

Response Aggregation to Obtain Truthful Answers to Sensitive Questions: Estimating the Prevalence of Illegal Purchases of Prescription Drugs

Journal of Marketing Research

1-18

© The Author(s) 2023



Article reuse guidelines:
sagepub.com/journals-permissions
 DOI: 10.1177/00222437231205252
journals.sagepub.com/home/mrj



Marco Gregori , Martijn G. De Jong, and Rik Pieters

Abstract

This research proposes a simple, new technique to obtain truthful answers to sensitive, categorical questions. The paired response technique (PRT) asks participants to merely report the sum of the answers to two paired questions, one baseline and one sensitive, with the answers to each question known only to the participants. The technique then statistically infers the prevalence of the sensitive characteristic and its potential drivers from the association of the baseline question with other questions in the survey. Monte Carlo simulations demonstrate the performance of the PRT under varying conditions. A representative survey ($n = 4,649$) in the Netherlands about legal and illegal purchases of prescription drugs to enhance sexual performance reveals that 17.4% of the target population has purchased medication to enhance sexual performance at least once. In contrast, in a control group surveyed with direct questioning, only 5.1% admit having done so. The great majority of these individuals opt to purchase illegally. Two further empirical applications, respectively, in the United States and the United Kingdom, show that the PRT reduces cognitive and affective costs of survey participation compared with a state-of-the-art randomized response technique for categorical questions.

Keywords

sensitive questions, privacy, survey methods, health, Bayesian econometrics

Online supplement: <https://doi.org/10.1177/00222437231205252>

Surveys are commonly used to investigate consumers' behavior and opinions, but participants may provide socially desirable answers when the research concerns a sensitive topic (Steenkamp, De Jong, and Baumgartner 2010; Tourangeau and Yan 2007). Examples of such sensitive topics are pornography viewership; smoking during pregnancy (De Jong and Pieters 2019); sexual desires and behavior (De Jong, Pieters, and Fox 2010); software, music, and movie piracy (Kwan, So, and Tam 2010); and self-reported information on charity donations or sustainable consumption (White, Habib, and Hardisty 2019).

This research proposes a new technique to enhance truth-telling for sensitive survey questions. Our paired response technique (PRT) offers distinct advantages over existing indirect question techniques to promote honest responses. The PRT (1) can accommodate all types of categorical questions, which broadens the researcher's toolbox, (2) is easy to understand for survey participants, (3) offers unconditional privacy protection to participants, and (4) does not rely on randomization devices or questions mimicking randomization devices. Jointly, this facilitates participants' compliance with the instructions. We examine the

properties of the new technique theoretically and empirically and show that the PRT outperforms a state-of-the-art randomized response technique (RRT) for categorical sensitive questions (De Jong, Pieters, and Fox 2010).

Figure 1 provides an example of the data collection for the PRT. It pairs a sensitive and a nonsensitive item. The instructions simply ask participants to report whether the sum of the answers to these two paired categorical items corresponds to one of the provided response options. The response to the sensitive question is known only to the survey participant and masked from the analyst. Data analysis for the PRT then uses

Marco Gregori is Assistant Professor of Marketing, Warwick Business School, University of Warwick, UK (email: marco.gregori@wbs.ac.uk). Martijn G. De Jong is Professor of Marketing Research and Tinbergen Research Fellow, Erasmus School of Economics, Erasmus University Rotterdam, The Netherlands (email: mgdejong@ese.eur.nl). Rik Pieters is Professor of Marketing, Universidade Católica Portuguesa, Portugal, and Eminent Research Scholar, Tilburg School of Economics and Management, Tilburg University, The Netherlands (email: pieters@uvt.nl)

*Below are two questions: question A and question B. Each question has three possible answer options, that are numbered as 1, 2 or 3.
We do not need to know what your answer to each question is. For each question, you need to select the number corresponding to your answer, and then sum the two answers.
For instance, if your answer to question A is 2. I have obtained such medication, and always legally with a doctor's prescription and 3. I love parties with many people, the sum is 5.*

| Question a. I have obtained prescription medication to enhance my sexual performance or alleviate erectile dysfunction, such as Viagra or Cialis | Question b. Do you enjoy parties with many people? |
|---|---|
| 1. Never | 1. I dislike them |
| 2. I have obtained such medication, and <u>always legally</u> with a doctor's prescription | 2. I am neutral |
| 3. I have obtained such medication, but <u>at least once illegally</u> (for instance, from a friend, an unregistered online pharmacy, or some other websites) | 3. I love them |

Compute the sum of the numbers paired with your answers. Please indicate whether the sum corresponds with either of the following options:

- A. The sum of the two answers is "2 or 5"*
- B. The sum of the two answers is "3 or 6"*
- C. The sum of the two answers is "4"*

Figure 1. PRT: Question Pair, Each with Three Response Options.

the association between other questions in the questionnaire and the nonsensitive baseline item to estimate the likely response to the sensitive question. The statistical model to analyze the data is based on item response theory (IRT), with Bayesian estimation for inference on the prevalence and correlates of the sensitive characteristic of interest. To facilitate implementation of the PRT in marketing research, we have posted the annotated Matlab and Python code and the data of the three empirical applications at <https://osf.io/verfg/>.

Substantively, our research uses the PRT to provide new, large-scale empirical evidence on a topic of major societal importance, namely, illegal purchases of prescription drugs. Consumers increasingly rely on self-medication (Goldman and Jena 2011), a trend that was likely accelerated by the COVID-19 pandemic (United Nations Conference on Trade

and Development 2020). As a result, more consumers might purchase prescription drugs online illegally. Medication purchased illegally online may entail substantial health risks as online pharmacies without a physical address are shown to sell counterfeit medications with improper ingredients or doses (Gelatti et al. 2013; Organisation for Economic Co-operation and Development [OECD] and European Union Intellectual Property Office [EUIPO] 2020; Venhuis et al. 2013; World Health Organization 2017). An early survey indicated that 25% of U.K. general practitioners had already treated patients for adverse reactions from illegal consumption of medications (Nuffield Council on Bioethics 2009). As a case in point, illegal consumption of a counterfeit drug for erectile dysfunction caused an outbreak of hypoglycemia in Southeast Asia in 2008–2009, leading to multiple deaths (Kao et al. 2009).

Our empirical context concerns consumer acquisition of medication to enhance sexual performance. Despite the potential health risks and side effects, medications to enhance sexual performance appear to be purchased frequently online and are commonly seized by customs controls (Koenraad and Van de Ven 2018; OECD and EUIPO 2020; Venhuis et al. 2014). In the United States, India, China, and most of Europe, a prescription is required to legally obtain medication to enhance sexual performance and alleviate erectile dysfunction. In our first empirical application, based on a representative Dutch sample ($n = 4,649$), the great majority of these medications appear to be purchased illegally, most likely online (Venhuis et al. 2013). Using the PRT, we estimate that approximately 3.1% of the Dutch population has purchased medications to enhance sexual performance at least once, but always with a medical prescription, and approximately 13.7% of the Dutch population has done so exclusively without a medical prescription (that is, only illegally). In addition, we identify several potent covariates, for instance, that illegal purchases are especially prevalent among survey participants with lower socioeconomic status. In contrast, using direct questions to survey this sensitive issue dramatically underestimates the prevalence of illegal purchases.

Two additional empirical applications provide evidence that, compared with direct questioning (DQ) and the RRT, the PRT lowers the perceived and experienced psychological costs of participating in surveys on sensitive topics. Furthermore, in a different country (specifically, the United Kingdom), Study 3 replicates the outcome observed in Study 1. DQ leads to a lower estimated prevalence of buying medication to enhance sexual performance compared with the PRT. However, the difference is smaller than in Study 1, most likely due to the distinct regulatory regime in the United Kingdom. Sildenafil is covered by public health insurance and may be bought over the counter without a prescription (Bell et al. 2021). The RRT models do not produce reasonable estimates, unless noncompliance is explicitly accounted for, which requires additional assumptions and increases the required sample size.

The rest of the article is organized as follows. We first present an overview of existing indirect question techniques and the psychological costs that these techniques incur on survey participants. We point out the benefits of the PRT in minimizing these psychological costs for participants. Next, we describe the PRT, the response aggregation function, the statistical model, and the fundamental trade-off between privacy protection and statistical accuracy in designing the PRT. Then, three empirical applications provide evidence of the effectiveness of the PRT, and we conclude. Web Appendix B presents the results of Monte Carlo simulations that demonstrate the performance of the PRT in recovering the average prevalence and correlates of the sensitive characteristic of interest, while protecting individual privacy.

Throughout the manuscript, the term “sensitive characteristic” refers to any state, trait, or behavior (pattern) that survey participants may under- or overreport. The term “sensitive item or question” then refers to the target sensitive question asked to survey participants.

Indirect Question Techniques

A range of indirect question techniques have been proposed to reduce socially desirable response tendencies via truth-telling incentives, including RRTs (Lensvelt-Mulders et al. 2005) and item count techniques (ICTs; Blair, Coppock, and Moor 2020). Compared with DQ, these techniques ensure more truthful reporting by participants when dealing with sensitive private information (Rosenfeld, Imai, and Shapiro 2016; Tourangeau and Yan 2007). However, such indirect question techniques impose additional psychological (cognitive, affective, and behavioral) costs on survey participants, which may limit their usefulness in large-scale marketing and population surveys. The PRT aims to minimize these psychological costs.

RRTs are common indirect question techniques (Chaudhuri and Christofides 2013; Warner 1965). The privacy of the survey participant is protected by using a randomization device, such as a coin, a die, or a spinner (real or electronic). The randomization device adds stochastic noise to hide the survey participant’s true response to the sensitive question from the analyst. At the sample or subgroup level, it is straightforward to make inferences about the true response to the sensitive question (Blair, Imai, and Zhou 2015; Chaudhuri and Christofides 2013). Although commonly used, these methodologies have also come under scrutiny (Coutts and Jann 2011; Umesh and Peterson 1991; Wolter and Preisendörfer 2013). Randomization devices are difficult for many participants to understand and trust (Krumpal and Voss 2020). Furthermore, the response procedure may require some participants without the sensitive characteristic to endorse the question about it, thereby seemingly incriminating themselves. These participants might refuse to comply with the instructions (John et al. 2018). Finally, these techniques can be cumbersome for participants due to the (sometimes repeated) use of a physical randomizer (die, coin, spinner) or electronic randomization device.

