Qual Quant
DOI 10.1007/s11135-017-0537-0

A mixed-mode sensitive research on cannabis use and sexual addiction: improving self-reporting by means of indirect questioning techniques

Pier Francesco Perri[1] · **Beatriz Cobo Rodríguez**[2] · **María del Mar Rueda García**[2]

Abstract In this article, we describe the methods employed and the results obtained from a mixed-mode "sensitive research" conducted in Spain to estimate certain aspects concerning patterns of cannabis consumption and sexual addiction among university students. Three different data-collection methods are considered and compared: direct questioning, randomized response technique and item sum technique. It is shown that posing direct questions to obtain sensitive data produces significantly lower estimates of the surveyed characteristics than do indirect questioning methods. From the analysis, it emerges that male students seem to be more affected by sex addiction than female students while for cannabis consumption there is no evidence of a predominant gender effect.

Keywords Bar-Lev et al. (2004) method · Item sum technique · Mixed-mode surveys · Privacy protection

✉ Pier Francesco Perri
pierfrancesco.perri@unical.it;
http://www.ecostat.unical.it/perri/

Beatriz Cobo Rodríguez
beacr@ugr.es

María del Mar Rueda García
mrueda@ugr.es

[1] Department of Economics, Statistics and Finance, University of Calabria, Via P. Bucci, 87036 Arcavacata di Rende (CS), Italy

[2] Department of Statistics and Operational Research, University of Granada, Campus Fuentenueva S/N, 18071 Granada, Spain

Journal : Small 11135 | Dispatch : 3-8-2017 | Pages : 19
Article No. : 537 | LE | TYPESET
MS Code : QUQU-D-17-00222 | CP | DISK

1 Introduction

Nowadays, large-scale surveys of human population delve increasingly into sensitive topics, which notoriously produce dishonest or misleading answers, and these, in turn, generate a well-known source of bias in survey, called *social desirability bias*, i.e. the tendency to present oneself in a positive light. Survey participants exhibit this bias when they overreport socially acceptable attitudes which conform to social norms (e.g., giving to charity, believing in God, church attendance, voting, healthy eating, doing voluntary work) and underreport socially disapproved, undesirable behaviours which deviate from social rules (e.g., xenophobia, anti-Semitism, gambling, consumption of alcohol, abortion, smoking among teens and by pregnant women, drug legalization). This type of bias generally produces an over/underestimation of the behaviour under study which may lead to inconsistent analyses and erroneous conclusions. Sometimes respondents may be reluctant to answer questions that do not specifically pertain to social desirability attitudes, for example concerning *taboo* topics which appear intrusive in some way of respondents' private sphere. Questions about income, sexual practices, domestic violence, stalking, political parties, religion and so on fall into this category and risk offending respondents regardless of their true status on the matter. Other questions may instead provoke concerns about the *threat of disclosure*, i.e., fears about the negative consequences that might occur to the respondents if confidential data collected by the researcher were released to third parties not directly involved in the survey, even if the protection of confidentiality and data nondisclosure were guaranteed. Questions falling in this case concern, for instance, illegal drug use and pushing, tax dodging, sexual abuses, and non-compliance with rules and regulations.

Doing "sensitive research" (see, e.g., Liamputtong 2007; Tourangeau and Yan 2007; Dickson-Swift et al. 2008) on stigmatizing, highly personal, embarrassing, threatening or even incriminating issues - especially by direct questioning (DQ) modes - is not an easy matter since it is likely to meet with three sources of errors: (1) refusal to cooperate (*unit-non-response*); (2) refusal to answer specific questions (*item-non-response*); (3) untruthful answers (*measurement error*). Refusal to answer and false information constitute non-sampling errors that are difficult to deal with and can seriously flaw the quality of the data, thus jeopardizing the usefulness of subsequent analyses, including the statistical inference on unknown characteristics of the population under study. Although these errors cannot be totally avoided, they may be mitigated by increasing respondents' cooperation, carefully considering key points such as the modes in which the survey is administered, the presence of the interviewer, whether it is the interviewer who poses the questions, the format of the questionnaire, the wording and the placing of the sensitive items in the questionnaire, the data-collection setting, the presence of other people and, above all, strongly assuring about anonymity and confidentiality (on this, see, e.g.,Tourangeau and Smith 1996; Groves et al. 2004).

Since the decision to cooperate honestly greatly depends on how survey participants perceive the possibility of their privacy being infringed, survey modes which ensure respondents' anonymity or, at least, a high degree of confidentiality, may go some way to improving cooperation and, consequently, to obtaining more reliable information on sensitive topics than can be gathered with DQ. In order to increase respondents' cooperation, many different strategies have been developed. One possibility for improving reporting on sensitive topics is to limit the influence of the interviewer in the question and answer process, as the presence of the interviewer tends to increase socially desirability effects. This goal is traditionally pursued by means of self-administered questionnaires (SAQs)

Journal : Small 11135 Dispatch : 3-8-2017 Pages : 19
Article No. : 537 LE TYPESET
MS Code : QUQU-D-17-00222 CP DISK

with paper and pencil, the interactive voice response (IVR) technique, computer-assisted telephone interviewing (CATI), computer-assisted self interviewing (CASI), audio computer-assisted self interviewing (ACASI) or by computer-assisted Web interviewing (CAWI).

