hilbig, B. E. (2016). Anonymity and incentives: An investigation of techniques to reduce socially desirable responding in the Trust Game. *Judgment and Decision Making, 11*, 527-536.

Tian, G.-L., & Tang, M.-L. (2014). *Incomplete Categorical Data Design: Non-Randomized Response Techniques for Sensitive Questions in Surveys*. Boca Raton, FL: CRC Press, Taylor & Francis Group.

Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin, 133*, 859-883. doi:https://doi.org/10.1037/0033-2909.133.5.85917723033

Ulrich, R., Schröter, H., Striegel, H., & Simon, P. (2012). Asking Sensitive Questions: A Statistical Power Analysis of Randomized Response Models. *Psychological Methods, 17*, 623-641. doi:https://doi.org/10.1037/A0029314

Umesh, U. N., & Peterson, R. A. (1991). A Critical Evaluation of the Randomized-Response Method - Applications, Validation, and Research Agenda. *Sociological Methods & Research, 20*, 104-138. doi:https://doi.org/10.1177/0049124191020001004

Wagner, U., & van Dick, R. (2001). Fremdenfeindlichkeit "in der Mitte der Gesellschaft": Phänomenbeschreibung, Ursachen, Gegenmaßnahmen [Xenophobia "in the middle of society": description of phenomena, causes, countermeasures]. *Zeitschrift für Politische Psychologie, 9*, 41-54.

Warner, S. L. (1965). Randomized-Response - a Survey Technique for Eliminating Evasive Answer Bias. *Journal of the American Statistical Association, 60*, 63-69.

Waubert de Puiseau, B., Hoffmann, A., & Musch, J. (2017). How indirect questioning techniques may promote democracy: A pre-election polling experiment. *Basic And Applied Social*

Psychology, 39, 209-217. doi:https://doi.org/10.1080/01973533.2017.1331351

World Medical Association. (2013). World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA, 310*, 2191-2194. doi:https://doi.org/10.1001/jama.2013.281053

Yu, J.-W., Tian, G.-L., & Tang, M.-L. (2008). Two new models for survey sampling with sensitive characteristic: design and analysis. *Metrika, 67*, 251-263. doi:https://doi.org/10.1007/s00184-007-0131-x

Zick, A., Hövermann, A., & Krause, D. (2012). Die Abwertung von Ungleichwertigen - Erklärung und Prüfung eines erweiterten Syndroms der Gruppenbezogenen Menschenfeindlichkeit [The devaluation of inferiors - Explaining and examining an expanded syndrome of group-focused enmity]. In W. Heitmeyer (Ed.), *Deutsche Zustände: Folge 10* (pp. 64-86). Berlin: Suhrkamp.

Zick, A., Küpper, B., & Hövermann, A. (2011). *Intolerance, Prejudice and Discrimination: A European Report*. Berlin: Friedrich-Ebert-Stiftung.