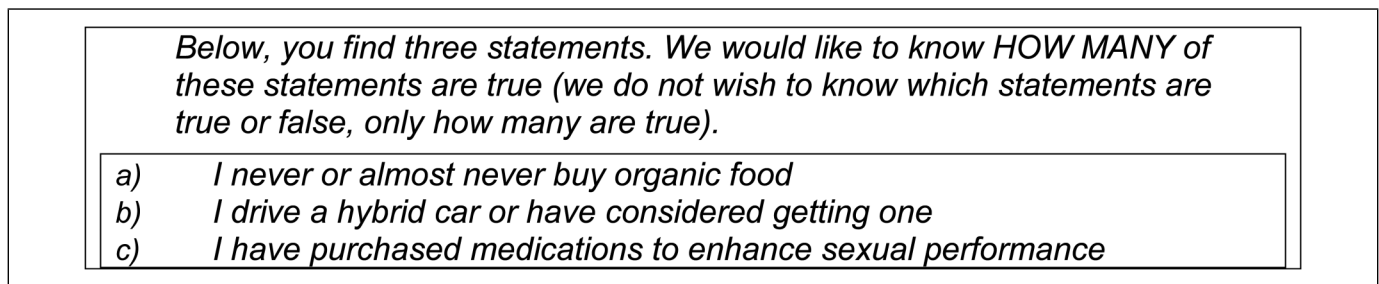
Response aggregation techniques offer an alternative to standard RRTs. Instead of relying on randomization devices that add stochastic noise, these techniques ask survey participants to provide a single response to a set of statements (i.e., to aggregate their responses). These techniques are typically easier to understand for participants and simpler to implement than RRTs are (Hoffman et al. 2017). Examples include the ICT (also called the “list experiment”), the crosswise model, or the triangular model. Table 1 provides an overview of popular existing response aggregation techniques. Response aggregation techniques are increasingly popular in political science, sociology, and related disciplines (Blair, Coppock, and Moor 2020; Blair and Imai 2012; Nepusz et al. 2014). An example in marketing is De Jong and Pieters (2019). Figure 2 illustrates the ICT with a binary target item. Table 2 presents an overview of the notation and acronyms.

The ICT is by far the most popular response aggregation technique, with more than 150 published applications before 2020 (Blair, Coppock, and Moor 2020). However, the ICT has important drawbacks. First, similar to other response

Table 1. Response Aggregation Techniques and Aggregation Functions.

| Technique Name | Response Aggregation Function | Example References | Example Instructions |
|-------------------------------------|--|---|--|
| M1. Item count technique (ICT) | $g(\{Z_1, \dots, Z_k\}) = Z_1 + \dots + Z_k = \sum Z_k$ | Blair and Imai (2012); De Jong and Pieters (2019) | HOW MANY of the next <number> statements are true? |
| M2. ICT (ceiling adjustment) | $g(Z_1, \dots, Z_k) = \begin{cases} \sum Z_k & \text{if } \sum Z_k \neq 0 \text{ or } K \\ 1(\{\sum Z_k = 0\} \vee \{\sum Z_k = K\}) & \text{other.} \end{cases}$ | Nepusz et al. (2014) | HOW MANY of the next <number> statements are true? |
| M3. Crosswise model | $g(\{Z_1, Z_2\}) = \begin{cases} \mathbb{1}(Z_1 = Z_2) \\ \mathbb{1}(Z_1 \neq Z_2) \end{cases}$ | Höglinger and Jann (2018); Yu, Tian, and Tang (2008) | Are your answers to the next two statements the SAME OR DIFFERENT? |
| M4. Triangular model | $g(\{Z_1, Z_2\}) = \begin{cases} \mathbb{1}(Z_1 + Z_2 > 0) \\ \mathbb{1}(\{Z_1 = 0, Z_2 = 0\}) \end{cases}$ | Hoffmann, Meisters, and Musch (2020); Yu, Tian, and Tang (2008) | Is your answer to AT LEAST ONE of the next two statements true? |
| M5. Paired response technique (PRT) | $g(\{Z_1, Z_2\}) = \begin{cases} \mathbb{1}(Z_1 + Z_2 = 2) \vee \mathbb{1}(Z_1 + Z_2 = L + 2) \\ \mathbb{1}(Z_1 + Z_2 = 3) \vee \mathbb{1}(Z_1 + Z_2 = L + 3) \\ \dots \\ \mathbb{1}(Z_1 + Z_2 = L + 1) \end{cases}$ | This research | What is the SUM of your responses to the next two statements? |

Notes: The PRT assumes both statements Z_1 and Z_2 have L different numbers of response options (usually $\mathbb{1} \in \{1, \dots, 5\}$).

**Figure 2.** ICT (List Experiment) with Three Items.

aggregation techniques, the ICT often results in substantial non-compliance (De Jong and Pieters 2019; Rosenfeld, Imai, and Shapiro 2016). Second, and in contrast to RRTs, the ICT is not suitable for categorical items (De Jong, Pieters, and Fox 2010). This is problematic because Likert and similar categorical items are the principal response format in marketing research and many of the social sciences in general (Bearden and Netemeyer 1999).

Criteria to Evaluate Indirect Question Techniques

Indirect question techniques such as RRTs, ICTs, and the PRT should statistically identify parameters of interest while minimizing the psychological costs (cognitive, affective, behavioral) of survey participation. In the following subsections, we describe each of these psychological costs of participation.

Cognitive costs. Cognitive costs concern the mental effort required to understand the text, follow the instructions, and implement one's choice when responding to survey questions

(Johnson and Payne 1985). To illustrate, participants must sum their responses to two or more items when using the PRT or ICT. With RRTs, participants require some basic understanding of probability, which is considered more demanding than summing items (Hoffmann et al. 2016; Wolter and Preisendörfer 2013). For instance, participants from a low socioeconomic background have shown more difficulty understanding instructions about randomization devices (Rosenfeld, Imai, and Shapiro 2016).

Affective costs. Affective costs concern the emotional consequences and hurdles of giving a certain sensitive answer. Affective costs are influenced by the design of the indirect question technique, such as the extent of privacy protection provided to survey participants (Krumpal and Voss 2020), whether participants have to give seemingly incriminating forced answers, and participants' perception of the baseline items.

Unconditional privacy protection implies that the participant's true response to the target item is concealed regardless of the answer to the indirect question technique. Failure to guarantee unconditional privacy protection implies that participants

Table 2. Terms and Notation.

| Notation | Description |
|-------------------------|--|
| PRT | Paired response technique |
| RRT | Randomized response technique |
| ICT | Item count technique |
| IRT | Item response theory |
| Y_i | Response to PRT for individual i |
| Z_i | Latent response to (nonsensitive) baseline item |
| U_i | Latent response to (sensitive) target item |
| $S_{ih}^{(b)}$ | Observed response to outside-the-pair item h |
| θ_i | Latent psychometric trait |
| α_h/α_{bas} | Discrimination parameter for outside-the-pair item (h)/ baseline item (bas) |
| γ_h/γ_{bas} | Threshold parameters for outside-the-pair item (h)/ baseline item (bas) |

might actually or seemingly incriminate themselves by endorsing a sensitive item (John et al. 2018; Wolter and Preisendörfer 2013). Figure 2 provides an example of the ICT, where unconditional privacy protection does not hold. In the ICT, if participants report that all items in the list are true (a “ceiling effect”), then they certainly have the sensitive characteristic. Likewise, if participants report that none of the items in the list is true (a “floor effect”), then they certainly do not have the sensitive characteristic. Thus, the ICT does not protect the privacy of participants unconditionally at the extremes. Participants who would truly have to respond at the extremes might then protect their privacy and avoid experiencing negative emotions such as shame and guilt by not adhering to the instructions and by reporting another response. This reduces the validity of the technique. Even if the sensitive item would not be affirmed, participants may report a different sum of affirmative responses to the list of items than the true sum to prevent any impression that they do (Blair and Imai 2012; Hoffmann, Meisters, and Musch 2020; Nepusz et al. 2014). Other common alternatives such as RRTs (John et al. 2018) also have challenges—for instance, when participants have to provide a forced answer that they feel uncomfortable with. This incurs affective costs for survey participants resulting in significant misreporting which may be difficult to detect and correct statistically.

For indirect question techniques relying on response aggregation, affective costs of participation also depend on the other, baseline items. These other items should be meaningful and consistent with the “conversational logic” of the questionnaire (Kuha and Jackson 2014; Schwarz 1995), with inconsistent items arousing suspicion and reducing trust among survey participants. Techniques such as the crosswise model, the triangular model, or some version of the ICT use baseline items such as “My father’s birthday falls on an uneven day” or “My house address begins with 6” (Höglinger and Jann 2018; Hoffmann, Meisters, and Musch 2020; Nepusz et al. 2014). These items aim at mimicking a randomization device (as their aggregate prevalence is known) and thus do not require between- or within-participation information for estimation. However, such items violate the conversational logic that the questions

in a survey are meaningful, and the resulting confusion and mistrust may increase accidental or systematic misreporting. Furthermore, such baseline items may make the sensitive question more salient, triggering defense mechanisms and leading to noncompliance (De Jong and Pieters 2019; Höglinger and Jann 2018; Kuha and Jackson 2014).

Behavioral costs. Behavioral costs relate to the practical implementation of a survey technique for participants, in terms of use of time, response devices, or other means. Behavioral costs are considerable with indirect question techniques that rely on randomized response. Such techniques require survey participants to obtain and/or use a physical randomizer, such as one or more dice (Blair, Imai, and Zhou 2015), or use a calculator to sum multiple responses in an ICT. With the PRT these and other costs are lower, as described next. Lower cognitive, affective, and behavioral costs facilitate the use of an indirect question technique in large-scale marketing and other surveys.

Paired Response Technique (PRT)

This section describes the PRT. We first formally define response aggregation, then discuss the data collection with the PRT and how it relates to the survey participation costs described previously. When using indirect question techniques, the analyst wants to draw inferences about participants’ responses to a sensitive item (survey question)—the target item—while protecting privacy. To achieve this with response aggregation, the target item is mixed with so-called baseline items into a set of items. Rather than directly asking for the response to the sensitive item, the analyst asks participants to provide an aggregate response for the item set as a whole, concealing potentially sensitive answers. Formally, given K items $1, \dots, k, \dots, K$, each item k is associated with a coded response Z_k (e.g., $Z_k = 1 [0]$ for true [false] responses; $Z_k = 1, 2, 3$ for agree, neutral, or disagree responses, respectively). Applying an aggregation function $g(\cdot)$ to the combination of coded responses to each item Z_1, \dots, Z_K outputs the response options of the indirect question technique (hereinafter denoted as Y). The aggregation function $g(\cdot)$ can be any mathematic operator, such as a sum, a product, or an indicator function. For instance, the ICT uses the sum $g(\cdot) = \sum(\cdot)$ as an aggregation function. Table 1 provides examples of response aggregation functions.

Data Collection

The data collection part of the PRT has the following defining features. First, participants indicate the sum of their responses to a pair of items: a target item and a baseline item. It shares the summing of responses with the ICT, and the use of a pair of items with the crosswise model. In the PRT, the baseline item comes from a validated multi-item scale. Other items from the multi-item scale, labeled “outside-the-pair items,” are asked elsewhere—for instance before or after the PRT in the current survey, or in a previous or follow-up survey. The statistical analysis of the PRT uses information from the outside-the-pair items to predict

the prevalence of the baseline item, from which the prevalence of the sensitive target item can be inferred. Hence, within-participant information is sufficient to probabilistically infer their likely response to the sensitive item.¹

As an example, suppose that the baseline item used in the PRT is “I enjoy parties with many people,” an item from the extraversion subscale of the Big Five personality inventory (Goldberg 1992). Additionally, assume that the following items are administered elsewhere (in the survey or in another survey) and answered directly: “I feel comfortable around people,” “I don’t talk a lot,” and “I keep in the background.” Since these items are all correlated, the outside-the-pair items predict the baseline item. These correlations are, however, not exceedingly large (e.g., 10%–30%), thereby guaranteeing privacy protection to the participant.

