Article

Sociological Methods & Research
2024, Vol. 53(1) 296–327

DOI: 10.1177/00491241211043138
journals.sagepub.com/home/smr

Self-protecting responses in randomized response designs: A survey on intimate partner violence during the coronavirus disease 2019 pandemic

Fabiola Reiber[1], **Donna Bryce**[1], **and Rolf Ulrich**[1]

Abstract

Randomized response techniques (RRTs) are applied to reduce response biases in self-report surveys on sensitive research questions (e.g., on socially undesirable characteristics). However, there is evidence that they cannot completely eliminate self-protecting response strategies. To address this problem, there are RRTs specifically designed to measure the extent of such strategies. Here we assessed the recently devised unrelated question model—cheating extension (UQMC) in a preregistered online survey on intimate partner violence (IPV) victimization and perpetration during the first contact restrictions as containment measures for the outbreak of the coronavirus disease 2019 pandemic in Germany in early 2020. The UQMC accounting for self-protecting responses described the data better than its predecessor model which assumes instruction adherence. The resulting three-month prevalence estimates were about 10% and we found a high proportion of self-protecting responses in the group of female participants queried about IPV victimization. However, unexpected results

[1]Department of Psychology, University of Tübingen, Tübingen, Germany

Corresponding Author:
Fabiola Reiber, Department of Psychology, University of Tübingen, Tübingen, Germany.
Email: fabiola.reiber@uni-tuebingen.de

concerning the differences in prevalence estimates across the groups queried about victimization and perpetration highlight the difficulty of investigating sensitive research questions even using methods that guarantee anonymity and the importance of interpreting the respective estimates with caution.

Keywords

Sensitive research questions, randomized response techniques, cheater detection, intimate partner violence, self-protecting responses

Many social and psychological phenomena of high societal relevance are difficult to investigate empirically because of their sensitive nature. For instance, the German news broadcaster Tagesschau recently reported an alarming increase in the incidence of intimate partner violence (IPV) in criminal statistics during the ongoing coronavirus disease 2019 (COVID-19) pandemic (Emundts 2020). However, criminal statistics are assumed to underestimate the actual numbers, because they only capture legally reported cases and the dark figure, that is, the number of non-registered cases might substantially exceed these numbers. Problematically, the dark figure of cases of IPV is difficult to investigate because it is a highly stigmatized topic (e.g., Ellsberg et al. 2001; Gracia 2004). Both victimization and perpetration of IPV are perceived as socially undesirable and reporting is associated with negative consequences (e.g., Schröttle 2015; Franke et al. 2004). Social desirability and fear of stigmatization or other negative consequences can influence response behavior in surveys and interviews (Tourangeau and Yan 2007). Specifically, survey respondents can be inclined not to respond at all, especially if they carry the investigated undesirable or stigmatized attribute, or to give an untruthful self-protecting response. Although both these behaviors are employed to disguise ones own individual status, they bias group-level estimates of dark figures as well. The consequence is that the extent of societal problems such as IPV can be underestimated by surveys (Tourangeau and Yan 2007). This impairment concerns a variety of research fields in the social sciences that address sensitive characteristics.

Randomized Response Techniques

To overcome self-protecting response strategies in surveys on sensitive attributes, randomized response techniques (RRTs; Warner 1965) were developed to assure the protection of the respondents' anonymity. Specifically, a randomization device (such as a die) is employed to ambiguate single responses and thus make them inconclusive toward the carrier status of a single respondent. For instance, in the unrelated question model (UQM; Greenberg et al. 1969) version of the RRT, a randomization device decides whether a respondent shall answer the sensitive question S of interest, such as "Have you ever been physically assaulted by a partner?" or an unrelated neutral question N, such as "Is your mother's birthday in the first half of the year?" In the case of employing a die as a randomization device, the instruction could be to answer the sensitive question S, if the die comes up 1 through 4, and the neutral question N, if it comes up 5 or 6. Importantly, only the response to either question but not the outcome of the randomization is reported. Therefore, a "Yes"-response could either mean that the respondent has been physically assaulted by a partner or that their mother's birthday is in the first half of the year. Consequently, it remains concealed whether a specific respondent was physically assaulted by a partner and, theoretically, respondents have no reason to employ self-protecting response strategies that could bias prevalence estimates. In the current study, we applied such a technique to estimate the prevalence of IPV during the first COVID-19 related contact restrictions in Germany in spring 2020.

Importantly, it is possible to compute these prevalence estimates using the known probabilities underlying the questioning design. Figure 1 depicts the probabilities underlying "Yes" and "No" responses in the UQM. A "Yes" response can come (a) from a respondent who was instructed to respond to the sensitive question S with probability p and carries the sensitive attribute with probability π and (b) a respondent who was instructed to respond to the neutral question N with probability $(1-p)$ and carries the neutral attribute with probability q. Therefore, the overall probability to respond "Yes" is

$$\lambda = p \cdot \pi + (1-p) \cdot q. \tag{1}$$

The randomization probability p is known—in the example using a die, above, it is $p = 4/6 = .67$. The neutral question can be chosen such that the neutral prevalence q is also known. In the example above it is $q \approx .5$, assuming a uniform distribution of birthdays across the year, which is a reasonable assumption based on the birthdate records over the last 50 years in Germany (Statistisches Bundesamt 2020). The probability λ to respond "Yes" can be estimated from the proportion of "Yes" responses in a sufficiently large sample such that equation (1) can be rearranged to estimate

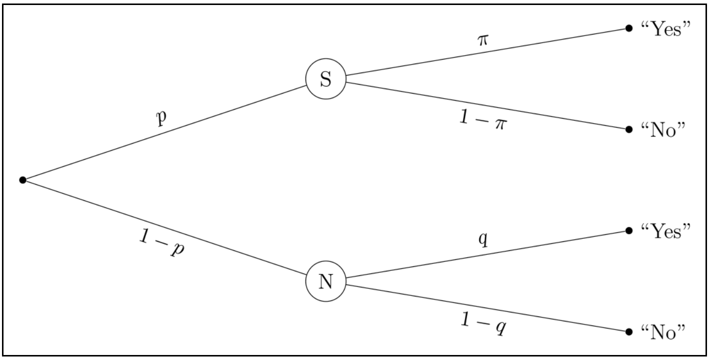

Figure 1. Probability tree of the unrelated question model (UQM).
Note. The sensitive question S and the neutral question N are randomly received by respondents with probability p and $1-p$, respectively. The probabilities of responding "Yes" and "No" to the neutral question N are q and $1-q$ and the probabilities of responding "Yes" and "No" to the sensitive question S are π and $1-\pi$. Adapted from Reiber et al. (2020).

the prevalence of the sensitive attribute:

$$\hat{\pi}_{UQM} = \frac{\hat{\lambda} - (1-p) \cdot q}{p}. \tag{2}$$

Because the respondents' anonymity is protected, this estimate is expected to be less biased due to self-protecting response strategies. In fact, there is evidence that RRT applications elicit prevalence estimates that are less biased toward the socially desirable response option (e.g., Moshagen et al. 2010; Ulrich et al. 2018; Wimbush and Dalton 1997) and closer to a known true prevalence (e.g., Horvitz et al. 1976; van der Heijden et al. 2000).

Non-Adherence to Instructions in RRTs

However, there are reasons to doubt that even with the RRT there is full honesty in responding. A number of studies did not find RRT estimates to be more valid than those from studies using direct questions (e.g., Höglinger and Diekmann 2017; Höglinger and Jann 2018; Holbrook and Krosnick 2010). A possible explanation for this finding is that the instructions of the RRT are difficult to understand (Hoffmann et al. 2017) and there is still a lack of trust in the anonymity protection (Höglinger et al. 2016). One way to address this problem is to increase comprehensibility (e.g., Meisters et al. 2020). Another way is to quantify the extent of non-compliance with instructions. In this vein, some RRT extensions, such as the cheater detection model (CDM; Clark and Desharnais 1998) or the stochastic lie detector (Moshagen et al. 2012) include parameters for specific types of instruction non-adherence. Especially the CDM has been applied in a number of studies (e.g., Elbe and Pitsch 2018; Moshagen et al. 2010; Ostapczuk et al. 2011; Pitsch et al. 2007; Schröter et al. 2016). It is based on another RRT variant, the forced response technique (Boruch 1971), which is similar to the UQM. The only difference is that the alternative to the sensitive question is not a neutral question but the instruction to respond "Yes." In the CDM, respondents are considered to be either honest and follow the instructions or to be cheaters and give a "No"-response irrespective of the outcome of the randomization and their carrier status. The latter can serve to evade being seen as a carrier of the sensitive attribute and has thus been termed a self-protective response strategy (Böckenholt and van der Heijden 2007). Based on this categorization, two parameters can be estimated: The proportion of cheaters γ and the proportion of honest carriers π_{CDM}.[1] To allow for the estimation of both these parameters, two independent estimates of the probability of a "Yes" response are required. To that end, two independent samples are assessed using varying levels of the randomization probability p. Studies applying this design found substantial proportions of cheating (Elbe and Pitsch 2018; Moshagen et al. 2010; Ostapczuk et al. 2011; Pitsch et al. 2007; Schröter et al. 2016). Thus, it seems to be reasonable to include a cheating parameter in RRTs.