Alternatively, since the 1960s a variety of questioning methods have been devised to ensure respondents' anonymity and to reduce the incidence of evasive answers and the over/underreporting of socially undesirable acts. These methods are generally known as indirect questioning techniques (IQTs; for a review see Chaudhuri and Christofides 2013) and they obey the principle that no direct question is posed to survey participants. Therefore, there is no need for respondents to openly reveal whether they have actually engaged in activities or present attitudes that are socially sensitive. Their privacy is protected because the responses remain confidential to the respondents and, consequently, their true status remains uncertain and undisclosed to both the interviewer and the researcher. Nonetheless, although the individual information, provided by the respondents according to the rules prescribed by the adopted IQT, cannot be used to discover their true status regarding the sensitive issues, the information gathered for all the survey participants can be profitably employed to draw inferences on certain parameters of interest for the study population, including the prevalence of a sensitive behaviour pattern, its frequency, the mean of a sensitive quantitative variable, the level of sensitivity of a question and so on.

The IQTs comprise various strategies for eliciting sensitive information, which mainly encompass these approaches: the randomized response (RR) technique (RRT), the item count technique (ICT) and the nonrandomized response technique (NRRT). In terms of the volume of research conducted in this field since Warner's (1965) pioneering work on indirect questioning, the RRT maintains a prominent position among IQTs. Fundamentally, the RRT employs (at least in its original formulation) a physical randomization device (decks of cards, coloured numbered balls, dice, coins, spinners, random number generators, etc.) which determines whether respondents should answer the sensitive question or another, neutral one or even provide a pre-specified response (e.g., "yes") irrespective of their true status concerning the sensitive behaviour. The randomization device generates a probabilistic relation between the sensitive question and the answer given, which is used to draw inferences on unknown parameters of interest. The rationale of the RRT is that the respondents are less inhibited when the confidentiality of their responses is guaranteed. This goal is achieved because all responses are given according to the outcome of the randomization procedure, which is unknown to both the interviewer and the researcher and, hence, respondents' privacy is preserved.

Similar protection is assured by the ICT (Miller 1984). Without loss of generality, by using this approach, the respondents receive a list of sensitive and innocuous items and are asked to report the total number of items that apply to them without revealing which item applies individually.

Finally, in the NRRT, no physical device is adopted, and neither are respondents asked to conduct a randomizing procedure (Tian and Tang 2014). Instead, the respondents answer according to their true beliefs regarding the sensitive question and to one or more non-sensitive variables.

In this article, we discuss the use of two IQTs in order to analyze some patterns of drug use and sexual behaviour which, traditionally, represent sensitive research fields that are difficult to investigate empirically. In recent years, although the IQTs have grown in popularity as effective methods for investigating the two issues, and various surveys have been conducted to measure the prevalence of drug use and sexual behaviour, very few studies have focused on estimating the characteristics of quantitative variables related to these topics. Therefore, we focus on the use of the RRT and the ICT in a real study conducted in Spain to investigate the frequency of certain sensitive phenomena concerning drug addiction and sexual behaviour among university students. In particular, given the quantitative nature of the variables surveyed, we use *ad hoc* procedures, termed the scrambling response method by Bar-Lev et al. (2004) and the recent variant of the ICT, termed the item sum technique (IST), proposed by Chaudhuri and Christofides (2013) and first employed by Trappmann et al. (2014) in a CATI survey. To the best of our knowledge, this is the first time that these two IQTs have been simultaneously employed to investigate cannabis consumption and sexual addiction, and both compared with the DQ method.

The motivating idea of the article is to compare the estimates obtained through DQ with those stemming from the above-described IQTs. The results of this study clearly show that DQ produces underreporting of the incidence of sensitive phenomena while the IQTs procure significantly larger estimates of the characteristics of interest, and at the same time enhance respondents' confidentiality and, thus, reduce nonresponse rates.

The article is also inspired by some considerations and suggestions given in Trappmann et al. (2014) who state (p. 68): "*Survey researchers aiming at measuring sensitive behaviors at a quantitative scale could therefore benefit from using the IST. Nonetheless, our study can only be regarded as a first step in the development and evaluation of the new technique*". The present paper is a step in this direction, providing empirical evidence of the effectiveness of the IST. The authors also affirm (p. 68): " *Although RRT schemes tailored to quantitative sensitive characteristics have been proposed in the literature [...] there is little evidence on how these techniques perform in practice*". Our contribution seeks to fill this gap, describing the practical implementation of the RRT for quantitative sensitive characteristics, making use of a smartphone mobile application, and evaluating the performance of the RRT and the estimates obtained.

The rest of the article is organized as follows. In Sect. 2, we introduce and discuss some issues related to cannabis consumption and sexual behaviour. Section 3 describes, in a general setting, the Bar-Lev et al. (2004) procedure (Sect. 3.1) and the IST (Sect. 3.2) used in the study. Section 4 is devoted exclusively to the description of our research. In particular, Sect. 4.1 outlines the main features and the fieldwork conducted in the survey, while Sect. 4.2 comments the results obtained for the sensitive characteristics investigated. In Sect. 5, we acknowledge a recent contribution concerning optimal sample size allocation in IST surveys, and investigate the improvement upon the efficiency of the estimates through a simulation study. Section 6 concludes the article with some final considerations.