Response Aggregation Function

To illustrate data collection and analysis for polytomous items, we use an example question from our empirical application on legally and illegally obtaining prescription drugs, shown in Figure 1. The target item U has three possible responses—hence $U \in \{1, 2, 3\}$. Under the aggregation function that we propose, the target item U is paired with a baseline item Z that also has three response options. The PRT uses the following response aggregation function, resulting in an answer Y :

$$Y = g_{\text{PRT}}(\{Z, U\}) = \begin{cases} \text{The sum is 2 or 5: } \mathbb{I}(\{U + Z = 2\} \vee \{U + Z = 5\}) \\ \text{The sum is 3 or 6: } \mathbb{I}(\{U + Z = 3\} \vee \{U + Z = 6\}). \\ \text{The sum is 4: } \mathbb{I}(\{U + Z = 4\}) \end{cases} \quad (1)$$

Or, equivalently:

$$Y = g_{\text{PRT}}(\{Z, U\}) = \begin{cases} \mathbb{I}(\{U = 1, Z = 1\} \vee \{U = 2, Z = 3\} \vee \{U = 3, Z = 2\}) \\ \mathbb{I}(\{U = 1, Z = 2\} \vee \{U = 2, Z = 1\} \vee \{U = 3, Z = 3\}). \\ \mathbb{I}(\{U = 1, Z = 3\} \vee \{U = 3, Z = 1\} \vee \{U = 2, Z = 2\}) \end{cases} \quad (2)$$

To appreciate how the aggregation provides unconditional privacy protection, note that every answer option is possible for the target item U : in Equations 1 and 2, each row includes each possible answer option (1, 2, or 3) to the target item. Thus, the analyst does not know the exact answer to the target item U . In contrast to other categorical techniques, survey participants do not provide seemingly incriminating responses. However, knowing the answer to the baseline item

makes retrieving the answer to the target item straightforward: note that each row in Equation 1 includes any possible answer to the baseline item Z only once. For instance, if the answer to the PRT is “the sum is 2 or 5” (first row of Equation 1), there are three possible response combinations, namely (1, 1), (2, 3), and (3, 2). If we knew that the answer to the baseline item is 2, then the answer to the target item must be 3, since the only possible combination of responses in Equation 1 is $\{U = 3, Z = 2\}$.

Implementing the model is straightforward by instructing participants to sum the two answers as shown in Figure 1. The PRT thus improves on other indirect question techniques in several important ways. First, the PRT reduces the cognitive and behavioral costs for participants by using “response aggregation” to mask the responses to sensitive questions, rather than by using randomization devices as the RRT does. The instructions are straightforward, and the task of summing up two small integers rests on simple cognitive operations (Johnson and Payne 1985), which facilitates participants’ comprehension and compliance. The PRT simplifies data collection as compared with the ICT, because only two items—including the target item—are required. Second, the PRT is suitable for all common categorical response specifications (binary, ordered, or multinomial responses). Third, the PRT reduces affective costs by ensuring *unconditional* privacy protection to survey participants. Participants can intuitively verify this because a low number is paired with a high number (e.g., “2 or 6”). Each sum option is not readily associated with a specific sensitive response, which reduces affective costs. This enhances truthful reporting and improves on techniques such as RRTs or the ICT. Table 3 provides the sum options satisfying the requirement of unconditional privacy protection given a target item with any number of options.

Unconditional privacy protection offered by the PRT has a price—that is, it results in a large posterior variance of the target item estimates. The statistical model introduced in the next section aims at reducing this variance while still ensuring unconditional privacy protection.

Statistical Model

We develop a formal statistical model for estimating the prevalence and drivers of categorical target items measured through the PRT. Participant i responds to H ordered baseline “outside-the-pair” items elsewhere in the questionnaire or in a separate survey. The baseline item inside and the outside-the-pair baseline items jointly reflect some latent individual trait, denoted θ_i for participant i . A “trait” denotes some underlying latent construct, such as personality traits, values, or norms (Fox 2010). The response to the outside-the-pair items $S_{ih}^{(\theta)}$ (for item $h \in \{1, \dots, H\}$) is modeled following a graded response model (Samejima 1969):

$$P(S_{ih}^{(\theta)} = c | \theta_i, \alpha_h, \gamma_h) = \Phi(\alpha_h \theta_i - \gamma_{h,c-1}) - \Phi(\alpha_h \theta_i - \gamma_{h,c}), \quad (3)$$

where $\Phi(\cdot)$ denotes the normal cumulative distribution function. The model specifies the conditional probability of response $S_{ih}^{(\theta)}$

¹ An alternative approach makes use of an extra sample/condition in which the baseline item from the PRT is asked directly. The between-participant information from this DQ sample/condition can then be used to estimate the prevalence of endorsing the sensitive item in the PRT group more efficiently, with the added financial costs of surveying a larger sample. See the section on “Identification and Efficiency.”

Table 3. Response Options and Aggregation Functions for the PRT.

| No. of Response Options of Items Z and U | Response Aggregation Function: Answer Options of Response Y to the PRT |
|--|---|
| 2 (1, 2) | $\mathbb{I}(\{U + Z = 2\} \vee \{U + Z = 4\}), \mathbb{I}(\{U + Z = 3\})$ |
| 3 (1, 2, 3) | $\mathbb{I}(\{U + Z = 2\} \vee \{U + Z = 5\}),$ $\mathbb{I}(\{U + Z = 3\} \vee \{U + Z = 6\}), \mathbb{I}(\{U + Z = 4\})$ |
| 4 (1, ..., 4) | $\mathbb{I}(\{U + Z = 2\} \vee \{U + Z = 6\}),$ $\mathbb{I}(\{U + Z = 3\} \vee \{U + Z = 7\}),$ $\mathbb{I}(\{U + Z = 4\} \vee \{U + Z = 8\}), \mathbb{I}(\{U + Z = 5\})$ |
| L (1, ..., L) | $\mathbb{I}(\{U + Z = 2\} \vee \{U + Z = L + 2\}),$ $\mathbb{I}(\{U + Z = 3\} \vee \{U + Z = L + 3\}),$ $\mathbb{I}(\{U + Z = 4\} \vee \{U + Z = L + 4\}), \dots,$ $\mathbb{I}(\{U + Z = L + 1\})$ |

Notes: Z = baseline item; U = target item. Each level for U and Z is a coded response (e.g., for the case “2 (1, 2)”: 1 = Yes, 2 = No; for the case “3 (1, 2, 3)”: 1 = Agree, 2 = Neither agree nor disagree, 3 = Agree). $\mathbb{I}(\cdot)$ is an indicator function.

in category $c \in \{1, \dots, C\}$. The specification in Equation 3 maps a latent trait parameter θ_i of participant i and question-specific item parameters α_h and γ_h onto an observed pattern of responses. The term discrimination parameter is applied for the parameter α_h and the term threshold or difficulty parameter is applied for γ_h .

The latent response to the sensitive target item for participant i is denoted as U_i , with J possible response options $\{1, \dots, j, \dots, J\}$. The response to the baseline item is denoted Z_i , with $K = J$ possible response options $\{1, \dots, k, \dots, K\}$. The probability of observing a graded response to the baseline item Z_i , similarly to the outside-the-pair items, depends on the latent trait θ_i and item-specific parameters α_{bas} and γ_{bas} :

$$P(Z_i = k | \theta_i, \alpha_{\text{bas}}, \gamma_{\text{bas}}) = \Phi(\alpha_{\text{bas}}\theta_i - \gamma_{\text{bas},k-1}) - \Phi(\alpha_{\text{bas}}\theta_i - \gamma_{\text{bas},k}). \quad (4)$$

The response to the target item U_i can depend on individual-level covariates \mathbf{X}_i and can be modeled with any categorical response specification (binary, ordered, or multinomial).

The response to the PRT Y_i with possible response options $\{1, \dots, l, \dots, L\}$ is observed and depends on the combination of responses to the baseline item Z_i and to the target item U_i . Assuming conditional independence of the baseline and target items Z_i and U_i , and conditioning on the remaining model parameters, the response probability to a certain combination of baseline and target item is

$$P(U_i = j, Z_i = k | \mathbf{X}_i, \boldsymbol{\beta}, \theta_i, \alpha_{\text{bas}}, \gamma_{\text{bas}}) = P(U_i = j | \mathbf{X}_i, \boldsymbol{\beta})P(Z_i = k | \theta_i, \alpha_{\text{bas}}, \gamma_{\text{bas}}). \quad (5)$$

Table 3 provides the final link between each combination of responses Z_i and U_i and response Y_i . Then, the likelihood of observing a certain paired response Y_i and outside-the-pair

items $S_{ih}^{(0)}$ with covariates \mathbf{X}_i is

$$L(Y_i, S_i^{(0)} | \mathbf{X}_i) = \int_{\Theta} \left[\prod_{h=1}^H \prod_{c=1}^C P(S_{ih}^{(0)} = c | \theta_i, \alpha_h, \gamma_h) \right]^{1[X_{ih}^{(0)}=c]} \times \prod_{l=1}^L P(Y_i = l | \mathbf{X}_i, \boldsymbol{\beta}, \theta_i, \alpha_{\text{bas}}, \gamma_{\text{bas}})^{1[Y_i=l]} f(\theta_i | \mu, \sigma) d\theta. \quad (6)$$

Estimation is carried out by using a Bayesian Markov chain Monte Carlo (MCMC) sampler, combining Gibbs and Metropolis–Hastings steps, to efficiently evaluate the multidimensional integrals of the likelihood. Details about the sampler appear in the Technical Appendix.

Identification and Efficiency

To identify the latent responses U_i , the following two conditions need to be satisfied. First, the number of response options presented in the PRT must be at least as large as the number of response options for the target item. Second, information on the unobserved baseline response Z_i is needed. The key challenge of the PRT is that the response to the baseline question Z_i for each individual is unknown to the analyst. However, under the proposed model, information about the latent responses Z_i is obtained from the latent trait parameter θ_i . Thus, to identify the parameters α_{bas} and γ_{bas} , one leverages either information from the functional form or information from external sources, such as information from an additional sample receiving the baseline items using DQ.² We discuss each identification condition in turn.

Condition 1. The first identification condition of the PRT is as follows: given a target and a single baseline item each with K response options, the prevalence of the target item is identified if the number of response options to the PRT is equal to or larger than K . For instance, in the PRT in Figure 1, the target item has three response options (1, 2, and 3) and the PRT has three possible responses (A, B, and C); hence, the model is just identified. Web Appendix A provides further details and examples illustrating response identification.

Condition 2. Upon satisfying condition 1, identification of the PRT relies on either functional form restrictions or between-participants information from a DQ sample. Specific priors for the baseline item parameters α_{bas} and γ_{bas} correspond to a functional form restriction that ensures model identification. An example is the noninformative priors specified in De Jong and Pieters (2019). However, noninformative priors result in large posterior variance of the estimated target item.