It is important to note, however, that the CDM still makes strong assumptions about the nature of instruction non-adherence. For instance, the varying levels of the randomization probability p, which are employed to enable the estimation of the cheating parameter, are assumed not to influence responses. In practice, different randomization probabilities could influence the subjective anonymity protection and thereby the probability to cheat.[2] Unfortunately, assumptions, such as this assumption of randomization probability independence, are not testable within the CDM. The variation of p across two independent samples allows for the estimation of both parameters. However, the resulting model is saturated, which means that it is not possible to assess model fit or, for instance, test whether cheating differs between the subsamples.[3]

The Unrelated Question Model—Cheating Extension

The recently proposed unrelated question model—cheating extension (UQMC; Reiber et al. 2020) transfers the CDM's concept of cheating to the UQM's design. The reason for devising this extension was that the psychological acceptability of the UQM has been found to be superior to that of the forced response method (Höglinger et al. 2016). As such, the UQM can be seen as less fallible to self-protecting responses, since there is no response option that clearly rules out being a carrier of the sensitive attribute (one could respond "No" to the neutral question and still be a carrier of the sensitive attribute). However, also in the UQM "No" can be seen as a self-protecting response since the conditional likelihood of being a carrier is always lower given a "No" than given a "Yes" response. Additionally, and probably more intuitively from a respondent's perspective, a "No" response to the neutral question can naively be interpreted as a response with which being a carrier of the sensitive attribute is negated. Thus, it is worthwhile to investigate whether cheating occurs in the UQM as well.

Another major advantage of embedding the cheating concept within the UQM is that it is possible to test the model's assumptions. In contrast to the CDM, the UQM incorporates a second design parameter that can be varied, namely the prevalence of the neutral attribute q. Therefore, four independent samples can be assessed and the gained degrees of freedom make the model testable.

From Figure 2, the probability of a "Yes" response in sample i in the UQMC is

$$\lambda_i = (1-\gamma) \cdot \left[p_i \cdot \epsilon + (1-p_i) \cdot q_i\right]. \tag{3}$$

Only respondents who do not cheat would respond "Yes." If they are assigned to the sensitive question S (with randomization probability p_i in sample $i \in \{1, 2, 3, 4\}$), honest respondents answer "Yes" with probability ϵ, that is, the true prevalence of the sensitive attribute. If they are assigned to the neutral question N (with probability $1 - p_i$ in sample i), honest respondents answer "Yes" with probability q_i, that is the prevalence of the neutral attribute in sample i. Following the logic of the CDM, $\pi_{\text{UQMC}} = (1-\gamma) \cdot \epsilon$ is the prevalence of honest carriers, that is, the joint probability of not cheating and of being a carrier of the sensitive attribute. Therefore,

$$\lambda_i = p_i \cdot \pi_{\text{UQMC}} + (1-\gamma) \cdot (1-p_i) \cdot q_i. \tag{4}$$

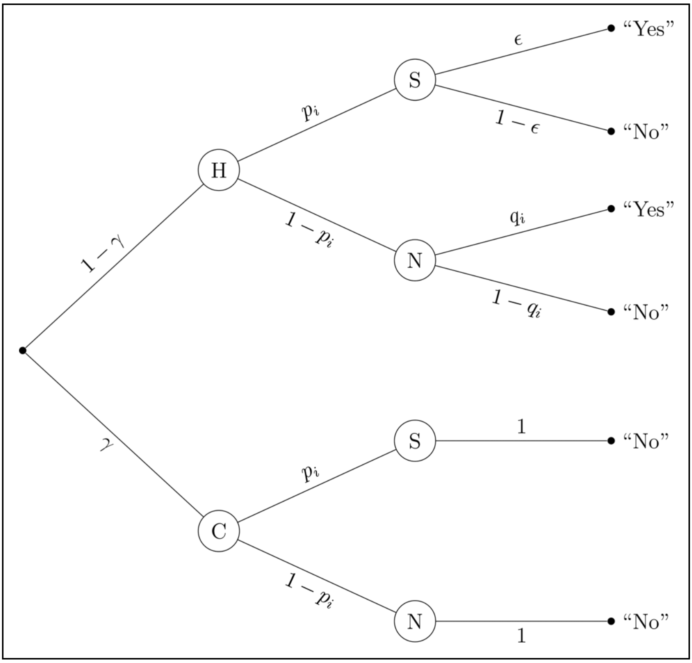

Figure 2. Probability tree of the unrelated question model—cheating extension (UQMC). *Note*. The prevalence of cheaters C is γ and the prevalence of honest participants H is $1-\gamma$. In both cases, the sensitive question S and the neutral question N are received by participants with probability p_i and $1-p_i$, respectively. The model assumes that cheaters always say “No” regardless of the question received. Honest participants respond “Yes” with probability q_i and “No” with probability $1-q_i$ if instructed to answer the neutral question N. They answer “Yes” with probability ϵ and “No” with probability $1-\epsilon$, if instructed to answer the sensitive question S. Thus, there are three groups of participants: (a) honest participants who are carriers of the sensitive attribute, who will respond “Yes” with probability $(1-\gamma)\cdot\epsilon=\pi$ if they receive S; (b) honest non-carriers of this attribute who will respond “No” with probability $(1-\gamma)\cdot(1-\epsilon)$ if they receive S; and (c) cheaters, who will respond “No” with probability γ regardless of whether they receive S or N. Adapted from Reiber et al. (2020).

There are no closed-form equations to compute estimates for π_{UQMC} and γ from the four samples’ estimated probabilities of a “Yes” response $\hat{\lambda}_i$.[4] Instead, the parameters must be estimated using numerical likelihood optimization.

The development and properties of the UQMC are described in more detail in Reiber et al. (2020). However, the validity of the model has so far not been investigated empirically. The aim of the present study was, therefore, to test the UQMC's validity in an empirical investigation and to assess whether it provides an advantage over its predecessor model, the original UQM.

Present Study

There are different approaches to assessing a model's validity. One widely accepted approach is to compare prevalence estimates with a known criterion, optimally on an individual level (e.g., Hoffmann et al. 2015). Unfortunately, the prevalences of highly sensitive topics are often not known, especially not on an individual level. Therefore, studies using this approach often use experimentally induced behaviors for sensitive characteristics, such as cheating for an extra pay-off in the survey (e.g., Hoffmann et al. 2015). However, these characteristics differ from those addressed in typical RRT applications because RRTs are most useful for investigating highly sensitive topics (see Lensvelt-Mulders et al. 2005). Therefore, in the present study, we chose another approach to assess the UQMC's validity. Specifically, to test whether the cheating extension provides a more realistic model than the original UQM, the occurrence of cheating in a survey sample was tested and the general model fit was assessed and compared to that of the original UQM. Since we wanted the study to resemble a typical RRT application, we assessed a highly sensitive characteristic, that is, IPV.

Intimate partner violence. The term IPV incorporates physical, sexual, and psychological violence and controlling behavior toward a former or current intimate partner (World Health Organization 2012). The present study focused on physical IPV because this facet is easiest to explain to respondents in an online survey using concrete examples of behavior (here "shoving, slapping, hitting, kicking, or punching"). Other forms of violence, such as "humiliation" as an example of psychological violence, can be much more difficult to identify as violence for survey respondents. The lifetime prevalence of physical and sexual IPV against women in the European Union was estimated to be 22% in a survey by the European Agency for Fundamental Rights (2014) and the 12-month prevalence to be 4%. The Federal Criminal Police Office reported 141,792 cases of attempted or committed IPV in Germany in 2019 (Bundeskriminalamt 2020), that is, 17.3% of all reported violent crimes (including non-partner violence). Of these, 61.2% were actual bodily harm ("einfache Körperverletzung"). Of all IPV victims in the criminal statistic, 26,889 were male and 114,903 female. Numbers such as these contribute to the assumption that IPV is mainly perpetrated by men against women. However, there is an ongoing debate about gender (a)symmetry with respect to varying characteristics of both the specific type of violence investigated and the survey method (e.g., Archer 2000; Johnson 2006; Kimmel 2002). For instance, the lifetime prevalence of physical IPV victimization in the US was estimated to be 30.6% among women and 31.0% among men in the National Intimate Partner and Sexual Violence Survey (Smith et al. 2018), whereas the prevalence of severe physical violence victimization was estimated to be 21.4% among women and 14.9% among men. Generally speaking, estimates vary strongly between studies due to differences in the applied measures and samples (see, e.g., Devries et al. 2013; Garcia-Moreno et al. 2006; Kimmel 2002; Waltermaurer 2005).