2 Measuring cannabis use and sexual behaviour

Illicit drugs use damages the health and well-being of millions people. Cardiovascular disease, stroke, cancer, HIV, hepatitis, respiratory diseases, neurological/mental or emotional disorders (agitation, aggression, psychosis and anxiety) can all be provoked or aggravated by drug use. Moreover, drugs have a severe impact in terms of social costs.

Estimating the prevalence of illicit drug use is a major concern for health and social operators, government agencies and policymakers seeking to evaluate the social and economic impact of illicit substances. Accurate data in this respect are needed to plan public intervention programmes, to promote drug prevention campaigns and to gauge progress towards improving the behavioural health of the population and towards reducing injurious effects and social costs.

Cannabis (or marijuana), the crude drug derived from *Cannabis Sativa L.* pistillate inflorescence, is the most widely-consumed illicit drug in the world, especially among young people. Although young males have historically had a higher prevalence of cannabis use, current results indicate that male-female differences in cannabis use are decreasing (Johnson et al. 2015).

Cannabis is often used for its mental and physical effects, such as heightened mood and relaxation, and it has been cited in the medical literature as a potential secondary treatment agent for severe pain, muscle spasticity, anorexia, nausea, sleep disturbance and numerous other conditions (Lamarine 2012). As with the majority of drugs, cannabis causes neurological effects both in the short term (alerted senses, changes in mood, insomnia, impaired body movement, difficulty in thinking and problem-solving, impaired memory) and in the long term (reduced cognitive, memory and learning functions). In addition, it may provoke mental consequences such as hallucinations, paranoia and schizophrenia.

There exists an enormous volume of government reports, medical and sociological research articles and data from various sources on the spread of cannabis, its determinants and effects. According to the latest data published by the European Monitoring Centre for Drugs and Drug Addiction (EMCDDA 2016) over 88 million adults, or just over a quarter of the EU population aged 15–64, are estimated to have tried illicit drugs at some point in their lives. Across all age groups, cannabis is the illicit drug most likely to be used. An estimated 16.6 million young Europeans aged 15–34 (13.3% of this age group) used cannabis in the last year before the survey, with 9.6 million of these aged 15-24 (16.4% of this age group). Cannabis accounts for the majority of illicit drug use among school-aged children.

Table 1 shows some data for Spain referred to year 2013. On average, 17% of young adults (23.6% of males and 10.3% of females) consumed cannabis at some time during the 12 months preceding the survey and, among all individuals aged 15–64, the estimated prevalence of those who have consumed cannabis at least once in their lifetime is nearly 30.4%. The use of cannabis is more prevalent among males than females.

Table 1 Prevalence (in percentage) of cannabis consumption among the Spanish population. *Source*: EMCDDA (2016)

	Use		
	Lifetime	Last year	Last month
All adults (15–64)			
Total	30.4	9.2	6.6
Males	37.9	12.9	9.8
Females	22.7	5.4	3.4
Young adults (15–34)			
Total	40.2	17.0	12.2
Males	47.8	23.6	17.7
Females	32.3	10.3	6.6
Young (15–24)			
Total	38.0	21.0	14.7
Males	43.9	27.2	20.1
Females	31.7	14.5	9.1

Journal : Small 11135 | Dispatch : 3-8-2017 | Pages : 19
Article No. : 537 | LE | TYPESET
MS Code : QUQU-D-17-00222 | CP | DISK

Levels and patterns of illicit drug use, their determinants, related behaviour and attitudes are traditionally measured through self-reporting methods of investigation. However, drug addiction is a sensitive topic that produces desirability bias and threat of disclosure, which can seriously flaw the validity of the results obtained by such methods. For this reason, the soundness of self-reported data has long been questioned (see, e.g., Harrison and Hughes 1997) and assessed by urine, blood or hair analyses. Although less intrusive survey methods, such as CATI, ACASI and CAWI, have also been used, in a bid to increase confidentiality, the results obtained continue to present errors, mostly due to misreporting. For instance, some studies show that individuals under criminal justice supervision are loath to report drug use on confidential and anonymous surveys, and others have observed that a non-negligible percentage of individuals who test positive for drugs by urinalyses deny having used drugs. Underreporting of drug consumption is therefore both evident and determined by threat of disclosure. Hence the need for alternative, indirect questioning methods to address the problem. In this respect, the RRT and its variants are increasingly employed in real-life studies of the use of drug, athletic and cognitive performance-enhancing substances. For instance, Kerkvliet (1994) used randomized response data in a logistic regression model, in which the academic performance of university students, their personal habits and socioeconomic characteristics were incorporated to estimate a logit model capable of predicting whether or not the students had consumed cocaine. Weissman et al. (1986) examined whether telephone interviewing could be a viable alternative to field interviewing as a method for eliciting drug use information. In this study, a variant of Warner's (1965) RR model was employed, and the telephone responses obtained with the RRT were compared with those obtained through DQ. Pitsch et al. (2007) used the RRT to examine whether the use of performance-assisting doping was prevalent in certain professional sports. Striegel et al. (2010) estimated the prevalence of doping and illicit drug abuse among athletes. In this study, the subjects were either asked to complete an anonymous standardized questionnaire or were interviewed using the RRT. According to this analysis, doping tests produced 0.81% positive test results, but the RRT showed that the prevalence was 6.8%. In another study, Dietz et al. (2013) reported that 20% of students used drugs in order to improve their cognitive performance. The authors concluded that the RRT revealed a high 12-month prevalence of cognitive-enhancing drug use by university students and suggested that other direct survey techniques might underestimate the use of these drugs, a fact which should be taken into consideration in the development of drug prevention programmes. Other studies related to the use of IQTs for investigating illicit drug consumption include Goodstadt and Gruson (1975), Simon et al. (2006), Stubbe et al. (2013) and Shamsipour et al. (2014).