² We thank an anonymous reviewer for pointing this out and for additional suggestions for this section.

An alternative approach is to administer the baseline item Z directly to a separate sample of participants. One could incorporate the responses of this separate sample in the likelihood when estimating Equations 3 and 4. The approach rests on the assumption of measurement invariance, that is, the item parameters $\alpha_{\text{bas}} = \alpha_{\text{bas}}^{\text{direct}}$ and $\gamma_{\text{bas}} = \gamma_{\text{bas}}^{\text{direct}}$, where the superscript “direct” indicates the item parameters of the baseline item in the separate sample. In other words, the baseline item is assumed to behave similarly in the PRT sample as in the separate sample. An undesirable effect of this approach is that information from the separate sample may then dominate α_{bas} and γ_{bas} . De Jong and Pieters (2019) show that survey participants might evaluate the baseline item differently when asked directly versus when asked as part of an ICT, where multiple questions are responded to indirectly rather a single question directly. Therefore, we opt for a different approach which allows for possible measurement noninvariance. This approach calibrates a prior for the baseline parameters α_{bas} and γ_{bas} based on the direct responses from the separate sample. This reduces the variance of the target item but still allows the baseline item to behave differently in the PRT versus the separate sample.

To complete the model identification, we impose the standard IRT restriction by setting the prior of the latent trait parameter θ_i equal to $N(0, 1)$ (Fox 2010).

ϵ -Differential Privacy

ϵ -differential privacy mathematically quantifies the privacy loss versus statistical accuracy (also known as “utility”) for an indirect question technique. If the so-called parameter ϵ is smaller, more privacy is guaranteed to a survey participant, but statistical estimates of interest, such as the prevalence of the target item, become less accurate (Dwork and Roth 2014; Schneider et al. 2017). In the case of the PRT, the formal definition is as follows. The PRT is ϵ_{PRT} -differentially private if, given the response aggregation function $g_{\text{PRT}}(\cdot)$ of the PRT, for any possible answer to the target item U_i , say j_a and j_b (with nonzero probability), and for any answer to the PRT Y_i , the following property holds:

$$\frac{\Pr(g_{\text{PRT}}(U_i = j_a, Z_i) = Y_i)}{\Pr(g_{\text{PRT}}(U_i = j_b, Z_i) = Y_i)} \leq \exp(\epsilon_{\text{PRT}}). \quad (7)$$

Web Appendix A shows that for the PRT:

$$\epsilon_{\text{PRT}} = \ln \frac{\max_k \Pr(Z_i = k)}{\min_k \Pr(Z_i = k)}. \quad (8)$$

The definition of ϵ_{PRT} has a straightforward interpretation: ϵ_{PRT} is larger and less privacy is provided to a survey participant if the baseline item is imbalanced, that is, the probability of selecting a certain answer to the baseline item $\Pr(Z_i = k)$ is rare or common. For instance, suppose that in a three-category PRT, $\max_k \Pr(Z_i = k) = 60\%$ and $\min_k \Pr(Z_i = k) = 10\%$. Then, the PRT is $\ln(6)$ -differentially private. Conversely, suppose that each answer to the baseline item is equally likely (that is $\Pr(Z_i = k) = 1/3$). Then, $\epsilon_{\text{PRT}} = 0$ and nothing is known

about the sensitive characteristic of the survey participant. In conclusion, ϵ_{PRT} quantifies the privacy loss for a survey participant. The value of ϵ_{PRT} in the sample depends on the discrimination parameter α_{bas} and the threshold parameters γ_{bas} of the baseline item, which determine the average probability of selecting a certain answer to the baseline item $\Pr(Z_i = k)$. Web Appendix B examines this relationship using Monte Carlo simulations.

Empirical Applications

Three empirical applications provide evidence for the effectiveness of the PRT in obtaining truthful responses to sensitive questions compared with DQ techniques and in reducing the psychological costs of participating in surveys on sensitive topics compared with a state-of-the-art RRT. The studies focus on illegal purchases of prescription drugs. Previous studies on this important societal issue rely on small, nonrepresentative samples using DQ techniques, which prevents firm conclusions about the prevalence and correlates of such purchases (e.g., Chan, Wood, and Dargan 2015; Koenraadt and Van de Ven 2018). Direct questions about acquisition and use of sexual enhancement drugs are likely to lead to biased estimates. Our empirical applications of the PRT aim to help close this knowledge gap. In particular, Study 1 is the first large-scale study to estimate the prevalence and correlates of purchases of prescription drugs to enhance sexual performance. Table 4 provides an overview of the studies.

Study 1: Purchases of Illegal Medication

Data for the first study were collected from an online panel administered in the Netherlands (LISS panel: <https://www.centerdata.nl/>). The sample is representative of the Dutch population over 16 years old. We have a two-wave longitudinal design. Responses to the target sensitive items are collected in the main wave, and responses to outside-the-pair items are collected in another wave.

Outside-the-pair items. Members of the LISS panel annually complete a questionnaire with a personality inventory. We had access to data from the May 2019 wave. All 6,218 panel members were sampled and 5,075 provided complete questionnaires (81.6%). This data set provides information about the “outside-the-pair” baseline items. We used four items from the Big Five International Personality Item Pool extraversion and conscientiousness subscales (Goldberg 1992), with five-point Likert response categories with endpoints at “Strongly disagree” and “Strongly agree.”

Due to the low incidence of some response options, we collapsed the “Strongly disagree” and “Disagree” categories into a single category. Furthermore, we recovered 182 additional complete questionnaires from the May 2020 wave, under the standard assumption of trait stability across a period of a few months (Gnambs 2014). Web Appendix C shows that the results are robust to excluding this additional sample.

Table 4. Overview of Empirical Applications.

| | Purpose | Sample | Outside-the-Pair Items | Prior Calibration | Techniques |
|---------|---|---|------------------------|---------------------------|--------------|
| Study 1 | Direct application of PRT | The Netherlands (representative; n = 4,649) | 4 items, separate wave | Based on DQ group | PRT, DQ |
| Study 2 | Comparison of cognitive and affective costs of PRT and RRT | The United States (nonrepresentative; n = 200) | N.A. | N.A. | PRT, RRT |
| Study 3 | Direct application of PRT and RRT, and comparison of “costs” of methods | The United Kingdom (nonrepresentative; n = 2,999) | 2 items, same wave | Based on PRT and DQ group | PRT, RRT, DQ |

Notes: N.A. = not applicable.

Paired response item. The main data wave (September 2019) included two PRT questions, one with three response options and one with four response options. The first PRT question was a placebo check used to assess whether the PRT provides valid responses in nonsensitive contexts. The target item for the second PRT was a sensitive question with four answer options about purchases of prescription drugs to enhance sexual performance, with the exact items shown in Table 7.

In the main wave, 5,946 panel members were included and 4,773 provided complete questionnaires (80.3%). After merging the three data sets (main wave and separate waves for outside-the-pair items), we have complete data for 4,649 participants. Participants in the main wave were randomly assigned to one of two conditions. Participants in the PRT condition answered the sensitive items with the PRT (80% of the total sample, n=3,687). Participants in the DQ condition received the PRT items, including the baseline item, using direct questions (20% of participants, n=962). Table 5 has descriptive statistics for both conditions. Participants in the two conditions do not differ in terms of observed covariates, which supports success of random assignment.

Covariates. As covariates to predict illegal purchases of medicine, we include two dummy variables for income (whether the participant earns a net monthly income lower than \$1,200 or a net monthly income higher than \$3,600) and information on education level (elementary education, vocational or technical education, university-level education). We include these measures based on initial evidence that individuals with lower socioeconomic status are more likely to make these purchases (Ivanitskaya et al. 2010; Napolitano et al. 2013). Furthermore, information was also available on the participant’s gender, age, relationship status (whether they currently have a partner or not), and psychometric measures of subjective well-being (Diener et al. 1985).

Data analysis was done in Matlab (version R2021a) using 50,000 draws for burn-in and 100,000 draws for computing posterior statistics. Code to replicate the analyses is provided at <https://osf.io/verfg/>.

Placebo test. The first PRT administered a nonsensitive target item with three answer options (“My birthday is between the

1st and the 20th of a month,” “My birthday is between the 21st and the 25th of a month,” and “My birthday is between the 26th and the 31st of a month”), which was then paired with a baseline question. If, as we expect, the PRT does not itself introduce bias but only enables participants to answer the sensitive target questions more truthfully, we should observe the following results. First, we should find no significant difference in prevalence of the behavior in the nonsensitive target item when obtained by PRT versus by DQ. Second, we should find a significantly higher prevalence of the behavior in the sensitive target item when it is obtained by PRT than by DQ.

The baseline item in the placebo test came from the Big Five scale of consciousness and was “if a project becomes too difficult, I don’t give up/sometimes I give up/I often give up.” For difference testing, we report the Bayesian p -value p_{Bayes} (i.e., the percentage of draws where a parameter estimated with the PRT is larger than the corresponding parameter estimated with DQ). Due to the low accuracy of the posterior estimates, the prior calibration procedure (mentioned in the identification section) is implemented using the baseline item administered directly in the DQ group. A prior sensitivity analysis is reported in Web Appendix C.

Table 6 has mean posterior estimates obtained for the PRT, DQ, and their difference. Web Appendix C has posterior estimates for the baseline item. Table 6 indicates that there is no evidence of possible overestimation or underestimation of response to the nonsensitive item due to using the PRT: none of the differences between PRT and DQ reaches statistical significance. As shown in Web Appendix C, this is true for various prior specifications. The PRT has larger variance compared with DQ, due to the response aggregation mechanism. Given the supportive evidence from the placebo test, we proceed with the sensitive target items.

Prescription drugs to enhance sexual performance. We use PRT to assess the prevalence of purchasing prescription drugs to enhance sexual performance. Table 7 shows that the estimated prevalence of buying medication to enhance sexual performance is much lower, namely 12.3% lower ($p_{\text{Bayes}} = .004$), when using DQ than when using PRT. That is, while under DQ 5.1% report to have purchased such medication at least

Table 5. Study I: Descriptive Statistics.

| | Outside-the-Pair (n = 5,075) | Main Survey (n = 4,837) | PRT (n = 3,687) | DQ (n = 962) |
|--------------------------------------|---------------------------------|----------------------------|----------------------------|----------------------------|
| Female | 53.8% | 53.6% | 53.1% | 53.1% |
| Male | 46.2% | 46.4% | 46.9% | 46.9% |
| Median age (years) | 55 | 56 | 56 | 56 |
| Median monthly net income (in euros) | Between 1,800 and 2,400 | Between 1,800 and 2,400 | Between 1,800 and 2,400 | Between 1,800 and 2,400 |
| Vocational education | 34.1% | 34.6% | 34.6% | 33.0% |
| University-level education | 37.5% | 37.9% | 37.7% | 37.2% |
| Has a partner | 68.4% | 68.7% | 69.2% | 68.0% |

Notes: Median income is net personal monthly income.