As outlined at the beginning of this paper, IPV is a highly sensitive topic, that is, exactly the kind of topic for which RRTs were developed and thus suitable for the present validation study. Furthermore, several articles in scientific journals and the media reported rising numbers of IPV in the context of the impact of the spread of COVID-19, which was declared a pandemic by the World Health Organization in March 2020 (e.g., Bradbury-Jones and Isham 2020; Emundts 2020; Jarnecke and Flanagan 2020). The pandemic and the measures implemented to contain it are believed to foster factors associated with IPV, such as increased material worries or restricted possibilities to avoid the perpetrator and seek help (Usher et al. 2020). The rising numbers of criminal reports corroborate this argumentation, highlighting the relevance of investigating the dark figure of IPV. Thus, we applied the UQMC to estimate the prevalence of physical IPV during the first COVID-19 contact restrictions in spring and early summer 2020 in Germany to assess the model's empirical adequacy in a context that is relevant and representative of RRT applications.

Sensitivity Manipulation. To further test the UQMC, we employed an experimental manipulation of the sensitivity of the question: respondents were either queried about their role as a victim of IPV or as a perpetrator of IPV. As mentioned before, both roles are associated with stigma and are perceived as socially undesirable. However, being a perpetrator is even legally incriminating and has been shown to have an even stronger association with social desirability (Sugarman and Hotaling 1997). We therefore expected the question on the perpetration of IPV to be more sensitive than the question on the victimization of IPV. Consequently, we expected cheating to be more pronounced in the subsample queried about perpetration. We restricted our sample to participants who were, at the time of the investigation, in a romantic relationship with exactly one person. This way, the true proportion of perpetrators and victims should be equal in our sample. Assuming that differences in honesty of responding would be captured by the cheating parameter, any differences between estimates of the prevalence of honest carriers should be reflected in complementary differences in cheating estimates. Specifically, we expected that, if there was significant cheating, (a) it would be estimated to be higher in the subsample queried about perpetration and that (b) the prevalence of honest carriers would be estimated to be lower in the subsample queried about perpetration. If there was no significant cheating, the prevalence of honest carriers was expected not to differ between the subsamples.

Objective. To summarize, the aim of the present study was to assess the empirical validity of the recently devised UQMC (Reiber et al. 2020) in a survey on the prevalence of IPV. To this end, the fit of the UQMC was compared to that of its predecessor, the original UQM, and the occurrence of cheating was tested. Additionally, the queried IPV role was experimentally manipulated to investigate the differential influence of the question sensitivity on cheating.

Methods

Participants

Participants were recruited from the participant panel of the market research institute respondiAG with a target sample size of 4800. Quotas to approximate population proportions were installed for gender, age, and highest educational achievement. The target quotas are depicted in Table 1.

To participate, respondents had to declare that they were at least 18 years old and currently in a relationship with one person. Participants who indicated that they were younger than 18 years or that they were in no romantic relationship or in a romantic relationship of equal importance with more than one person, were screened out before answering the questionnaire. Participants who fell into an age, gender, or education level category for which the quota was already full were also screened out. To ensure data quality, an attention check question was included in the questionnaire and participants who failed to answer this question correctly were screened out before finishing the questionnaire. Section A of the online supplemental materials contains information on participant dropout per page.

Table 1. Sample Demographics.

	Target quota in percent	Sample size (percentage)		
		All	Victimization	Perpetration
Gender				
Male	49.60	1553 (47.13)	752 (46.48)	801 (47.76)
Female	50.40	1732 (52.56)	862 (53.28)	870 (51.88)
Diverse	—	10 (0.30)	4 (0.25)	6 (0.36)
Age				
[18,30)	18.50	640 (19.42)	328 (20.27)	312 (18.60)
[30,40)	16.50	510 (15.48)	251 (15.51)	259 (15.44)
[40,50)	17.00	558 (16.93)	280 (17.31)	278 (16.58)
[50,60)	21.20	689 (20.91)	342 (21.14)	347 (20.69)
[60,80)	26.80	898 (27.25)	417 (25.77)	481 (28.68)
Highest educational achievement*				
Less than primary school	—	12 (0.36)	6 (0.37)	6 (0.36)
Primary/lower secondary education	—	485 (14.72)	220 (13.60)	265 (15.80)
→ Low	32.90	497 (15.08)	226 (13.97)	271 (16.16)
Middle secondary education	—	1043 (31.65)	512 (31.64)	531 (31.66)
→ medium	32.40	1043 (31.65)	512 (31.64)	531 (31.66)
High secondary education	—	269 (8.16)	141 (8.71)	128 (7.63)
Apprenticeship	—	868 (26.34)	415 (25.65)	453 (27.01)
University degree	—	618 (18.76)	324 (20.02)	294 (17.53)
→ High	34.70	1755 (53.26)	880 (54.39)	875 (52.18)

Note. Displayed are the target quotas and acquired subsample sizes (percentages in parentheses) for all participants, and those in the victimization and perpetration conditions separately, for gender, age, and highest educational achievement. *Target quotas for highest educational achievement referred to the summary categories "low," "medium," and "high." The percentages in the raw categories and the summary categories each sum up to 100%. The raw categories of highest educational achievement are translated from the German categories "Kein Schulabschluss," "Grund-/Hauptschulabschluss," "Realschule (Mittlere Reife)," "Gymnasium (Abitur)," "Abgeschlossene Ausbildung," and "(Fach-)Hochschulabschluss" in this order.

The total data set consisted of 4804 participants who reached the last page of the questionnaire. Of these, 1326 participants who failed to answer the second of two training questions correctly, described in more detail later, and 183 participants with a mean response time of less than half of the median of each page (relative speed index, RSI; Leiner 2019) were excluded from the analysis.

After exclusion, the final sample consisted of 3295 participants with a mean age of 47.35 years ($SD = 15.44$); 1732 (52.56%) indicated that their gender was female and 10 (0.30%) indicated diverse. Age and gender categories approximated the target quotas very well as can be seen in Table 1. The target quotas for education could not be attained because too few participants in the lower education level groups were reached, as can be seen in the table. More specifically, people with a high education level were over-represented in the sample.

Of the final sample, 1618 (49.10%) answered the question on the victimization of IPV. They did not differ from those who answered the question on perpetration with respect to age, $t(3293) = 1.79$, $p = .073$, gender, $p = .623$, Fisher's exact test, or highest educational achievement, $\chi^2(5) = 7.22$, $p = .205$.

Design

The prevalence of IPV was assessed using one of two sensitive questions. Participants were either asked if they had experienced IPV (victimization role) or if they had committed IPV (perpetration role). They were randomly assigned to either of these role conditions, which only differed in the phrasing of the sensitive question itself. The sensitive question in the victimization condition read: "Have you, in your current relationship, since 23 March, been intentionally physically assaulted by your partner?"[5] In the perpetration condition it read: "Have you, in your current relationship, since 23 March, intentionally physically assaulted your partner?"[6] The date 23 March was chosen because it marks the date on which contact restrictions as a means of containing the spread of COVID-19 were officially announced in Germany. Participants were reminded of this context before answering the IPV question.

The sensitive question was presented within a UQMC design. Specifically, participants were instructed to think of a person whose birthday they knew and keep that birthday in mind. If the birthday was within a certain range of days in a month, they were asked to respond to a neutral question A and if it was in the remaining days of a month they were asked to respond to the sensitive question B. This range of days in a month determined the

Table 2. Condition Allocation.

	Victimization		Perpetration	
p \ q	0.25	0.75	0.25	0.75
1/3	420	407	405	425
2/3	402	389	426	421

Note. Number of participants per combination of the factors role, p and q.

randomization probability p to respond to the sensitive question B. It was varied between participants on two levels: 1st to 10th day ($p_1 = 2/3$) or 1st to 20th day ($p_2 = 1/3$). The sensitive question B was the question on IPV. The neutral question A asked whether the memorized birthday was in a certain range of months in the year. This question was also varied between participants on two levels to obtain two neutral prevalences q: January to September ($q_1 = .75$) or January to March ($q_2 = .25$). The birthday probabilities were reconciled with German birth rate records since 1950 (Statistisches Bundesamt 2020). Participants were instructed to mark their response ("Yes" or "No") to the question they were assigned to and reminded that only they knew which question they were responding to.