The transition from childhood to adulthood normally marks the beginning of sexual behaviour. In this stage of life, important behavioural patterns are formed and may become lifelong. Improper sexual behaviours, too, often begin at this stage of life. In some countries, rapid economic and social changes have strongly contributed to modifying sexual culture, leading to more frequent and different forms of sexual violence (Aggleton et al. 2006) and unconventional sexual behaviour (exhibitionism, voyeurism, masturbation, pornography, cybersex, commercial sex involvement, swapping partners, anonymous or group sex, etc.). In the spectrum of problematic sexual behaviour, the impact of sexual addiction has increased notably in recent years and, for the serious psychological and social problems that it poses to sex addicts, has attracted the attention of mental health practitioners which are engaged in the assessment, diagnosis and clinical treatment of this mental disorder. Sex addiction is a chronic, relapsing disorder in which repeated, compulsive sexual stimulation persists despite serious negative consequences. Sexual arousal induces

Journal : Small 11135 | Dispatch : 3-8-2017 | Pages : 19
Article No. : 537 | ☐ LE | ☐ TYPESET
MS Code : QUQU-D-17-00222 | ☑ CP | ☑ DISK

pleasant states (euphoria in the initial phase) and relieves stress. On the other hand, it can lead to dependence, craving and relapse. In the nervous system, sexual addiction produces the same effects as cocaine, amphetamines and compulsive gambling and is dangerous in the same way as is heroin (Levine 2010). Moreover it often coexists with substance addiction (alcohol, drugs, etc.). Studies of the prevalence of sex addiction have reported questionable results, partly due to the use of imprecise subjective methods to estimate behaviour patterns, or in other cases to the use of (unreliable) self-reported survey data. To the best of our knowledge, very few studies have discussed the use of IQTs in the investigation of sexual behaviour. Among these few are LaBrie and Earleywine (2000) and Walsh and Braithwaite (2008) who used IQTs to investigate risky sexual activity. Miner (2008) explored the use of the RRT for estimating the mean number of sexual offences taking place and found that RRT estimates were significantly higher than the official figures (2.20 vs. 0.51). The use of the RR estimates was, therefore, recommended, rather than data from official records, in order to evaluate sex offender treatment interventions. Krebs et al. (2011) applied the ICT to measure the prevalence of sexual assaults. Jong et al. (2012), incorporating different RR methods, examined permissive sexual attitudes and risky sexual behaviour among samples of adults from different countries, including Spain. Geng et al. (2016), employing different RR methods for quantitative and qualitative variables, investigated the behavioural risk profile of men who had homosexual relations. This research focused on estimating the mean number of male sex partners, the mean age at first homosexual encounter and the prevalence of condom use. Srivastava et al. (2015) discussed the use of a multi-proportion RR method to assess the extent of sexual abuse among children.

3 Methodological aspects: indirect questioning techniques

In this section, we describe the methodological aspects of the data-collection techniques we used in our study to investigate cannabis consumption and sexual addiction. In particular, we illustrate the RR method proposed by Bar-Lev et al. (2004; hereafter BarLev) to scramble the responses for sensitive quantitative variables, and the IST. Our analysis is conducted under a generic sampling design in order to provide the methodological framework for obtaining estimates and variance estimation for a wide class of survey designs. It is assumed that the reader is familiar with basic sampling elements (see, e.g., Cochran 1977).

Without loss of generality, let $U = \{1, \ldots, N\}$ be a finite population consisting of N different and identifiable units. Let y_i be the value of the sensitive variable under study, namely $\mathcal{Y}$, for the ith population unit. Suppose that $\mathcal{Y}$ is quantitative and its population mean, $\bar{Y} = N^{-1}\sum_{i\in U} y_i$, is unknown and must be estimated on the basis of a sample s of fixed size n selected from U according to a generic sampling design $p(\cdot)$ which admits positive first- and second-order inclusion probabilities, $\pi_i = \sum_{s\ni i} p(s)$ and $\pi_{ij} = \sum_{s\ni i,j} p(s)$ with $i,j \in U$. For the sake of notation, let $d_i = \pi_i^{-1}$, $\check{y}_i = d_i y_i$, $\Delta_{ij} = \pi_{ij} - \pi_i\pi_j$ and $\check{\Delta}_{ij} = \Delta_{ij}/\pi_{ij}$. Under a DQ survey mode, let $\widehat{\bar{Y}}_{\mathrm{HT}}$ denote the well-known Horvitz-Thompson estimator (hereafter HT-estimator; Horvitz and Thompson 1952) of $\bar{Y}$