Table 6. Study I Placebo Test: Comparing PRT with DQ on the Nonsensitive Target Item.

| Target Item Response Options | PRT | DQ | Difference | Bayesian p-Value |
|--|---------------|---------------|---------------|---------------------|
| a. My birthday is between the 1st and the 20th of a month | 66.0 (1.6) | 64.6 (1.5) | 1.4 (2.2) | .252 |
| b. My birthday is between the 21st and the 25th of a month | 16.7 (3.5) | 18.5 (1.3) | -1.9 (3.7) | .313 |
| c. My birthday is between the 26th and the 31st of a month | 17.3 (2.6) | 16.9 (1.2) | .4 (2.9) | .473 |

Notes: n = 3,687 (PRT) and n = 962 (DQ). Posterior standard errors in parentheses.

once, whereas under PRT that share is 17.4%, which is a striking difference. As for the placebo item, we tested the robustness of the results to various prior specifications. The results are indeed robust (see Web Appendix C).

Under PRT, 3.1% of the participants disclose having purchased such prescription drugs legally, while with DQ 2.2% of the participants indicate having done so ($p_{\text{Bayes}} = .421$). With DQ, 2.0% of the participants report having purchased such prescription drugs illegally, whereas an almost seven-times higher 13.7% of the participants report having done so under PRT ($p_{\text{Bayes}} = .008$). Differences between questioning methods for the last response option (“having purchased prescription drugs both legally and illegally”) are not significant ($p_{\text{Bayes}} = .266$).

Using the MCMC sampler we estimate the proportion of exclusively illegal purchases to be 77.8% (95% CI: [41.3%, 97.1%]) of the total purchase incidence. This estimate is consistent with quite different, unobtrusive but admittedly indirect evidence collected earlier in the country. Venhuis et al. (2014) extracted the amount of sildenafil (an excreted ingredient of Viagra) present in sewage water in three major Dutch cities (Amsterdam, Eindhoven, and Utrecht) during a one-week window. Then, they estimated the percentage of this sildenafil that had been regularly prescribed and purchased

at legal pharmacies countrywide. From the difference, they concluded that at least 60% of the Viagra consumed in these cities had been purchased illegally. We estimate from the current survey that approximately 80% of consumers who buy prescription drugs to improve their sexual life carry out these purchases illegally. This high proportion of illegal purchases and the potential health hazards of using illegally acquired prescription medication suggests that such purchases pose a serious public health issue. We next examine the prevalence of buying the target medication for different demographic subgroups.

Gender effects. Results stratified by gender are in Tables 8 and 9. Under the PRT, 7.2% of the male sample reports having bought prescription drugs to enhance sexual performance legally and 13.5% reports having bought illegally. In contrast, female respondents report almost never purchasing prescription drugs to improve their sexual life. Indeed, there are no approved drugs in the Netherlands for enhancing sexual performance for women, and only few approved drugs for female sexual arousal disorder in the United States (e.g., Addyi, Vyleesi). Still, the survey question asked whether participants had bought prescription drugs to enhance sexual life, which they may have purchased for themselves or on behalf of their partner.

Correlates of illegal purchases. Table 10 illustrates how the PRT can be used to predict the unobserved sensitive trait from the covariates described previously. Due to the low percentage of legal purchases, the focus is on participants who buy only illegally versus those who do not buy. The quadratic specification for age implies that the lowest likelihood of purchasing medications online illegally is approximately between 35–45 years of age in this sample. Low socioeconomic status increases the probability that participants engage in illegal purchases of the medication. In particular, the prevalence of illegal purchases is estimated at 27.9% for individuals who only completed elementary education, at 13.1% for individuals who received vocational or technical education, and at only 5.9% for individuals who received university-level education.

Table 7. Study 1: Comparing PRT and DQ on the Sensitive Target Item.

| Target Item Response Options | PRT | DQ | Difference | Bayesian <i>p</i> -Value |
|---|---------------|--------------|----------------|--------------------------|
| a. I have <i>never</i> purchased prescription drugs that improve my sexual life | 82.6 (3.3) | 94.9 (.7) | -12.3 (3.4) | .004 |
| b. I have purchased <i>at least once</i> prescription drugs that improve my sexual life, but always <i>with</i> a medical prescription | 3.1 (2.3) | 2.2 (.5) | .9 (2.3) | .421 |
| c. I have purchased <i>at least once</i> prescription drugs that improve my sexual life, <i>without</i> a medical prescription | 13.7 (3.8) | 2.0 (.4) | 11.7 (3.8) | .008 |
| d. I have purchased <i>at least once</i> prescription drugs that improve my sexual life, <i>with and without</i> a medical prescription | .6 (.6) | .9 (.3) | -.3 (.7) | .266 |

Notes: *n* = 3,687 (PRT) and *n* = 962 (DQ). Posterior standard errors in parentheses.

Table 8. Study 1: Comparing PRT and DQ (Male Subgroup).

| Target Item Response Options | PRT | DQ | Difference | Bayesian <i>p</i> -Value |
|---|---------------|---------------|----------------|--------------------------|
| a. I have <i>never</i> purchased prescription drugs that improve my sexual life | 77.7 (3.4) | 89.9 (1.4) | -12.2 (3.7) | .001 |
| b. I have purchased <i>at least once</i> prescription drugs that improve my sexual life, but always <i>with</i> a medical prescription | 7.2 (3.9) | 4.2 (.9) | 2.9 (4.0) | .254 |
| c. I have purchased <i>at least once</i> prescription drugs that improve my sexual life, <i>without</i> a medical prescription | 13.5 (4.0) | 3.7 (.9) | 9.7 (4.0) | .017 |
| d. I have purchased <i>at least once</i> prescription drugs that improve my sexual life, <i>with and without</i> a medical prescription | 1.5 (1.3) | 2.0 (.7) | -.5 (1.5) | .307 |

Notes: *n* = 1,728. Posterior standard errors in parentheses.

Table 9. Study 1: Comparing PRT and DQ (Female Subgroup).

| Target Item Response Options | PRT | DQ | Difference | Bayesian <i>p</i> -Value |
|---|---------------|--------------|----------------|--------------------------|
| a. I have <i>never</i> purchased prescription drugs that improve my sexual life | 81.7 (2.9) | 98.7 (.6) | -17.0 (3.0) | <.001 |
| b. I have purchased <i>at least once</i> prescription drugs that improve my sexual life, but always <i>with</i> a medical prescription | 1.5 (1.4) | .6 (.3) | 1.0 (1.4) | .280 |
| c. I have purchased <i>at least once</i> prescription drugs that improve my sexual life, <i>without</i> a medical prescription | 15.3 (2.7) | .6 (.3) | 14.7 (2.7) | <.001 |
| d. I have purchased <i>at least once</i> prescription drugs that improve my sexual life, <i>with and without</i> a medical prescription | 1.5 (1.4) | .2 (.2) | 1.3 (1.4) | .111 |

Notes: *n* = 1,959. Posterior standard errors in parentheses.

The next two empirical applications zoom in on the prevalence of buying such medications illegally and show how the PRT is superior to the RRT using both a within-subjects design (Study 2: lay beliefs about PRT and RRT) and a between-subjects design (Study 3: prevalence and evaluation of PRT and RRT).

Study 2: Lay Beliefs About PRT and RRT

In Study 2, we surveyed 200 participants from Prolific on their lay beliefs (e.g., Bolton, Warlop, and Alba 2003) about the psychological costs of indirect question techniques to participants. Completion rate was approximately 91%. Participants were instructed to imagine working for an organization that plans to conduct a survey about a sensitive topic. The participant's

task was to judge the psychological costs of two indirect question techniques (the PRT and the RRT) that can be used to protect survey participants' privacy, to evaluate each technique, and then to choose the technique that they considered to maximize the likelihood that survey participants would respond truthfully to the questions. Thus, participants did not have to answer the sensitive questions themselves, but only to evaluate the survey techniques and select the best one, based on their lay beliefs. Each indirect question technique was illustrated using the example "purchasing prescription drugs without a medical prescription" with three response options ("never," "occasionally," and "often"); the illustrated PRT is as depicted in Figure 1. The RRT was implemented using an electronic die, which randomly shows one of six die faces upon clicking on a "roll die" button (as in Höglinger and Jann 2018). Each participant evaluated

Table 10. Study 1: Multinomial Logit PRT for Sensitive Target Question on Illegal Medication Purchases.

| Variable | Coefficient | Bayesian p-Value |
|--------------------------------------|---------------|------------------|
| Intercept | -.365 (.747) | .314 |
| Female | -.256 (.281) | .170 |
| Has a partner | .015 (.299) | .494 |
| Age | -.092 (.037) | .001 |
| Age squared | .001 (.0004) | .004 |
| Net monthly income less than \$1,200 | -.306 (.363) | .194 |
| Net monthly income more than \$3,600 | -.796 (.567) | .060 |
| Vocational education | -.717 (.311) | .005 |
| University-level education | -1.738 (.486) | <.001 |
| Subjective well-being | .206 (.136) | .044 |

Notes: $n = 3,687$ (PRT). Posterior standard errors in parentheses. Criterion variable is illegally purchasing prescription drugs to enhance sexual performance. Not purchasing (at all) is the reference category.

both the PRT and the RRT on five criteria, shown in Table 11 (questions a to e), with a slider ranging in values from 0 to 10. The evaluation criteria cover both cognitive costs (questions a and b) and affective costs (questions c to e). Afterward, there was a voluntary free-form evaluation. Finally, survey participants were asked whether they would recommend the technique to ask sensitive questions, using a slider from 0 to 100 (item 6 in Table 11). The presentation order of question techniques was randomized across participants. After evaluating both techniques, participants recommended implementing either the PRT or the RRT.

The PRT scored higher than the RRT on all five criteria (paired t-test for difference in means: $p < .02$ for all items). The overall recommendation score for the PRT is much larger than for the RRT (item f): 71.5% of the sample recommend implementing the PRT over the RRT ($p < .001$, exact binomial test). The differences in average score are much larger for items c and e (as compared with the other items).

Study 3: Prevalence and Evaluation of PRT and RRT

Study 3 builds on the previous two studies by exploring alternative designs to implement the PRT and by comparing the PRT with the RRT on the experienced psychological costs of participation. Study 1 presented the outside-the-pair items in a separate wave to minimize common method bias and hypothesis guessing by participants; however, such a design is often not feasible. Therefore, Study 3 includes the outside-the-pair in the same wave. Further, Study 2 examined lay beliefs (in a within-subjects design) rather than the actual experienced costs of participating in a survey. Hence, Study 3 implements the PRT, RRT, and DQ using a single-wave between-subjects design and asks participants about the psychological costs after they completed the survey. The study uses a convenience sample of $N = 2,999$ male participants in the United Kingdom, recruited from

Prolific (PRT: 52%; RRT: 32%; DQ: 16%). Participants were recruited from five age blocks: 18–27 years old (16.7%), 28–37 years old (18%), 38–47 years old (16.8%), 48–57 years old (20.1%) and 58+ years old (28.5%). The age blocks oversampled older people relative to the standard Prolific population, as purchases of prescription drugs are more common among older people. We compare the PRT with the RRT method (De Jong, Pieters, and Fox 2010) and DQ. Participants evaluated the psychological costs of participating in the PRT and the RRT after completing the survey, but we also report on prevalence estimates under PRT, RRT, and DQ.