The combination of the factors role condition, p, and q resulted in eight groups, which are depicted in Table 2. This design was implemented to allow for testing the UQMC's assumptions and model fit.

Procedure

The questionnaire was created using the software SoSci Survey (Leiner 2020). The survey administration period lasted from 29 June 2020 to 15 July 2020.

On being directed to the survey via a link distributed by respondiAG, participants received general information about the study and were asked to confirm their informed consent. Only participants who did so were directed to the following pages of the questionnaire. First, they answered demographic questions on age, gender, highest educational achievement, and relationship status for screening and quota checks. Then they received detailed instructions on the UQMC together with an example involving the abuse of illicit drugs as a sensitive question. All participants completed two UQMC training questions. In each one, they received a vignette of a fictional person who is asked whether they took illicit drugs within a UQMC design. This design was, for each participant, exactly the same as in the question on IPV, with the difference that the sensitive question was on taking illicit drugs instead of IPV and that participants did not have to answer for themselves but for the fictional person. This way, it was possible to provide feedback on the response, because the correct answer was known from the vignette. In both cases the correct response was “Yes” but only once because the fictional person had taken illicit drugs. In the other case, the correct response was “Yes” as an answer to the neutral question although the person had not taken illicit drugs. This was meant to demonstrate anonymity protection. Participants who did not respond correctly to the second training question were later excluded from the analysis. After completing the training questions participants were informed about the definition of physical IPV and the relevant time period beginning 23 March, that is, during the first contact restrictions due to the COVID-19 pandemic in Germany. Participants then completed the IPV question within the UQMC design in one of the above described eight conditions. On the following two pages, participants were asked to provide information on their living conditions during the considered time period. A list of the questions is in Section B of the online supplemental materials. Among the additional questions was an attention check (“Which of the following cities is not in Germany?”—Berlin, Hamburg, Cologne, *London*, Frankfurt, Munich). Participants were expected to be able to answer this question if they were paying attention and, thus, participants who failed to answer correctly were excluded from the survey. On the last survey page, participants were provided helpline information for victims and perpetrators of IPV before being redirected to the site of respondiAG.

Data Analysis[7]

Data exclusion. Participants who responded incorrectly to the second of two UQMC training questions were excluded from the analysis. Because the training questions were very similar to the IPV question, failing to answer the second training question correctly was taken as an indicator for unreliable statements in the IPV question. We excluded 1326 participants, that is 27.60%, because they did not meet this criterion. This is a surprisingly high number, especially because only 866, that is 18.03%, failed to respond correctly to the first training question. Even though it is unclear why so many participants failed to answer the second training question correctly, this casts doubts on the validity of this criterion. However, the main results of this study are not strongly affected by inclusion or exclusion of the respective participants. Section C of the online supplemental materials contains the results of the analyses including participants who answered the second training question incorrectly. Differences between the two analyses are largely explainable by differences in power.

Additionally, 183 fast respondents with an RSI (Leiner 2019) above 2.00 were excluded from the main analysis. The RSI measures the participants' screen processing times relative to the screens' median processing times averaged across all screens. Therefore, an RSI above two indicates that the participant, on average, proceeded to the next screen twice as fast as the median of respondents. This can be used as an indicator for careless responding (Leiner 2019).

The participants' gender was included as a control variable in most analyses because of the inconclusive findings in the literature concerning its association with IPV. Whenever it was included, participants who indicated diverse gender were excluded from the analyses because the group was too small to be included as a separate factor level.

Parameter Estimation and Assessment of Model fit. All models were fitted by optimizing the G^2 statistic, which is a measure for the deviance of observed and model-predicted response frequencies, using the method by Nelder and Mead (1965) implemented in the function *optim* provided in the *stats* R-package. In the first step, the sample was split into four subsamples following from the combination of the two factors gender (excluding diverse gender) and role. For each of these subsamples, the IPV prevalence π_{UQM} in the UQM and the prevalence of honest carriers π_{UQMC} and the cheating prevalence γ in the UQMC were estimated separately. The model fit of both models in the four subsamples was assessed using the asymptotically χ^2-distributed G^2 statistic. Additionally, the overall fit of both models was assessed by summing the G^2 values from the subsamples and thereby making use of the additivity property of χ^2-distributed values. The fit of the UQM and UQMC was compared using G^2 difference tests and the Akaike and Bayesian information criterion (AIC and BIC), which set the model fit in relation to model complexity using penalty terms depending on the number of free parameters.

Analysis of Role Conditions. To test the influence of the role manipulation within the UQMC, a full logistic model including baseline cheating and honest carrier prevalence parameters as well as parameters for the factor Role (victimization vs. perpetration), the factor Gender (male vs. female), and interaction terms was fitted by optimizing the G^2 statistic:

$$\text{logit}(\gamma) = \gamma_0 + \gamma_1 \cdot \text{Gender} + \gamma_2 \cdot \text{Role} + \gamma_3 \cdot \text{Gender} \cdot \text{Role}, \qquad (5)$$

$$\text{logit}(\pi) = \pi_0 + \pi_1 \cdot \text{Gender} + \pi_2 \cdot \text{Role} + \pi_3 \cdot \text{Gender} \cdot \text{Role}. \qquad (6)$$

The factor Role was dummy coded with victimization as the reference category. Therefore, the effects of Role (γ_2 and π_2) can be interpreted as the difference in γ and π_{UQMC} between the victimization and perpetration conditions. Gender was included as a control variable and was effect coded in order that the mean effects of Role across the levels of Gender could be estimated. This full model was compared to restricted models using G^2 difference tests. Specifically, we successively restricted the interaction effects of Role and Gender π_3 and γ_3 and the main effects of Role π_2 and γ_2 to be equal to 0.[8] We compared each resulting model to the previous more complex model with respect to G^2, AIC, and BIC differences.

All analysis scripts and a preregistration of the study are on the Open Science Framework (OSF; https://osf.io/9bna3/).

Results

Estimation and Model fit

Table 3 depicts UQM and UQMC parameter estimates and their standard errors for the four subsamples following from the allocated role condition and participant gender. Due to the beginning of the contact restrictions on 23 March and the survey administration period from 29 June to 15 July, the estimates refer to 3 to 3.5 month IPV prevalences. The estimates for physical IPV without accounting for cheating, that is $\hat{\pi}_{\text{UQM}}$, lie between 7.43% and 11.56%. Applying the UQMC, in three of the four subsamples cheating is estimated to be close to 0 and, correspondingly, the prevalence estimates differ only slightly between the UQM and the UQMC. Only the IPV prevalence estimates among female participants queried about IPV victimization differ strongly between the models, with an honest carrier prevalence estimate of $\hat{\pi}_{\text{UQMC}} = 17.56\%$ and a cheating estimate of $\hat{\gamma} = 30.17\%$.

The latter outcome is consistent with the results of the model comparison in Table 4. The model fit of the UQMC is better than that of the UQM with respect to all model comparison criteria only for this subsample. Within this subsample, the UQM's G^2 statistic is highly significant, indicating insufficient model fit. The G^2 statistic indicates sufficient model fit for both models in all other subsamples.

Table 3. Model Estimates.

	N	$\hat{\pi}$ (SE)	$\hat{\gamma}$ (SE)	$\hat{\pi}+\hat{\gamma}$ (SE)
Victimization male				
UQMC	752	.14 (.03)	.08 (.06)	.23 (.08)
UQM	752	.12 (.02)	-	-
Victimization female				
UQMC	862	.18 (.03)	.30 (.06)	.48 (.08)
UQM	862	.07 (.02)	-	-
Perpetration male				
UQMC	801	.11 (.03)	.05 (.05)	.16 (.08)
UQM	801	.09 (.02)	-	-
Perpetration female				
UQMC	870	.08 (.03)	.00 (.06)	.08 (.08)
UQM	870	.08 (.02)	-	-

Note. Estimates for the prevalence of IPV (prevalence of honest carriers) π in the UQM (UQMC), the prevalence of cheating γ and the upper bound of the prevalence of IPV $\pi+\gamma$ in the UQMC along with their estimated standard errors in parentheses. UQM = unrelated question model; UQMC = unrelated question model—cheating extension; IPV = intimate partner violence.