Journal : Small 11135 | Dispatch : 3-8-2017 | Pages : 19
Article No. : 537 | ☐ LE | ☐ TYPESET
MS Code : QUQU-D-17-00222 | ☑ CP | ☑ DISK

$$\hat{\bar{Y}}_{\mathrm{HT}} = \frac{1}{N}\sum_{i\in s}\check{y}_i. \tag{1}$$

The estimator is unbiased and has variance

$$\mathbb{V}(\hat{\bar{Y}}_{\mathrm{HT}}) = \frac{1}{N^2}\sum_{i\in U}\sum_{j\in U}\Delta_{ij}\check{y}_i\check{y}_j,$$

which can be unbiasedly estimated by

$$\widehat{\mathbb{V}}(\hat{\bar{Y}}_{\mathrm{HT}}) = \frac{1}{N^2}\sum_{i\in s}\sum_{j\in s}\check{\Delta}_{ij}\check{y}_i\check{y}_j. \tag{2}$$

3.1 The BarLev model

Let us consider a generic RR model which induces a scrambled response z_i and, hence, a revised randomized response r_i which is an unbiased estimation of y_i, $\mathbb{E}_{\mathrm{R}}(r_i) = y_i$, $\forall i \in s$ (see Chaudhuri and Christofides 2013). Then, in this RR framework, the HT-estimator for the $\bar{Y}$ takes the form

$$\hat{\bar{Y}}_{\mathrm{RRT}} = \frac{1}{N}\sum_{i\in s}\check{r}_i, \tag{3}$$

with variance

$$\mathbb{V}(\hat{\bar{Y}}_{\mathrm{RRT}}) = \frac{1}{N^2}\sum_{i\in U}d_i\mathbb{V}_{\mathrm{R}}(r_i) + \mathbb{V}(\hat{\bar{Y}}_{\mathrm{HT}}),$$

where $\check{r}_i = d_i r_i$ while $\mathbb{V}_{\mathrm{R}}(r_i)$ denotes the variance of r_i induced by the specific randomization mechanism adopted to mask the true value y_i. The variance of $\hat{\bar{Y}}_{\mathrm{RRT}}$ is unknown and can be unbiasedly estimated by

$$\widehat{\mathbb{V}}(\hat{\bar{Y}}_{\mathrm{RRT}}) = \frac{1}{N^2}\left(\sum_{i\in s}d_i\hat{v}_{\mathrm{R}}(r_i) + \sum_{i\in s}\sum_{j\in s}\check{\Delta}_{ij}\check{r}_i\check{r}_j\right), \tag{4}$$

where $\hat{v}_{\mathrm{R}}(r_i)$ denotes the estimated variance of r_i which becomes explicit only after the RR mechanism is chosen.

In order to introduce the BarLev method, let $\mathcal{S}$ denote an innocuous quantitative variable unrelated to $\mathcal{Y}$ and assume that its distribution, mean μ_s and variance σ_s^2 are all known. The BarLev procedure works as follows: with probability q the ith respondent is asked to release the true value of the sensitive variable y_i, whereas with probability $1-q$ he or she is asked to generate a number s_i from $\mathcal{S}$ and multiply it by y_i. Hence, the observed randomized response for the ith respondent will be

$$z_i = \begin{cases} y_i & \text{with probability } q \\ y_i s_i & \text{with probability } 1-q. \end{cases}$$

Here, q denotes a *design parameter* which is controlled by the researcher. Consequently, the revised response r_i under the BarLev method easily follows as

$$r_i = \frac{z_i}{q + (1-q)\mu_s},$$

and the expression of $\widehat{\widehat{Y}}_{\mathrm{RRT}}$ is determined accordingly.

It is straightforward to prove that r_i is a RR-unbiased estimator of y_i, while simple algebra yields the expression of its variance

$$\mathbb{V}_{\mathrm{RRT}}(r_i) = \frac{(1-q)[q(1-\mu_s^2) + \sigma_s^2]}{[q + (1-q)\mu_s]^2} y_i^2,$$

which is estimated by

$$\widehat{\mathbb{V}}_{\mathrm{RRT}}(r_i) = \frac{(1-q)[q(1-\mu_s^2) + \sigma_s^2]}{[q + (1-q)\mu_s]^2} r_i^2.$$

Hence, the estimated variance of the BarLev estimator easily follows.

We note that computing $\widehat{\mathbb{V}}(\widehat{\widehat{Y}}_{\mathrm{HT}})$ and $\widehat{\mathbb{V}}(\widehat{\widehat{Y}}_{\mathrm{RRT}})$ requires knowledge of the second-order inclusion probabilities for each pair of sampled units. In a complex sampling design, variance estimation may be an hard matter to deal with that, however, can be achieved by using resampling procedures like bootstrap or jackknife (see, e.g., Wolter 2007). Resampling methods for BarLev variance estimation have been recently implemented in the R package RRTCS by Cobo et al. (2015).

3.2 The item sum technique

The IST is a variant of the well-known and widely used ICT, which was proposed by Chaudhuri and Christofides (2013) to deal with quantitative sensitive variables. Due to its very recent introduction, this technique for conducting sensitive research is as yet little known among survey practitioners. Up to now, to the best of our knowledge, only Trappmann et al. (2014) used the technique to estimate the amount of undeclared work performed in Germany. Surely, it is the first time in the literature that the procedure is employed to investigate cannabis consumption and sexual addiction and compared with another indirect questioning method.