The application in the United Kingdom is interesting for regulatory and marketing reasons (Bell et al. 2021). Purchases of sildenafil are covered by public health insurance, and it is legal to purchase Viagra over the counter after consultation with a pharmacist. Furthermore, direct-to-consumer advertising of Viagra sold over the counter is legal, potentially increasing take-up of similar medications (Kim and KC 2020). Therefore, prescription of drugs for erectile dysfunction is substantially more common compared with countries such as the Netherlands (in Study 1).³ We focus on prescription-based purchases because these are more common, being cheaper and often reimbursed by public health insurance (Bell et al. 2021).

The PRT was implemented as follows. Participants first received instructions, followed by the PRT question with a sensitive target item, namely “I have obtained prescription medication to enhance my sexual performance or to alleviate erectile dysfunction, such as Viagra or Cialis.” The item had three response options: “Never,” “I have obtained such medication, and always legally with a doctor’s prescription,” and “I have obtained such medication, but at least once illegally (for instance, from a friend, an unregistered online pharmacy, or some other website).” Figure 1 shows the PRT used by participants.

Four different baseline items were used inside the PRT question, which were selected from two validated personality scales. Respondents in the PRT group randomly got one of the four baseline questions, with the remaining three baseline items potentially asked directly elsewhere in the survey (depending on the subgroup). Thus, there were four experimental subgroups within the PRT group. The advantage of this design is that we also collect direct responses to baseline items used in the PRT question, improving the accuracy of the PRT estimates. Web Appendix C provides the baseline items and randomization scheme for the four groups.

The RRT was implemented using an electronic die to facilitate participation. To reduce cognitive costs, additional text instructions were shown on the screen depending on the die outcome (e.g., “Roll die again,” “Answer truthfully”). After

³ Bell et al. (2021) report that 4,505,623 items for erectile dysfunction were prescribed in the United Kingdom in 2019, corresponding to a rate of .175 prescription per adult man (given a 2019 male population of 25.8 million). In contrast, there were approximately 272,000 items for erectile dysfunction prescribed to 112,000 men in the Netherlands in 2018 (Pharmaceutisch Weekblad 2019), corresponding to a rate of .043 prescription per adult man (given a 2018 male population of 6.4 million).

Table 11. Study 2: Comparing PRT and RRT on Psychological Costs.

| Evaluation Criteria/Items | PRT | RRT | Paired t-Test (p-Value) | Effect Size (Cohen's d) |
|---|---------------|---------------|----------------------------|----------------------------|
| a. The instructions on how to respond to this method are clear for the participants | 68.8 (1.8) | 63.4 (2.1) | 2.64 (.009) | .19 |
| b. I am confident that the participant can follow the instructions of this method correctly | 67.0 (1.9) | 61.0 (2.2) | 2.76 (.006) | .20 |
| c. I understand how this method aims to hide the true answers of the survey participants to the sensitive question | 74.3 (1.7) | 61.0 (2.3) | 5.79 (<.001) | .41 |
| d. I am confident that with this method only the survey participant knows the true answer to the sensitive question | 65.9 (1.8) | 60.0 (2.1) | 2.42 (.017) | .17 |
| e. With this method, survey participants would answer the sensitive question honestly | 68.2 (1.6) | 50.1 (2.1) | 7.98 (<.001) | .57 |
| f. Rate whether you would recommend using this method to ask sensitive questions to survey participants | 50.6 (2.2) | 31.2 (2.1) | 7.46 (<.001) | .53 |
| Which method would you recommend implementing in practice? | 71.5% | 28.5% | | |

Notes: n = 200 (within design). Mean scores. SEs in parentheses.

responding to the sensitive question, participants in the PRT and other (RRT and DQ) conditions evaluated the technique on psychological cost criteria, shown in Table 12, with a five-point scale (1 = “Strongly disagree,” and 5 = “Strongly agree”). Furthermore, we recorded time to completion.

Results: Evaluation. Table 12 summarizes the evaluation of three survey methods by participants. As expected, the PRT outperforms RRT, here on four of the six criteria, namely, criteria a (“instructions are clear”), b (“followed instructions correctly”), c (“understand how answers are hidden”), and f (“worried about privacy”) ($p < .02$). This is consistent with the lay beliefs assessed in Study 2: survey participants find the PRT instructions more comprehensible and feel less worried about their privacy when answering the sensitive questions, compared with survey participants in the RRT condition. As we expected, participants found DQ instructions easier to understand and follow than PRT instructions (criteria a and b), but participants in the DQ condition also felt a bit more uncomfortable and worried about their privacy (criterion e: $p = .026$; criterion f: $p = .011$). Furthermore, completion of the PRT is faster than for the RRT, although not as fast as answering questions directly. Participants require 43.7 seconds on average to answer the PRT versus 52.9 seconds to answer the RRT ($p < .001$). Thus, to the extent that time to respond captures the underlying cognitive processes (Johnson and Payne 1985), participants find it easier and faster to engage with the instructions of the PRT than the RRT.

Results: Prevalence. Prevalence estimates for the three techniques are reported in Table 13. The results indicate that under the PRT, participants are more likely to select sensitive response options (obtaining such medications either legally or illegally). Two PRT models are tested, that is, a “simple” PRT and a PRT where the dependent variable is predicted using a multinomial logit model (“PRT + covs.”), with covariates such as age and

education level (reported in Web Appendix C). Predicting the dependent variable with these covariates increases accuracy.

More specifically, concerning purchases of prescription drugs to enhance sexual performance, survey participants report a significantly *higher* likelihood of picking the option “never” with the RRT (88.8%) as compared with the PRT (77.7%; $\Delta = 11.0\%$ with $p_{\text{Bayes}} = .004$) or PRT + covs. (77.3%; $\Delta = 11.5\%$ with $p_{\text{Bayes}} = .002$). The estimated prevalence of the RRT is even higher than DQ (83.6%; $\Delta = 5.2\%$ with $p_{\text{Bayes}} = .049$), which implies that a substantial number of survey participants did not comply with the RRT instructions. Using covariates to predict the RRT responses reduces the difference with PRT, but it is still quite large (PRT: $\Delta = 9.2\%$ with $p_{\text{Bayes}} = .013$; PRT + covs.: $\Delta = 9.6\%$ with $p_{\text{Bayes}} = .006$). Participants might feel uncomfortable abiding by the instructions when asked to select a sensitive option (John et al. 2018), or they might not trust the use of an electronic randomization device that they cannot control (Coutts and Jann 2011). De Jong, Pieters, and Stremersch (2012) estimate non-compliance with RRT instructions in a U.K. sample to be 15%, although their design relied on a physical die as randomizer.⁴ In contrast to Study 1, the difference between DQ and the PRT in selecting the response option “Never” is smaller, but still approximately 6% (comparison with PRT: $\Delta = 5.9\%$ with $p_{\text{Bayes}} = .052$; comparison with PRT + covs.: $\Delta = 6.3\%$ with $p_{\text{Bayes}} = .029$).

⁴ Web Appendix C reports the results of simulations to examine the potential effect of noncompliance on the RRT results. The simulations vary the percentage of the sample assumed to systematically select the nonsensitive answer option “never,” regardless of the die outcome. The level of noncompliance is unknown, but may be as high as 30% (Reiber, Bryce, and Ulrich 2022). The simulations show that if 10%–20% of the RRT sample would not comply with the instructions, the RRT prevalence estimates would be close to the PRT estimates. However, although corrections for noncompliance are possible, they require additional statistical assumptions and typically result in at least a 2.5-fold increase in variance (e.g., Reiber, Bryce, and Ulrich 2022). The PRT is therefore more efficient than RRT models with noncompliance correction.

Table 12. Study 3: Comparing PRT, RRT and DQ on Psychological Costs of Participation.

| Evaluation Criteria/Time to Response | PRT | RRT | t-Test (p) | DQ | t-Test (p) |
|--|-----------------|-----------------|----------------------|-----------------|----------------------|
| a. The instructions on how to respond on this survey are clear | 4.43 (.02) | 4.24 (.02) | 6.70 ($<.001$) | 4.65 (.02) | -6.72 ($<.001$) |
| b. I am confident that I followed the instructions of this survey correctly | 4.58 (.01) | 4.47 (.02) | 4.72 ($<.001$) | 4.69 (.02) | -3.47 ($<.001$) |
| c. I understand how my true answers to the sensitive questions were hidden | 4.30 (.02) | 3.99 (.03) | 8.51 ($<.001$) | | |
| d. I am confident that only I know my true answer to the sensitive questions | 4.03 (.02) | 4.01 (.03) | .47 (.639) | | |
| e. I felt uncomfortable when answering the sensitive question | 1.85 (.03) | 1.94 (.04) | -1.68 (.094) | 2.00 (.06) | -2.23 (.026) |
| f. I felt worried about my privacy when answering the sensitive question | 1.62 (.02) | 1.71 (.03) | -2.57 (.010) | 1.74 (.05) | -2.56 (.011) |
| Response time (in seconds): | 43.70 (1.00) | 52.87 (1.13) | -5.90 ($<.001$) | 27.24 (1.16) | 8.54 ($<.001$) |

Table 13. Study 3: Comparing PRT, RRT, and DQ on the Sensitive Target Item.

| Target Item | PRT | PRT + Covs. | RRT | RRT + Covs. | DQ |
|--|---------------|---------------|---------------|---------------|---------------|
| I have obtained prescription medication to enhance my sexual performance or to alleviate erectile dysfunction, such as Viagra or Cialis | | | | | |
| Never | 77.7 (3.1) | 77.3 (2.8) | 88.8 (2.6) | 86.9 (2.5) | 83.6 (1.7) |
| I have obtained such medication, and always legally with a doctor's prescription | 14.3 (2.3) | 14.3 (2.2) | 9.4 (2.4) | 9.4 (2.3) | 12.0 (1.5) |
| I have obtained such medication, but at least once illegally (for instance, from a friend, an unregistered online pharmacy, or some other website) | 7.9 (2.9) | 8.4 (2.5) | 1.8 (1.4) | 3.7 (1.4) | 4.4 (1.0) |

Notes: n = 1,572 (PRT), n = 955 (RRT) and n = 472 (DQ). Posterior SEs in parentheses.

Conclusion

We introduced the paired response technique (PRT), a novel methodology to obtain truthful answers to sensitive survey questions. The technique guarantees unconditional privacy protection to participants and is easy to implement in large-scale marketing and other social science studies for any type of sensitive question. Our studies provide evidence that the PRT has several advantages compared with the randomized response technique (RRT), a leading indirect questioning technique for categorical questions (De Jong, Pieters, and Fox 2010). We empirically show that participants prefer the PRT to the RRT, finding it easier to understand and more pleasant to use. It is also faster to administer and provides prevalence estimates that are more realistic and precise, especially when RRT models suffer from noncompliance. To facilitate adopting the PRT in future marketing research, we have provided annotated Matlab and Python code on OSF.