The last two rows in Table 4 depict the overall model fit of the UQM and the UQMC using the G^2 sums over the subsamples. The UQMC's G^2 test indicates a reasonable model fit, whereas the UQM's G^2 value is significant, indicating insufficient model fit. The significant G^2 difference test supports the superiority of including the UQMC's cheating parameter.

When cheating is taken into account, the plausible range of estimates for the prevalence of IPV is indicated by the interval $[\hat{\pi}_{UQMC}; \hat{\pi}_{UQMC}+\hat{\gamma}]$. As it is typical for RRTs, the standard errors of both bounds are quite high, despite the large sample size. To accommodate this uncertainty, the confidence intervals (CIs) of the bounds need to be taken into account. Descriptively, the resulting range is highest in the subsample of female participants queried about victimization, with 95% CIs ranging from 12.15 to 62.88, and the lowest in the subsample of female participants queried about perpetration, with 95% CIs ranging from 2.70 to 22.60. The subsamples of male participants are very similar with respect to the UQMC's estimates with 95% CIs ranging from 8.66 to 39.06 in the victimization condition and from 5.83 to 31.32 in the perpetration condition. Note that these intervals are relatively wide because they incorporate the cheating estimates. Thus, they do not only indicate unsystematic uncertainty in the estimates but the systematic influence of a specific response style on the estimates. Therefore, despite

Table 4. Comparison of the UQM and the UQMC.

	Model fit				Model comparison						
	N	G^2	df	p	ΔG^2	df	p	AIC	ΔAIC	BIC	ΔBIC
Victimization male											
UQMC	752	5.61	2	.061				9.61		18.85	
UQM	752	6.61	3	.086	1.00	1	.317	8.61	−1.00	13.23	−5.62
Victimization female											
UQMC	862	3.90	2	.142				7.90		17.42	
UQM	862	19.10	3	<.001	15.19	1	<.001	21.10	13.19	25.86	8.44
Perpetration male											
UQMC	801	0.76	2	.683				4.76		14.13	
UQM	801	1.24	3	.745	0.47	1	.492	3.24	−1.53	7.92	−6.21
Perpetration female											
UQMC	870	0.37	2	.830				4.37		13.91	
UQM	870	0.37	3	.946	0.00	1	>.999	2.37	−2.00	7.14	−6.77
Overall											
UQMC	3285	10.64	8	.223							
UQM	3285	27.31	12	.007	16.67	4	.002				

Note. Model fit and model comparison of the UQMC and the UQM using G^2-tests and the AIC and BIC. UQM = unrelated question model; UQMC = unrelated question model—cheating extension; AIC=Akaike information criterion; BIC=Bayesian information criterion.

being wide these confidence intervals are indicative of relevant information, which is ignored by the original UQM and most RRTs as well as direct questioning techniques.

The estimates do not indicate a clear gender effect. An effect of the role condition is more apparent in the UQMC's estimates, especially in the subsample of female participants. Consequently, the effects of the role condition on the UQMC's estimates and their interactions with gender were tested in a logistic model. The main effects of gender were not specifically tested because there were no founded expectations due to the inconclusive findings on the gender differences in IPV.

Analysis of the Role Condition

The results of testing the effects of the role condition on the IPV prevalence and cheating are in Table 5. Each row of this table includes G^2, AIC, and BIC values of two models and their differences between both models. The parameter representing the respective effect is estimated freely in the "free" model and restricted to 0 in the "restricted" model. None of the fit statistics indicate that excluding an interaction term (π_3 or γ_3) for participant gender and role condition leads to a relevant decrease in model fit. Restricting the main effect of role condition on the honest carrier IPV prevalence π_2 to equal 0 does not lead to a decrease in model fit. Only the AIC favors the model allowing π_2 to differ from 0, and only if the restriction is introduced before the restriction on the main effect on cheating. The effect size estimate is $\hat{\pi}_2 = -0.50$ on the logit scale, which means that the odds of reporting IPV are estimated to be $e^{-0.50} = 0.61$ times as high for participants queried about perpetration as compared to victimization (i.e., taking the inverse, 1.64 times as high for participants queried about victimization). Restricting the main effect of role condition on the cheating prevalence γ_2 to equal 0 leads to a decrease in model fit according to the G^2 test, if it is restricted before the main effect on the IPV prevalence is restricted. However, this effect is not significant if multiple testing is taken into account using Holmes–Bonferroni corrections on the p-values. The AIC favors the unrestricted model both if the restriction of the main effect on cheating is introduced before and after the main effect on the honest carrier IPV prevalence is restricted. The effect size estimate is $\hat{\gamma}_2 = -2.31$ on the logit scale, which means that the odds of cheating are estimated to be $e^{-2.31} = 0.10$ times as high for participants queried about perpetration as compared to victimization (i.e., taking the inverse, 10.11 times as high for participants queried about victimization).

Table 5. Tests for Differences Between the Role Conditions.

	G^2						AIC			BIC		
	Free	Restr.	Δ	*df*	*p*	p_{corr}	Free	Restr.	Δ	Free	Restr.	Δ
Role × gender interaction effects												
π_3 (Restricted first)	10.68	11.58	0.89	1	.345	>.999	26.68	25.58	−1.11	75.46	68.26	−7.21
γ_3 (Restricted after π_3)	11.58	13.13	1.55	1	.213	.852	25.58	25.13	−0.45	68.26	61.71	−6.55
γ_3 (Restricted first)	10.68	11.13	0.44	1	.505	>.999	26.68	25.13	−1.56	75.46	67.81	−7.65
π_3 (Restricted after γ_3)	11.13	13.13	2.00	1	.158	.788	25.13	25.13	0.00	67.81	61.71	−6.10
Role effects												
π_2 (Restricted first)	13.13	15.71	2.59	1	.108	.647	25.13	25.71	0.59	61.71	56.20	−5.51
γ_2 (Restricted after π_2)	15.71	19.54	3.83	1	.050	.352	25.71	27.54	1.83	56.20	51.93	−4.27
γ_2 (Restricted first)	13.13	19.47	6.34	1	.012	.094	25.13	29.47	4.34	61.71	59.95	−1.76
π_2 (Restricted after γ_2)	19.47	19.54	0.08	1	.784	.784	29.47	27.54	−1.92	59.95	51.93	−8.02

Note. N = 3285. The interaction effects of Role and Gender and the main effects of Role on cheating γ and the honest carrier prevalence π are tested using G^2-difference tests and AIC and BIC differences. By row, single parameters are successively restricted to 0. Each resulting restricted ("Restr.") model is compared to a more complex model, in which the respective parameter is estimated freely ("Free"). p_{corr} refers to p-values adjusted for multiple testing using the Holmes–Bonferroni correction. The models for testing the interaction effects are derived from the full logistic model: $\text{logit}(\gamma) = \gamma_0 + \gamma_1 \cdot \text{Gender} + \gamma_2 \cdot \text{Role} + \gamma_3 \cdot \text{Gender} \cdot \text{Role}$; $\text{logit}(\pi) = \pi_0 + \pi_1 \cdot \text{Gender} + \pi_2 \cdot \text{Role} + \pi_3 \cdot \text{Gender} \cdot \text{Role}$. The models for testing the Role effects are derived from the main effects model: $\text{logit}(\gamma) = \gamma_0 + \gamma_1 \cdot \text{Gender} + \gamma_2 \cdot \text{Role}$; $\text{logit}(\pi) = \pi_0 + \pi_1 \cdot \text{Gender} + \pi_2 \cdot \text{Role}$. AIC=Akaike information criterion; BIC=Bayesian information criterion.

To summarize, we found (a) no interaction of role condition and participants' gender and (b) no significant main effect of role condition on the honest carrier prevalence or cheating. Specifically, contrary to our expectations, the prevalence of cheaters γ is not estimated to be higher in the group queried about perpetration than in the group queried about victimization. Additionally, numerically, the effect of role condition on γ, indicated by a small AIC difference, even goes in the opposite direction (i.e., γ is estimated to be higher in the group queried about victimization). Moreover, the effect of the role condition on the prevalence of honest carriers π_{UQMC} is not complementary to the effect on cheating. An effect of role condition on π_{UQMC} is only indicated by a small AIC difference and, numerically, the effect goes in the same direction as the effect on cheating (i.e., π_{UQMC} is estimated to be higher in the group queried about victimization).

From the effect size estimates, separate predictions for the UQMC parameters for both role conditions can be derived. For IPV victimization, the predicted honest carrier prevalence is $\pi_{\text{vict}} = 0.14$ and the predicted cheating prevalence is $\gamma_{\text{vict}} = 0.05$, both pooled across gender. For IPV perpetration, the predicted honest carrier prevalence is $\pi_{\text{perp}} = 0.09$ and the predicted cheating prevalence is $\gamma_{\text{perp}} = 0.01$, both pooled across gender.