The IST, like the ICT, requires the selection of two independent samples. Therefore, with the same notation discussed above, let s_1 and s_2 be two samples of size n_1 and n_2, respectively, selected from U according to the sampling design $p(\cdot)$. Without loss of generality, assume that units belonging to s_1 are given a questionnaire with a long list (LL) of items containing $G+1$ questions of which G refer to nonsensitive characteristics and one pertains to the sensitive variable $\mathcal{Y}$ under investigation. The units sampled in s_2 are provided with a short list (SL) of items containing only the G innocuous questions present in the LL-sample. All the items refer to quantitative variables, possibly measured on the same scale as the sensitive one. The units in both samples are requested to report the total score of their answers to all the questions in their list without revealing the individual score for each question.

Let $\mathcal{T}$ be the variable denoting the total score applicable to the G nonsensitive questions, and $\mathcal{Z} = \mathcal{Y} + \mathcal{T}$ be the total score applicable to the nonsensitive questions and the sensitive question. Hence, the answer of the ith respondent will be

$$z_i = \begin{cases} y_i + t_i & \text{if } i \in s_1 \\ t_i & \text{if } i \in s_2. \end{cases}$$

Under the design $p(\cdot)$, let $\widehat{\widehat{Z}}_{\text{HT}}$ and $\widehat{\widehat{T}}_{\text{HT}}$ be the HT-estimators of $\bar{Z} = N^{-1}\sum_{i\in U}(y_i + t_i)$ and $\bar{T} = N^{-1}\sum_{i\in U} t_i$, respectively. Hence, a HT-type estimator of $\bar{Y}$ under the IST can be easily obtained as

$$\widehat{\widehat{Y}}_{\text{IST}} = \widehat{\widehat{Z}}_{\text{HT}} - \widehat{\widehat{T}}_{\text{HT}}. \tag{5}$$

From the unbiasedness of $\widehat{\widehat{Z}}_{\text{HT}}$ and$\widehat{\widehat{T}}_{\text{HT}}$, it readily follows that the estimator $\widehat{\widehat{Y}}_{\text{IST}}$ is unbiased for $\bar{Y}$. The variance of $\widehat{\widehat{Y}}_{\text{IST}}$, as long as the two samples s_1 and s_2 are independent, can be expressed as

$$\mathbb{V}(\widehat{\widehat{Y}}_{\text{IST}}) = \frac{1}{N^2}\left(\sum_{i\in U}\sum_{j\in U}\Delta_{ij}\check{z}_i\check{z}_j + \sum_{i\in U}\sum_{j\in U}\Delta_{ij}\check{t}_i\check{t}_j\right) \tag{6}$$

and unbiasedly estimated by

$$\widehat{\mathbb{V}}(\widehat{\widehat{Y}}_{\text{IST}}) = \frac{1}{N^2}\left(\sum_{i\in s_1}\sum_{j\in s_1}\check{\Delta}_{ij}\check{z}_i\check{z}_j + \sum_{i\in s_2}\sum_{j\in s_2}\check{\Delta}_{ij}\check{t}_i\check{t}_j\right), \tag{7}$$

where the meaning of $\check{z}_.$ and $\check{t}_.$ is clear.

4 Estimating patterns of cannabis consumption and sexual addiction: some evidence from a real study

In this section, we describe the results obtained and the salient aspects of a mixed-mode survey conducted in two Spanish universities to investigate patterns of cannabis consumption and sexual addiction. In particular, we aim to evaluate the effectiveness of the above-described IQTs in comparison with the DQ survey mode. It should be noted that these two topics have different degrees of sensitivity. While the use of cannabis is widely accepted nowadays and is commonly experienced by younger people, unconventional sexual behaviour is much more sensitive and continues to represent a taboo for young people.

4.1 The survey design

The survey was carried out at the universities of Granada and Murcia during the academic year 2015/2016. The data-collection and the fieldwork were performed by the FQM356 research group as part of the Andalusian Research Plan, University of Granada.

A stratified sample of 2398 students enrolled in different faculties were selected such that degree programs and year of degree were represented in proportion to their total numbers of students.

Moving from Trappmann et al. (2014), and from some budget, time and fieldwork constraints, we firstly decided to recruit 500 students by the DQ method and then to oversize the samples of students to assign to the BarLev and the IST survey modes due to the lower statistical power of the two IQTs. In particular, the size of the sample to be

Journal : Small 11135 | Dispatch : 3-8-2017 | Pages : 19
Article No. : 537 | ☐ LE | ☐ TYPESET
MS Code : QUQU-D-17-00222 | ☑ CP | ☑ DISK

surveyed by using the BarLev method was increased at a ratio of 1.20 to 1 (DQ) while the size of the IST sample was increased at a ratio of 2.5 to 1 (DQ) in order to have enough students to assign to the LL-sample and SL-sample. Additionally, we increased the size of the LL-sample size at a ratio of 1.5 to 1 (SL-sample) in order to compensate for the larger variability of the estimates. The students were contacted in class and randomly assigned to one of the three survey modes. Some extra students, recruited in a second moment during an academic event, were added to the survey and assigned to the BarLev method (25%) and the IST (75%). At the end of the fieldwork, 492 students were surveyed using DQ, 613 using the BarLev method and 1293 with the IST (773 in the LL-sample and 520 in the SL-sample). To motivate students' participation, the scientific nature of the survey was emphasized. No incentives of any kind were provided. The questionnaires were distributed during the class time break to the students who provided signed informed consent to participate in the study. The classroom setting facilitated cooperation and no objection to the survey was raised.