The empirical applications of PRT produced noteworthy, new results on consumers purchasing drugs to enhance sexual performance, with important theoretical and policy implications. From the empirical applications, we estimate that approximately 17.4% of the population in the Netherlands obtained drugs to enhance sexual performance, and the majority of these consumers

bought the drugs illegally. Standard DQ methods dramatically underestimate this. The differences between PRT and DQ are not due to the technique itself, as statistical placebo tests indicate.

We further estimate that in the United Kingdom, where regulation for purchasing medications to enhance sexual performance is less strict, 22.3% of a nonprobability sample obtained such medications. However, 14.3% of this sample obtained such medications with a prescription, whereas only 7.9% did so at least once illegally. In contrast to the Dutch sample, the difference between the PRT and DQ was smaller in the U.K. sample, perhaps due to regulatory differences, cross-cultural differences in the acceptability of using such medication, or differences between the probability and nonprobability samples. Nonetheless, the PRT appears more suitable and less taxing to participants than the RRT to investigate these issues.

Research on the incidence and determinants of legal and illegal purchases of prescription drugs is urgent, in view of the health threats of consuming counterfeit medicine, incorrect medication dosing by consumers and medicine consumption untraced by regular health care systems (Kao et al. 2009; OECD and EUIPO 2020; World Health Organization 2017). Koenraadt and Van de Ven (p. 353, 2018) stress that "future research should monitor the demand and sale of lifestyle drugs as well as further

differentiating between the motives and methods for purchasing various substances.” Yet, reporting on such purchases is sensitive, which might lead to deliberate underreporting if using direct questions with potential grave implications for policy making and evaluation. Moreover, professional research firms are held to exercise special care when the nature of the research is sensitive and data collection may upset or disturb participants (e.g., ICC/ESOMAR 2016, article 1), which may limit or complicate data collection using DQ on such sensitive issues. The proposed PRT may open new avenues for research on sensitive issues. Such research should delve more deeply into the drivers of these sensitive behaviors, ideally further explaining which consumers are more likely to make them and proposing and evaluating policies to discourage them from doing so.

In closing, the PRT is a new, straightforward technique to ask sensitive questions in surveys. It is easy to implement and comprehend and promises to provide new insights into topics that are relevant to policy and theory, but about which consumers may choose not to speak truthfully if asked directly.

Technical Appendix

The conditional posterior distributions of the model based on the likelihood are given here. We present only the sampler for estimating the prevalence of the target item (defined as β), but we provide codes implementing the PRT as well for relating the sensitive target item to other observed covariates.

1. Draw W_{ih} for $i = 1, \dots, N$ and $h = 1, \dots, H$ and γ_h . A data augmentation step is used for the IRT structure of Equations 3 and 4, as in Albert and Chib (1993). For the outside-the-pair items $S_{ih}^{(0)}$ we introduce variables W_{ih} , such that $S_{ih}^{(0)} = c$ if $\gamma_{h,c-1} < W_{ih} < \gamma_{h,c}$. The full conditional posterior of W_{ih} is then⁵

$$W_{i,h}|S_{ih}^{(0)}, \theta_i, \alpha_h, \gamma_h \sim \mathbb{I}(\gamma_{h,c-1} < W_{ih} < \gamma_{h,c})N(\alpha_h\theta_i, 1) \quad (9)$$

$$\text{if } S_{ih}^{(0)} = c.$$

2. Draw $W_{i,bas}$ for $i = 1, \dots, N$ given $Z_i, \theta_i, \alpha_{bas}$ and γ_{bas} . We introduce variables $W_{i,bas}$, such that $Z_i = k$ if $\gamma_{bas,k-1} < W_{i,bas} < \gamma_{bas,k}$. The full conditional posterior of $W_{i,bas}$ is then

$$W_{i,bas}|Z_i, \theta_i, \alpha_{bas}, \gamma_{bas} \sim \mathbb{I}(\gamma_{bas,k-1} < W_{i,bas} < \gamma_{bas,k})N(\alpha_{bas}\theta_i, 1) \quad (10)$$

$$\text{if } Z_i = k.$$

3. Draw θ_i for $i = 1, \dots, N$ given $W_{i,h}, W_{i,bas}, \alpha_h, \alpha_{bas}$. The full conditional posterior of θ_i is

$$\theta_i|W_{i,h}, W_{i,bas}, \alpha_h, \alpha_{bas} \sim N((\alpha^T \alpha + 1)^{-1}, \alpha^T \mathbf{w}_i), \quad (11)$$

with $\alpha = [\alpha_1, \dots, \alpha_H, \alpha_{bas}]$ and $\mathbf{w}_i = [W_{i,1}, \dots, W_{i,H}, W_{i,bas}]$.

4. Draw α_h for $h = 1, \dots, H$ given $W_{i,h}$ and θ_i . The conditional posterior of α_h is proportional to $\pi(\alpha_h) \prod_i N(W_{i,h}|\alpha_h\theta_i, 1)$, where $\pi(\alpha_h)$ is the chosen prior distribution. The parameter is sampled using a Metropolis–Hastings algorithm, with proposal density $\mathbb{I}(\alpha_h^* > 0) \sim N(\alpha_h^{(m-1)}, \sigma_{\alpha, MH})$, where $\alpha_h^{(m-1)}$ is the previous draw and α_h^* is the proposal draw.
5. Draw α_{bas} given $W_{i,bas}$ and θ_i . Sampled as in Step 4.
6. Draw γ_h for $h = 1, \dots, H$ given $S_{ih}^{(0)}, \theta_i$, and α_h . The conditional posterior of γ_h is proportional to

$$\pi(\gamma_h) \prod_i (\Phi(\alpha_h\theta_i - \gamma_{h,c-1}) - \Phi(\alpha_h\theta_i - \gamma_{h,c}))^{1(S_{ih}^{(0)}=c)}, \quad (12)$$

where $\pi(\gamma_h)$ is the chosen prior distribution. The parameter is sampled using a Metropolis–Hastings algorithm, with proposal density:

$$\mathbb{I}(\gamma_{h,c-1}^* < \gamma_{h,c}^* < \gamma_{h,c+1}^{(m-1)}) \sim N(\gamma_h^{(m-1)}, \sigma_{\gamma, MH}^2). \quad (13)$$

7. Draw γ_{bas} given Z_i, θ_i , and α_{bas} . Sampled as in Step 6.
8. Draw U_i and Z_i for $i = 1, \dots, N$. The joint posterior probability of U_i and Z_i conditional on the observed response to the PRT Y_i is

$$P(Z_i = k, U_i = j|Y_i = 1) = \frac{P(Y_i = 1|U_i = j, Z_i = k)P(U_i = j)P(Z_i = k)}{P(Y_i = 1)}. \quad (14)$$

These probabilities can be computed using Equation 4 (to compute $P(Z_i = k)$), Step 9 below (to compute $P(U_i = j)$), and Table 3 (to compute $P(Y_i = 1|U_i = j, Z_i = k)$ and $P(Y_i = 1)$). The probabilities are then used to compute the marginal probabilities $P(Z_i = k)$ and $P(U_i = j)$. For instance, in the example with three response options, for the case $Z_i = 1$:

$$P(Z_i = 1) = P(Z_i = 1, U_i = 1|Y_i = 1)\mathbb{I}(Y_i = 1) + P(Z_i = 1, U_i = 2|Y_i = 2)\mathbb{I}(Y_i = 2) + P(Z_i = 1, U_i = 3|Y_i = 3)\mathbb{I}(Y_i = 3), \quad (15)$$

where $\mathbb{I}(\cdot)$ is an indicator function. Notice that Table 3 matches Y_i to a specific response combination Z_i and U_i ; for instance $Y_i = 1$ corresponds to $\{U_i + Z_i = 2\} \vee \{U_i + Z_i = 5\}$. Then,

⁵ We use the notation $x \sim \mathbb{I}(a < x < b)N(\mu, \sigma)$ to indicate that variable x has a normal distribution with mean μ and variance σ , but truncated between lower bound a and upper bound b . $\mathbb{I}(\cdot)$ denotes an indicator function.

these probabilities are used to draw U_i and Z_i :

$$\begin{aligned} Z_i &\sim \text{Multinomial}(P(Z_i = 1), \dots, P(Z_i = K)) \\ U_i &\sim \text{Multinomial}(P(U_i = 1), \dots, P(U_i = J)). \end{aligned} \quad (16)$$

9. Draw β given U_i . We define the prevalence of the target item $P(U_i = j) = \beta_j$; we assume a noninformative Dirichlet prior $\text{Dirichlet}(1, \dots, 1)$. Hence, the full conditional posterior of β is

$$\beta|U_i \sim \text{Dirichlet}\left(\sum_i \mathbb{I}(U_i = 1) + 1, \dots, \sum_i \mathbb{I}(U_i = J) + 1\right).$$

where $\mathbb{I}(U_i = j)$ is an indicator for whether individual i has latent response $U_i = j$. The sampler can be extended to predict the target item U_i using covariates when in Step 9. For the multinomial logit model this is implemented using Pólya–gamma data augmentation (Polson, Scott, and Windle 2013).

Acknowledgments

The authors are very grateful to the JMR review team for many constructive comments that significantly improved the article. The first author thanks the Scientific Computing Research Technology Platform of the University of Warwick for providing access to computing facilities.

Coeditor

Sachin Gupta

Associate Editor

Neeraj Arora


Declaration of Conflicting Interests

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

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Adri Bakker Foundation.