Discussion

The current study was conducted to assess the validity of the UQMC (Reiber et al. 2020) in an applied setting. To that end, we investigated IPV in a UQMC design in an online survey. We assessed the fit of the UQMC and compared it to the fit of the UQM not accounting for cheating. Additionally, respondents were either queried about IPV victimization or perpetration because we expected this manipulation of question sensitivity to influence cheating. In light of the inconclusive prior findings on gender differences, we either conducted the analyses separately for male and female respondents or included gender as a control variable.

The overall model fit of the UQMC is acceptable and it is superior to the fit of the UQM, which cannot account for cheating. The biggest advantage is observable in the subsample of female participants queried about the victimization of IPV. In this group, the prevalence estimate of cheating is 30%. Thus, especially in this group of respondents, accounting for cheating allows responses to be more accurately described.

However, the effects of the IPV role condition manipulation are not as expected. Contrary to our expectations, cheating is estimated to be higher in the subsamples queried about victimization. Also, in the logistic model, the observed effect of the IPV role condition on cheating is not as expected and numerically even opposite to our expectations. Theoretically, this could mean that perpetrators are less reluctant to report their behavior than victims[9], but this is not in line with the previous literature, which showed that reporting of perpetration is stigmatized (e.g., Schröttle 2015; Franke et al. 2004) and associated even stronger with social desirability than victimization (Sugarman and Hotaling 1997). Moreover, if perpetrators were open to reporting their behavior and victims were reluctant to do so, the honest carrier prevalence of perpetration should be higher than that of victimization. Specifically, because the sample only consists of persons in an exclusive relationship, the true prevalence of IPV victimization and perpetration should be the same and, therefore, differences in the honest carrier prevalence should result from complementary differences in cheating. However, the prevalence estimate of honest carriers is not lower but numerically even higher in the subsample queried about victimization. In other words, the manipulation of the IPV role did not affect the parameters in different directions, indicating that the model's parameters are not complementary. This is not in line with the reasoning behind these parameters.

To summarize, although the general model fit is good (and therefore the model's assumptions seem to hold), we were not able to differentially manipulate the model parameters. In the following, three possible explanations for this inconsistency are outlined.

First, the inconsistency might be due to selective sampling. The expectations concerning the parameter relationship between role conditions are based on the assumption that the true prevalence of IPV perpetration and victimization in the assessed sample is the same. Yet, this does not necessarily have to be true. For example, IPV perpetrators could have decided to abort the survey more often than victims of IPV once they realized the content of the question. This would mean that the honest carrier and cheating prevalence are not complementary and explain how both can be higher in the victimization condition. However, the general dropout rates are not high enough to completely explain the inverted data pattern (see Section A of the online supplemental materials). Especially the dropout rates on the screen on which the queried role became apparent are very low (victimization: $N = 9$; perpetration: $N = 13$). A higher dropout among perpetrators before this point, which is independent of the role condition, could only have such a large impact on parameter estimates if the true prevalence of IPV was much higher than estimated in either of the conditions. Therefore, selective dropout is an unlikely explanation for the unexpected finding of model parameters being non-complementary. However, selective participation could still explain the results pattern, if IPV perpetrators were generally less likely to participate in surveys or be part of respondiAG's participant panel.

Second, there could be violations of the model assumptions which are mathematically consistent with the UQMC and thus not detectable merely by computational tests of model fit. For example, the UQMC inherited the assumption from the CDM that cheating is equally likely among respondents instructed to respond to the sensitive question and respondents instructed to respond to the neutral question. However, this need not be the case. Therefore, in the original presentation of the UQMC (Reiber et al. 2020), the possibility of *partial cheating* was outlined. In this framework, in addition to the two categories of respondents defined in the UQMC, that is, honest respondents and cheaters, there is a third category termed partial cheaters. This group of respondents would respond honestly if directed to answer the neutral question but give a self-protecting "No" response if directed to the sensitive question. Interestingly, following this logic, the estimation of the model parameters does not change. Specifically, the prevalence of cheating γ and the prevalence of honest carriers π_{UQMC} are estimated like in the UQMC that only allows for complete cheaters. The only thing that changes is the interpretation of the remainder category. In the UQMC, like in the CDM, the remainder category, $1 - \pi_{\mathrm{UQMC}} - \gamma$, is interpreted as the prevalence of honest non-carriers. However, in the framework of partial cheating, this remainder category also entails the partial cheaters.[10] In light of this idea, the results of the study could be interpreted differently: there could be partial cheaters in the subsample queried about perpetration, who cannot be detected by the model but their presence would explain the unexpected differences between estimates in the perpetration and victimization conditions.

Third, following a more substantive line of reasoning, differences in the individual interpretations of IPV by the participants could account for the data pattern. The UQMC is only capable of detecting deliberate cheating. Therefore, the hypotheses depend on the assumption that not only the true prevalence of IPV victimization and perpetration is equal, but also the perceived prevalence. However, it has been proposed that perpetrators and victims judge the same instance of IPV differently (see Follingstad and Rogers 2013). Specifically, the same situation can be reported as violent by the suspected victim but not by the suspected perpetrator. In such a case, a perpetrator not admitting to a violent act, which was perceived as violent by the victim, would not be a cheater in the sense of the UQMC. We decided to assess only physical IPV and provided specific examples in the instructions to minimize the likelihood of self-deception. However, it might still have played a role. This would explain why the lower estimated perpetration prevalence in the current study is not explainable by higher cheating.

Apart from these accounts there are limitations of the present study which might have influenced the results. On the one hand, it was crucial for the premises of our experimental manipulation that the participants were in a relationship with exactly one person. However, the relationship status in itself is a sensitive topic since in most social groups being in a committed relationship with one person still constitutes the norm. By only contrasting "being in a committed relationship with one person" to "not being in a relationship" or "being in more than one relationship of equal importance" in the respective screening question, we tried to minimize social desirability bias. However, it is still possible that some respondents chose to respond that they were in a committed relationship with one person although they were not. Nevertheless, this would only influence the results pattern if the likelihood of this response tendency differed strongly between perpetrators and victims of IPV.

On the other hand, there was a high proportion of respondents (27.60%) who did not respond correctly to the second training question. This calls into doubt that the instructions were sufficiently understood. Given that the probability to guess the correct response is 50%, this would in the worst case mean that another 30% did not fully understand the instructions. However, this seems unlikely because the rate of incorrect responses to the first training question was much lower (18.03%). Instead, since respondents did not know that an incorrect response to the second training question would lead to an exclusion of their data, they might not have paid attention to this question after correctly answering the first one. Therefore, the high proportion of incorrect responses could be not as much indicative of a major problem with understanding the instructions but rather that this preregistered exclusion criterion was sub-optimal. Still, this exclusion criterion did not substantively influence the results pattern either, as indicated by the additional analyses in Section C of the online supplemental materials.

Whether any of these accounts are actually responsible for the observed inconsistencies in the data pattern is, of course, not testable using the given data. However, the applied design enabled us to detect these inconsistencies and come up with plausible explanations. Surveys using direct questions or a simple RRT design are probably also affected by unexpected response patterns. In these cases, however, the inconsistencies do not become visible. Using the design applied in this study, we could, first, measure a specific type of instruction non-adherence, namely cheating, and the results indicate that especially among female participants queried about IPV victimization cheating is highly prevalent. Second, the unexpected effects of experimentally manipulating the queried IPV role indicated that additional factors influence the estimates. Although we can only speculate about these factors, detecting inconsistencies itself has important implications. It shows that the estimates need to be treated with caution—something that is arguably true for any survey on IPV.

All of the outlined explanations suggest that the IPV prevalence estimates in this study rather represent a lower limit to the true prevalence of IPV during the period of about three months starting with the initiation of the first contact restrictions due to the COVID-19 pandemic in Germany. However, even the lower limit estimates of about 10% are already very high for such a short time period. Therefore, although the exact numbers need to be interpreted carefully and, of course, a direct comparison to other time periods is not possible, the presented results are in line with the literature reporting high numbers of IPV in the context of the COVID-19 pandemic and the related containment measures (e.g., Steinert and Ebert 2020).

Conclusion

The purpose of the current study was to validate the UQMC, an extension of the UQM, to account for self-protecting responses. To that end, we conducted an online survey on IPV during the first contact restrictions due to the COVID-19 pandemic in Germany. The UQMC provides a reasonable account of the data, which is superior to that of the UQM. The data indicate a high prevalence of IPV, which is in line with the increase in IPV related to the COVID-19 pandemic reported by many other sources. Some unexpected data patterns emerged, highlighting once more the difficulty of investigating sensitive research topics and the need for treating the respective estimates with caution. Nevertheless, testable RRT designs accounting for instruction non-adherence can provide more insight into the response process and, thereby, a better understanding of sensitive research topics.