Except for some differences stemming from the different ways of providing the sensitive information, all students received the same questionnaire covering academic items and personal characteristics. The sensitive questions for the DQ survey mode and the experimental section for the IQTs were positioned at the end of the questionnaire.

In the DQ survey mode, the questionnaire had a block containing four sensitive questions:

Q1: How many cannabis cigarettes did you consume last year?
Q2: Over the past 90 days, how many days did you consume cannabis?
Q3: Over the past 90 days, how many times have you had trouble stopping your sexual behaviour when you knew it was inappropriate?
Q4: Over the past 90 days, how many times has sex been an escape from your problems?

Questions Q1 and Q2 concerning cannabis consumption were taken from the questionnaire on drug addiction given in Miller and Rollnick (2015), while the sensitive questions Q3 and Q4 referring to sexual behaviour were freely adapted from Carnes' Sexual Addictions Screening Test (Carnes 1989).

To collect sensitive information using the BarLev method, we used as a randomizing device the smartphone application of the "Baraja Española", a deck composed of 40 cards, divided into four families or suits, each numbered from 1 to 7, and three figures for the each suit. The students assigned to this method were requested to install the application on their smartphone. The application is very simple to use: the user touches the screen and a card is shown. For each sensitive question, the students were asked to run the application and to give the true sensitive response if the card shown was a figure. If the screen did not show a figure, the students were asked to multiply the real sensitive value of the response by the number shown on the card. In this way, the design parameter q of the BarLev model was set to $q = 3/10$. All the explanations on how to proceed were clearly given in the questionnaire and a blank space was provided in which to write the responses.

For the IST, four different nonsensitive questions, each corresponding to one of the sensitive questions, were formulated. For cannabis use, the student "Selectivity" mark[1] was used as an innocuous variable. Hence, the students who were assigned to the IST received two different questionnaires, depending on whether they belonged to the SL-

[1] The Selectivity mark is the score obtained in the university entrance examination. It is computed by summing the marks of two phases, the general and the specific. The general phase consists of four tests, and is scored from 0 to 10. The specific phase consists of two tests and is scored from 0 to 4.

Journal : Small 11135 | Dispatch : 3-8-2017 | Pages : 19
Article No. : 537 | ☐ LE | ☐ TYPESET
MS Code : QUQU-D-17-00222 | ☑ CP | ☑ DISK

sample or the LL-sample. The IST described in Sect. 3.2 was repeated four times by the students, one run for each of the sensitive questions Q1–Q4.

The students in the SL-sample received the questionnaire with the following innocuous questions:

IQ1: What was your general mark in the Selectivity exam, without counting specific subjects? (Value between 0 and 10)
IQ2: What was your Selectivity mark counting specific subjects? (Value between 0 and 14)
IQ3: What is the number of subjects in which you have enrolled in the academic year?
IQ4: What is the final digit of your mobile phone number?

The students in the LL-sample received a questionnaire with text explaining the IST procedure followed by a block consisting of pairs of questions, the sensitive question and the corresponding nonsensitive question. More precisely, the sensitive question Q1 was paired with the innocuous question IQ1, Q2 with IQ2, Q3 with IQ3, and Q4 with IQ4. For each pair of questions, the students were asked to report the sum of the scores of the two questions, without revealing the individual responses.

For both the BarLev method and the IST, when the questionnaires were distributed, the students were assured of the confidentiality of their responses. It was emphasised that the investigators would not be able, from the responses given, to discover their true status with respect to the sensitive characteristic being investigated, since they would not know which card was generated by the mobile application or the individual score to the LL-questions. Moreover, in order to reassure the students and to maximize response rates, it was stressed that, although individual responses could not be used to infer any personal and confidential status, the responses of all of them could be used to produce collective knowledge of the phenomena under study.

4.2 Results

In this section we present and analyze the results of our research. The main aim is the show how the reported amount of the four investigated sensitive characteristics depend on the data-collection method. Given the sensitive nature of the issues in question, we expected a systematic underreporting of cannabis use and sexual behaviour in the DQ survey. Hence, according to the “more-is-better” assumption (Lensvelt-Mulders et al. 2005), the data-collection method that provided higher estimates of the sensitive characteristics was considered to be the more valid one.

Table 2 Percentage nonresponse rates for DQ, BarLev and IST survey modes

Question	Direct questioning	BarLev method	Item Sum Technique
Q1	10.96	14.03	$1.93^{**}_{\star\star\star}$
Q2	11.79	4.40^{*}	1.55^{**}
Q3	21.14	6.69^{**}	$0.15^{***}_{\star\star}$
Q4	16.67	6.20^{**}	$0.23^{***}_{\star\star}$

One-tailed t-test for difference between two proportions: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$ for IQTs versus DQ, and $^{\star}p < 0.05$, $^{\star\star}p < 0.01$, $^{\star\star\star}p < 0.001$ for IST versus BarLev

The first notable result that emerges from the study is the significant reduction in the nonresponse rate in the case of the IQTs. Table 2 shows the nonresponse rates (in percentage) for the four questions under the three data-collection methods. As expected, the DQ nonresponse rate is higher for questions Q3 and Q4 than for Q1 and Q2. This is probably due to the fact that sexual matters are much more confidential and intrusive of the personal sphere than are patterns of cannabis consumption, among university students. Both IQTs obtained a significantly higher level of cooperation than the DQ method, except the BarLev model for Q1. There was a remarkable reduction in the nonresponse rate for question Q3, which seems to be the most sensitive one. Comparison of the two IQTs reveals that the IST nonresponse rate for questions Q1, Q3 and Q4 is statistically lower than that of the BarLev method. In general, the IST yielded a very low nonresponse rate, no more than 2% for any of the questions.