ORCID iD

Marco Gregori  <https://orcid.org/0000-0002-2956-4526>

References

- Albert, James H. and Siddhartha Chib (1993), “Bayesian Analysis of Binary and Polychotomous Response Data,” *Journal of the American Statistical Association*, 88 (422), 669–79.
- Bearden, William O. and Richard G. Netemeyer (1999), *Handbook of Marketing Scales: Multi-Item Measures for Marketing and Consumer Behavior Research*. Sage Publications.
- Bell, Colette, Muhammad Abdul Hadi, Saval Khanal, and Vibhu Paudyal (2021), “Prescribing Patterns and Costs Associated with Erectile Dysfunction Drugs in England: A Time Trend Analysis,” *BJGP Open*, 5 (2), 1–8.
- Blair, Graeme, Alexander Coppock, and Margaret Moor (2020), “When to Worry About Sensitivity Bias: A Social Reference Theory and Evidence from 30 Years of List Experiments,” *American Political Science Review*, 114 (4), 1–19.
- Blair, Graeme and Kosuke Imai (2012), “Statistical Analysis of List Experiments,” *Political Analysis*, 20 (1), 47–77.
- Blair, Graeme, Kosuke Imai, and Yang-Yang Zhou (2015), “Design and Analysis of the Randomized Response Technique,” *Journal of the American Statistical Association*, 110 (511), 1304–19.
- Bolton, Lisa E., Luk Warlop, and Joseph W. Alba (2003), “Consumer Perceptions of Price (Un) Fairness,” *Journal of Consumer Research*, 29 (4), 474–91.
- Chan, Wui Ling, David Michael Wood, and Paul Ivor Dargan (2015), “Significant Misuse of Sildenafil in London Nightclubs,” *Substance Use & Misuse*, 50 (11), 1390–4.
- Chaudhuri, Arijit and Tasos C. Christofides (2013), *Indirect Questioning in Sample Surveys*. Springer-Verlag.
- Coutts, Elisabeth and Ben Jann (2011), “Sensitive Questions in Online Surveys: Experimental Results for the Randomized Response Technique (RRT) and the Unmatched Count Technique (UCT),” *Sociological Methods & Research*, 40 (1), 169–93.
- De Jong, Martijn G. and Rik Pieters (2019), “Assessing Sensitive Consumer Behavior Using the Item Count Response Technique,” *Journal of Marketing Research*, 56 (3), 345–60.
- De Jong, Martijn G., Rik Pieters, and Jean-Paul Fox (2010), “Reducing Social Desirability Bias Through Item Randomized Response: An Application to Measure Underreported Desires,” *Journal of Marketing Research*, 47 (1), 14–27.
- De Jong, Martijn G., Rik Pieters, and Stefan Stremersch (2012), “Analysis of Sensitive Questions Across Cultures: An Application of Multigroup Item Randomized Response Theory to Sexual Attitudes and Behavior,” *Journal of Personality and Social Psychology*, 103 (3), 543–64.
- Diener, Edward D., Robert A. Emmons, Randy J. Larsen, and Sharon Griffin (1985), “The Satisfaction with Life Scale,” *Journal of Personality Assessment*, 49 (1), 71–5.
- Dwork, Cynthia and Aaron Roth (2014), “The Algorithmic Foundations of Differential Privacy,” *Foundations and Trends in Theoretical Computer Science*, 9 (3–4), 211–407.
- Fox, Jean-Paul (2010), *Bayesian Item Response Modeling: Theory and Applications*. Springer-Verlag.
- Gelatti, Umberto, Roberta Pedrazzani, Claudio Marcantoni, Silvia Mascaretti, Carla Repice, Laura Filippucci, I. Zerbini, M. Dal Grande, G. Orizio, and Donatella Feretti (2013), “‘You’ve Got M@il: Fluoxetine Coming Soon!’: Accessibility and Quality of a Prescription Drug Sold on the Web,” *International Journal of Drug Policy*, 24 (5), 392–401.
- Gnambs, Timo (2014), “A Meta-Analysis of Dependability Coefficients (Test–Retest Reliabilities) for Measures of the Big Five,” *Journal of Research in Personality*, 52, 20–28.
- Goldberg, Lewis R. (1992), “The Development of Markers for the Big-Five Factor Structure,” *Psychological Assessment*, 4 (1), 26–42.
- Goldman, Dana P. and Anupam B. Jena (2011), “Growing Internet Use May Help Explain the Rise in Prescription Drug Abuse in the United States,” *Health Affairs*, 30 (6), 1192–99.

- 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, 1470–83.
- Hoffmann, Adrian, Julia Meisters, and Jochen Musch (2020), "On the Validity of Non-Randomized Response Techniques: An Experimental Comparison of the Crosswise Model and the Triangular Model," *Behavior Research Methods*, 52, 1768–82.
- Höglinger, Marc and Ben Jann (2018), "More Is Not Always Better: An Experimental Individual-Level Validation of the Randomized Response Technique and the Crosswise Model," *PLOS One*, 13 (8), e020177.
- ICC/ESOMAR (2016), "ICC/ESOMAR International Code on Market, Opinion and Social Research and Data Analytics," International Chamber of Commerce/ESOMAR, <https://esomar.org/code-and-guidelines/icc-esomar-code>.
- Ivanitskaya, Lara, Jodi Brookins-Fisher, Irene O'Boyle, Danielle Vibbert, Dmitry Erofeev, and Lawrence Fulton (2010), "Dirt Cheap and Without Prescription: How Susceptible Are Young US Consumers to Purchasing Drugs from Rogue Internet Pharmacies?" *Journal of Medical Internet Research*, 12 (2), e11.
- John, Leslie K., George Loewenstein, Alessandro Acquisti, and Joachim Vosgerau (2018), "When and Why Randomized Response Techniques (Fail to) Elicit the Truth," *Organizational Behavior and Human Decision Processes*, 148, 101–23.
- Johnson, Eric J. and John W. Payne (1985), "Effort and Accuracy in Choice," *Management Science*, 31 (4), 395–414.
- Kao, Shih Ling, Cheng Leng Chan, Belinda Tan, Tchoyoson C.C. Lim, Rinkoo Dalan, Daphne Gardner, Edward Pratt, Marilyn Lee, and Kok Onn Lee (2009), "An Unusual Outbreak of Hypoglycemia," *New England Journal of Medicine*, 360 (7), 734–35.
- Kim, Tongil "TI" and Diwas KC (2020), "Can Viagra Advertising Make More Babies? Direct-to-Consumer Advertising on Public Health Outcomes," *Journal of Marketing Research*, 57 (4), 599–616.
- Koenraadt, Rosa and Katinka van de Ven (2018), "The Internet and Lifestyle Drugs: An Analysis of Demographic Characteristics, Methods, and Motives of Online Purchasers of Illicit Lifestyle Drugs in the Netherlands," *Drugs: Education, Prevention and Policy*, 25 (4), 345–55.
- Krumpal, Ivar and Thomas Voss (2020), "Sensitive Questions and Trust: Explaining Respondents' Behavior in Randomized Response Surveys," *SAGE Open*, 10 (3), 2158244020936223.
- Kuha, Jouni and Jonathan Jackson (2014), "The Item Count Method for Sensitive Survey Questions: Modelling Criminal Behaviour," *Journal of the Royal Statistical Society: Series C: Applied Statistics*, 63 (2), 321–41.
- Kwan, Samuel S.K., Mike K.P. So, and Kar Yan Tam (2010), "Research Note—Applying the Randomized Response Technique to Elicit Truthful Responses to Sensitive Questions in IS Research: The Case of Software Piracy Behavior," *Information Systems Research*, 21 (4), 941–59.
- 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 & Research*, 33 (3), 319–48.
- Lipsitch, Marc, Eric Tchetgen Tchetgen, and Ted Cohen (2010), "Negative Controls: A Tool for Detecting Confounding and Bias in Observational Studies," *Epidemiology*, 21 (3), 383–88.
- Napolitano, Francesco, Maria Teresa Izzo, Gabriella Di Giuseppe, and Italo F. Angelillo (2013), "Frequency of Inappropriate Medication Prescription in Hospitalized Elderly Patients in Italy," *PLOS One*, 8 (12), e82359.
- Nepusz, Tamás, Andrea Petróczy, Declan P. Naughton, Tracy Epton, and Paul Norman (2014), "Estimating the Prevalence of Socially Sensitive Behaviors: Attributing Guilty and Innocent Noncompliance with the Single Sample Count Method," *Psychological Methods*, 19 (3), 334–55.
- Nuffield Council on Bioethics (2009), "Medical Profiling and Online Medicine: the Ethics of 'Personalised Healthcare' in a Consumer Age," research report, Nuffield Council on Bioethics (July 6).
- OECD and EUIPO (2020), *Illicit Trade: Trade in Counterfeit Pharmaceutical Products*. OECD Publishing.
- Pharmaceutisch Weekblad (2019), "Erectiepillen meest verstrekt rond de pensioenleeftijd," (May 31), <https://www.pw.nl/vaste-rubrieken/sfk/2019/erectiepillen-meest-verstrekt-rond-de-pensioenleeftijd>.
- Polson, Nicholas G., James G. Scott, and Jesse Windle (2013), "Bayesian Inference for Logistic Models Using Pólya–Gamma Latent Variables," *Journal of the American Statistical Association*, 108 (504), 1339–49.
- Reiber, Fabiola, Donna Bryce, and Rolf Ulrich (2022), "Self-Protecting Responses in Randomized Response Designs: A Survey on Intimate Partner Violence During the Coronavirus Disease 2019 Pandemic," *Sociological Methods & Research* (published online January 17), <https://doi.org/10.1177/004912412111043138>.
- Rosenfeld, Bryn, Kosuke Imai, and Jacob N. Shapiro (2016), "An Empirical Validation Study of Popular Survey Methodologies for Sensitive Questions," *American Journal of Political Science*, 60 (3), 783–802.
- Samejima, Fumiko (1969), *Estimation of Latent Ability Using a Response Pattern of Graded Scores (Psychometric Monograph No. 17)*. Psychometric Society.
- Schneider, Matthew J., Sharan Jagpal, Sachin Gupta, Shaobo Li, and Yan Yu (2017), "Protecting Customer Privacy When Marketing with Second-Party Data," *International Journal of Research in Marketing*, 34 (3), 593–603.
- Schwarz, Norbert (1995), "What Respondents Learn from Questionnaires: The Survey Interview and the Logic of Conversation," *International Statistical Review/Revue Internationale de Statistique*, 63 (2), 153–68.
- Steenkamp, Jan-Benedict E.M., Martijn G. De Jong, and Hans Baumgartner (2010), "Socially Desirable Response Tendencies in Survey Research," *Journal of Marketing Research*, 47 (2), 199–214.
- Tourangeau, Roger and Ting Yan (2007), "Sensitive Questions in Surveys," *Psychological Bulletin*, 133 (5), 858–83.
- Umesh, Uchila N. and Robert A. Peterson (1991), "A Critical Evaluation of the Randomized Response Method: Applications, Validation, and Research Agenda," *Sociological Methods & Research*, 20 (1), 104–38.
- United Nations Conference on Trade and Development (2020), *COVID-19 and E-commerce. Findings from a Survey of Online Consumers in 9 Countries*. United Nations Conference on Trade and Development.
- Venhuis, Bastiaan J., Pim de Voogt, Erik Emke, Ana Causanilles, and Peter H.J. Keizers (2014), "Success of Rogue Online Pharmacies:

- Sewage Study of Sildenafil in the Netherlands,” *British Medical Journal*, 349, g4317.
- Venhuis, Bastiaan J., Ruth Mosimann, Lynda Scammell, Domenico Di Giorgio, Roy Van Cauwenberghe, Marcel G.A.M. Moester, Mickey Arieli, and Sabine Walser (2013), “Identification of Health Damage Caused by Medicrime Products in Europe: An Exploratory Study,” Research Report No. 040003001, RIVM (August 8).
- Warner, Stanley L. (1965), “Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias,” *Journal of the American Statistical Association*, 60 (309), 63–69.
- White, Katherine, Rishad Habib, and David J. Hardisty (2019), “How to SHIFT Consumer Behaviors to Be More Sustainable: A Literature Review and Guiding Framework,” *Journal of Marketing*, 83 (3), 22–49.
- WHO (2017), *Global Surveillance and Monitoring System for Substandard and Falsified Medical Products*. World Health Organization.
- Wolter, Felix and Peter Preisendörfer (2013), “Asking Sensitive Questions: An Evaluation of the Randomized Response Technique Versus Direct Questioning Using Individual Validation Data,” *Sociological Methods & Research*, 42 (3), 321–53.
- Yu, Jun-Wu, Guo-Liang Tian, and Man-Lai Tang (2008), “Two New Models for Survey Sampling with Sensitive Characteristic: Design and Analysis,” *Metrika*, 67 (3), 251–63.