Authors' Note

Data and all analysis scripts are available on the Open Science Framework (https://osf.io/9bna3/).

Declaration of Conflicting Interest

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was funded by the Deutsche Forschungsgemeinschaft (DFG), grant 2277, Research Training Group "Statistical Modeling in Psychology" (SMiP).

ORCID iDs

Fabiola Reiber https://orcid.org/0000-0002-5654-4985
Donna Bryce https://orcid.org/0000-0001-8311-4457

Supplemental Material

Supplemental material for this article is available online.

Notes

1. Note that the interpretation of π_{CDM} differs from that of π_{UQM} as it denotes the combined probability of being an honest respondent and a carrier of the sensitive attribute.
2. The assumption that a higher probability to receive the sensitive question would lead to a lower proportion of honest admissions was tested by experimentally manipulating p in a UQM survey (Dietz et al. 2018). A difference in the expected direction was observed but it was not significant, possibly due to a lack of power.
3. This would require estimating a model with separate cheating parameters for each subsample. Such a model, however, is underdetermined.
4. Closed form equations for a UQMC implementation using only two subsamples are provided in Reiber et al. (2020). However, this approach does not allow for the assessment of model fit.
5. Adapted from Moshagen et al. (2012) and translated from German.
6. Adapted from Moshagen et al. (2012) and translated from German.
7. We used R (Version 4.0.5; R Core Team 2021) and the R-packages dplyr (Version 1.0.5; Wickham et al. 2021), forcats (Version 0.5.0; Wickham 2020), ggplot2 (Version 3.3.2; Wickham 2016), kableExtra (Version 1.1.0; Zhu 2019), papaja (Version 0.1.0.9997; Aust and Barth 2020), purrr (Version 0.3.4; Henry and Wickham 2020), readr (Version 1.3.1; Wickham et al. 2018), stringr (Version 1.4.0; Wickham 2019), tibble (Version 3.1.0; Müller and Wickham 2021), tidyr (Version 1.1.3; Wickham 2021), and tidyverse (Version 1.3.0; Wickham et al. 2019) for all our analyses.
8. To facilitate estimation we used the estimates from simpler models as starting values for the more complex models.

9. A reviewer suggested they could even be boastful instead of ashamed about their controlling behavior.
10. For an outline of the logic behind this conclusion see Reiber et al. (2020).

References

Archer, John. 2000. “Sex Differences in Aggression Between Heterosexual Partners: A Meta-Analytic Review.” *Psychological Bulletin* 126:651–680. doi: 10.1037/0033-2909.126.5.651.

Aust, Frederik and Marius Barth. 2020. *papaja: Create APA Manuscripts with R Markdown*.

Boruch, Robert F. 1971. “Assuring Confidentiality of Responses in Social Research: A Note on Strategies.” *The American Sociologist* 6:308–311.

Böckenholt, Ulf and Peter G. M. van der Heijden. 2007. “Item Randomized-Response Models for Measuring Noncompliance: Risk-return Perceptions, Social Influences, and Self-Protective Responses.” *Psychometrica* 72:245–262. doi: 10.1007/s11336-005-1495-y.

Bradbury-Jones, Caroline and Louise Isham. 2020. “The Pandemic Paradox: The Consequences of COVID-19 on Domestic Violence.” *Journal of Clinical Nursing* 29:2047–2049. doi: 10.1111/jocn.15296.

Bundeskriminalamt. 2020. *Partnerschaftsgewalt Kriminalstatistische Auswertung – Berichtsjahr 2019*.

Clark, S. J. and R. A. Desharnais. 1998. “Honest Answers to Embarrassing Questions: Detecting Cheating in the Randomized Response Model.” *Psychological Methods* 3:160–168.doi: 10.1037/1082-989x.3.2.160.

Devries, K. M., J. Y. T. Mak, C. García-Moreno, M. Petzold, J. C. Child, G. Falder, S. Lim, L. J. Bacchus, R. E. Engell, L. Rosenfeld, C. Pallitto, T. Vos, N. Abrahams, and C. H. Watts. 2013. “The Global Prevalence of Intimate Partner Violence Against Women.” *Science* 340:1527–1528. doi: 10.1126/science.1240937.

Dietz, P., A. Quermann, M. N. Maria van Poppel, H. Striegel, H. Schröter, R. Ulrich, and P. Simon. 2018. “Physical and Cognitive Doping in University Students Using the Unrelated Question Model (UQM): Assessing the Influence of the Probability of Receiving the Sensitive Question on Prevalence Estimation.” *PLoS One* 13: e0197270. doi: 10.1371/journal.pone.0197270.

Elbe, Anne-Marie and Werner Pitsch. 2018. “Doping Prevalence Among Danish Elite Athletes.” *Performance Enhancement and Health* 6:28–32. doi: 10.1016/j.peh.2018.01.001.

Ellsberg, Mary, Lori Heise, Rodolfo Peña, Sonia Agurto, and Anna Winkvist. 2001. “Researching Domestic Violence Against Women: Methodological and Ethical

Considerations." *Studies in Family Planning* 32:1–16. doi: 10.1111/j.1728-4465.2001.00001.x.

Emundts, Corinna. 2020. "Gewalt in Beziehungen: Die Zahlen sind schockierend." *Tagesschau*, November 10, 2020 (https://www.tagesschau.de/inland/beziehungsgewalt-bka-statistik-101.html)

European Agency for Fundamental Rights. 2014. *Violence Against Women: An EU-wide Survey*.

Follingstad, Diane R. and M. J. Rogers. 2013. "Validity Concerns in the Measurement of Women's and Men's Report of Intimate Partner Violence." *Sex Roles* 69:149–167. doi: 10.1007/s11199-013-0264-5.

Franke, B., D. Seifert, S. Anders, J. Schröer, and A. Heinemann. 2004. "Gewaltforschung zum Thema "häusliche Gewalt" aus kriminologischer Sicht." *Rechtsmedizin* 14:193–198. doi: 10.1007/s00194-004-0263-5.

Garcia-Moreno, Claudia, Henrica A. F. M. Jansen, Mary Ellsberg, Lori Heise, and Charlotte H. Watts. 2006. "Prevalence of Intimate Partner Violence: Findings From the WHO Multi-Country Study on Women's Health and Domestic Violence." *The Lancet* 368:1260–1269. doi: 10.1016/S0140-6736(06)69523-8.

Gracia, Enrique. 2004. "Unreported Cases of Domestic Violence Against Women: Towards an Epidemiology of Social Silence, Tolerance, and Inhibition." *Journal of Epidemiology and Community Health* 58:536–537. doi: 10.1136/jech.2003.019604.

Greenberg, Bernard G., Abdel-Latif A. Abul-Ela, Walt R. Simmons, and Daniel G. Horvitz. 1969. "The Unrelated Question Randomized Response Model: Theoretical Framework." *Journal of the American Statistical Association* 64:520–539. doi: 10.1080/01621459.1969.10500991.

Henry, Lionel and Hadley Wickham. 2020. *Purrr: Functional Programming Tools*.

Hoffmann, Adrian, Birk Diedenhofen, Bruno Verschuere, and Jochen Musch. 2015. "A Strong Validation of the Crosswise Model Using Experimentally-Induced Cheating Behavior." *Experimental Psychology* 62:403–414. doi: 10.1027/1618-3169/a000304.

Hoffmann, Adrian, Berenike Waubert De Puiseau, Alexander F. Schmidt, and Jochen Musch. 2017. "On the Comprehensibility and Perceived Privacy Protection of Indirect Questioning Techniques." *Behavior Research Methods* 49:1470–1483. doi: 10.3758/s13428-016-0804-3.

Holbrook, Allyson L. and Jon A. Krosnick. 2010. "Measuring Voter Turnout by Using the Randomized Response Technique - Evidence Calling Into Question the Methods's Validity." *Public Opinion Quarterly* 74:328–343. doi: 10.1093/poq/nfq012.

Horvitz, D. G., B. G. Greenberg, and J. R. Abernathy. 1976. "Randomized Response: A Data-Gathering Device for Sensitive Questions." *International Statistical Review* 44:181–196. doi: 10.2307/1403276.

Höglinger, Marc and Andreas Diekmann. 2017. "Uncovering a Blind Spot in Sensitive Question Research: False Positives Undermine the Crosswise-Model RRT." *Political Analysis* 25:131–137. doi: 10.1017/pan.2016.5.

Höglinger, Marc and Ben Jann. 2018. "More Is Not Always Better: An Experimental Individual-Level Validation of the Randomized Response Technique and the Crosswise Model." *PLoS One* 13:e0201770. doi: 10.1371/journal.pone.0201770.

Höglinger, Marc, Ben Jann, and Andreas Diekmann. 2016. "Sensitive Questions in Online Surveys: An Experimental Evaluation of Different Implementations of the Randomized Response Technique and the Crosswise Model." *Survey Research Methods* 10:171–187. doi: 10.18148/srm/2016.v10i3.6703.

Jarnecke, Amber M. and Julianne C. Flanagan. 2020. "Staying Safe During COVID-19: How a Pandemic Can Escalate Risk for Intimate Partner Violence and What Can Be Done to Provide Individuals with Resources and Support." *Psychological Trauma: Theory, Research, Practice, and Policy* 12:202–204. doi: 10.1037/tra0000688.

Johnson, Michael P. 2006. "Conflict and Control: Gender Symmetry and Asymmetry in Domestic Violence." *Violence Against Women* 12:1003–1118. doi: 10.1177/1077801206293328.

Kimmel, Michael S. 2002. "Gender Symmetry in Domestic Violence: A Substantive and Methodological Research Review." *Violence Against Women* 8:1332–1363. doi: 10.1177/107780102237407.

Leiner, D. J. 2020. SoSci Survey (Version 3.2.12).

Leiner, Dominik J. 2019. "Too Fast, Too Straight, Too Weird: Non-Reactive Indicators for Meaningless Data in Internet Surveys." *Survey Research Methods* 13:229–248. doi: 10.18148/srm/2019.v13i3.7403.

Lensvelt-Mulders, Gerty J. L. M., Joop J. Hox, Peter G. M. van der Heijden, and Cora J. M. Maas. 2005. "Meta-Analysis of Randomized Response Research: Thirty-five Years of Validation." *Sociological Methods and Research* 33:319–348. doi: 10.1177/0049124104268664.

Meisters, Julia, Adrian Hoffmann, and Jochen Musch. 2020. "Can Detailed Instructions and Comprehension Checks Increase the Validity of Crosswise Model Estimates?." *PLoS One* 15:e0235403. doi: 10.1371/journal.pone.0235403.

Moshagen, Morten, Jochen Musch, and Edgar Erdfelder. 2012. "A Stochastic Lie Detector." *Behavior Research Methods* 44:222–231. doi: 10.3758/s13428-011-0144-2.

Moshagen, Morten, Jochen Musch, Martin Ostapczuk, and Zengmei Zhao. 2010. "Reducing Socially Desirable Responses in Epidemiologic Surveys: An

Extension of the Randomized-Response Technique." *Epidemiology* 21:379–382. doi: 10.1097/EDE.0b013e3181d61dbc.

Müller, Kirill and Hadley Wickham. 2021. *Tibble: Simple Data Frames*.

Nelder, J. A. and R. Mead. 1965. "A Simplex Method for Function Minimization." *The Computer Journal* 7:308–313. doi: 10.1093/comjnl/7.4.308.

Ostapczuk, Martin, Jochen Musch, and and Morten Moshagen. 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: 10.1177/0962280210372843.

Pitsch, Werner, Eike Emrich, and Markus Klein. 2007. "Doping in Elite Sports in Germany: Results of a Www Survey." *European Journal for Sport and Society* 4:89–102. doi: 10.1080/16138171.2007.11687797.

R Core Team. 2021. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Reiber, Fabiola, Harrison Pope, and Rolf Ulrich. 2020. "Cheater Detection Using the Unrelated Question Model." *Sociological Methods and Research*. Advance online publication. doi: 10.1177/0049124120914919.

Schröter, Hannes, Beatrix Studzinski, Pavel Dietz, Rolf Ulrich, Heiko Striegel, and Perikles Simon. 2016. "A Comparison of the Cheater Detection and the Unrelated Question Models: A Randomized Response Survey on Physical and Cognitive Doping in Recreational Triathletes." *PLoS One* 11:e0155765. doi: 10. 1371/journal.pone.0155765.

Schröttle, Monika. 2015. "Sexuelle Gewalt und Gewalt in Paarbeziehungen." Pp. 81-210 in *Viktimisierungsbefragungen in Deutschland*, 47.1 ed., edited by R. Mischkowitz. Wiesbaden: Bundeskriminalamt.

Smith, S. G., X. Zhang, K. C. Basile, M. T. Merrick, J. Wang, M. Kresnow, and J. Chen. 2018. *National Intimate Partner and Sexual Violence Survey (NISVS): 2015 Data Brief-Update Release*.

Statistisches Bundesamt. 2020. Statistik der Geburten—Lebendgeborene: Deutschland, Monate, Geschlecht.

Steinert, Janina and Cara Ebert. 2020. *Gewalt an Frauen und Kindern in Deutschland während COVID-19-bedingten Ausgangsbeschränkungen: Zusammenfassung der Ergebnisse*.

Sugarman, David B. and Gerald T. Hotaling. 1997. "Intimate Violence and Social Desirability: A Meta-Analytic Review." *Journal of Interpersonal Violence* 12:275–290. doi: 10.1177/088626097012002008.

Tourangeau, Roger and Ting Yan. 2007. "Sensitive Questions in Surveys." *Psychological Bulletin* 133:859–883. doi: 10.1037/0033-2909.133.5.859.

Ulrich, Rolf, Harrison G. Pope, Léa Cléret, Andrea Petróczi, Tamás Nepusz, Jay Schaffer, Gen Kanayama, R. Dawn Comstock, and Perikles Simon. 2018.

"Doping in Two Elite Athletics Competitions Assessed by Randomized-Response Surveys." *Sports Medicine* 48:211–219. doi: 10.1007/s40279-017-0765-4.

Usher, Kim, Navjot Bhullar, Joanne Durkin, Naomi Gyamfi, and Debra Jackson. 2020. "Family Violence and COVID-19: Increased Vulnerability and Reduced Options for Support." *International Journal of Mental Health Nursing* 29:549–552. doi: 10.1111/inm.12735.

van der Heijden, Peter G. M., Ger van Gils, Jan Bouts, and Joop J. Hox. 2000. "A Comparison of Randomized Response, Computer-Assisted Self-Interview, and Face-to-Face Direct Questioning." *Sociological Methods & Research* 28:505–537. doi: 10.1177/0049124100028004005.

Waltermaurer, Eve. 2005. "Measuring Intimate Partner Violence (IPV): You May Only Get What You Ask For." *Journal of Interpersonal Violence* 20:501–506. doi: 10.1177/0886260504267760.

Warner, S. L. 1965. "Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias." *Journal of the American Statistical Association* 60:63–66. doi: 10.1080/01621459.1965.10480775.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag

Wickham, Hadley. 2019. *Stringr: Simple, Consistent Wrappers for Common String Operations*.

Wickham, Hadley. 2020. *Forcats: Tools for Working with Categorical Variables (Factors)*.

Wickham, Hadley. 2021. *Tidyr: Tidy Messy Data*.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, Alex Hayes, Lionel Henry, Jim Hester, Max Kuhn, Thomas Lin Pedersen, Evan Miller, Stephan Milton Bache, Kirill Müller, Jeroen Ooms, David Robinson, Dana Paige Seidel, Vitalie Spinu, Kohske Takahashi, Davis Vaughan, Claus Wilke, Kara Woo, and Hiroaki Yutani. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. doi: 10.21105/joss.01686.

Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2021. *Dplyr: A Grammar of Data Manipulation*.

Wickham, Hadley, Jim Hester, and Romain Francois. 2018. *Readr: Read Rectangular Text Data*.

Wimbush, James C. and Dan R. Dalton. 1997. "Base Rate for Employee Theft: Convergence of Multiple Methods." *Journal of Applied Psychology* 82:756–763. doi: 10.1037//0021-9010.82.5.756.

World Health Organization. 2012. *Intimate Partner Violence*.

Zhu, Hao. 2019. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*.

Author Biographies

Fabiola Reiber is a postdoctoral research assistant at the University of Tübingen, Germany. Her research focuses on psychological research methods, especially in the field of investigating sensitive topics in surveys.

Donna Bryce is a Junior Group Leader for Cognitive Development at the University of Tübingen, Germany. Her main research interests are cognitive control, metacognition and their development.

Rolf Ulrich is senior professor of cognitive psychology at the University of Tübingen, Germany. His main research focuses on mathematical psychology and cognition.