Table 3 summarizes the main results of our study. It includes the estimated means of the number of cannabis cigarettes smoked in the last year, of days during the past 90 in which cannabis was consumed, of number of times during the past 90 days that students had difficulty in halting inappropriate sexual behaviour and of the number of times during the past 90 days when sex was used to escape from personal problems. To get the estimates, the estimators $\widehat{\widehat{Y}}_{\mathrm{HT}}$, $\widehat{\widehat{Y}}_{\mathrm{RRT}}$ and $\widehat{\widehat{Y}}_{\mathrm{IST}}$ given in (1), (3) and (5) were applied under the proportional-allocation stratified sampling design. The estimated standard error of the estimators was calculated from expressions (2), (4) and (7), together with the 95% Wald confidence interval (CI) for the unknown means and the length (L) of the interval. The normality of the estimates under the three survey methods was ascertained by investigating the sampling distribution of the estimators using a bootstrap resampling procedure.

The estimates are reported for the entire sample and for subgroups by gender (males and females). Prior to this analysis, we assessed whether the random assignment of the students to the three survey modes produced comparable groups of respondents by gender. The Chi-squared test of independence confirmed the effectiveness of the random assignment.

The results obtained reflect the impact of the different survey methods on the estimates. As expected, the DQ method produced an underestimation of the sensitive characteristics investigated. Thus, the DQ estimates were statistically lower than the IQT ones, apart from question Q4 under the BarLev method, where no statistical evidence of underreporting was ascertained. The BarLev estimates were statistically higher than the IST ones for questions Q2 and Q3, and lower for question Q4, while no significant difference was ascertained for question Q1. Therefore, according to the “more-is-better” assumption, both of the IQTs outperform the DQ method, but there is no evidence of a uniform superiority of one indirect questioning method over the other.

We note that the lower limit of the confidence interval for direct question Q2 in the female group was negative. This does not make sense, of course. Nonetheless, we observe that there is sufficient statistical evidence to consider that the estimated mean was not significantly different from zero. For the remaining cases, the confidence intervals obtained under the three methods show that all the estimates were different from zero.

With respect to accuracy, the IST estimates presented lower standard errors and narrower confidence intervals than the BarLev method, except for question Q4. As expected, the DQ estimates were more precise than the IQT ones, except for question Q4. The latter, in fact, are in general affected by an extra source of variability induced by masking the responses, other than that inherent to the sampling design.

An in-depth analysis of these results indicates that patterns of sexual addiction are present in the population of students, with a slight predominance in the male group. The

Journal : Small 11135 Dispatch : 3-8-2017 Pages : 19
Article No. : 537 ☐ LE ☐ TYPESET
MS Code : QUQU-D-17-00222 ☑ CP ☑ DISK

Table 3 Mean estimates and accuracy measures under the three data-collection methods

Question	Direct questioning					BarLev method					Item sum technique				
	n (%)	Mean	SD	CI	L	n (%)	Mean	SD	CI	L	n (%)	Mean	SD	CI	L
Total															
Q1	492 (100%)	3.11	0.60	[1.93; 4.30]	2.37	613 (100%)	13.07^{**}	3.38	[6.43; 19.70]	13.27	1293 (100%)	14.93^{***}	2.53	[9.97; 19.89]	9.93
Q2		1.41	0.40	[0.63; 2.19]	3.05		9.33^{***}	1.28	[6.82; 11.84]	5.02		$3.72^{***}_{\star\star\star}$	0.47	[2.80; 4.65]	1.84
Q3		0.23	0.07	[0.10; 0.36]	0.26		2.12^{***}	0.42	[1.31; 2.94]	1.63		$1.11^{***}_{\star\star}$	0.29	[0.53; 1.68]	1.15
Q4		2.52	0.66	[1.23; 3.81]	2.58		3.46	0.55	[2.38; 4.53]	2.14		$7.60^{***}_{\star\star\star}$	0.70	[6.24; 8.97]	2.73
Males															
Q1	211 (42.89%)	6.35	1.43	[3.54; 9.15]	5.61	252 (41.11%)	21.14^{*}	7.20	[7.03; 35.25]	28.22	597 (46.17%)	24.65^{***}	4.69	[15.47; 33.84]	18.37
Q2		2.23	0.76	[0.74; 3.72]	2.98		8.85^{***}	1.67	[5.58; 12.12]	6.54		$5.51^{***}_{\star}$	0.81	[3.92; 7.09]	3.17
Q3		0.48	0.17	[0.15; 0.81]	0.66		2.73^{**}	0.90	[0.97; 4.48]	3.51		1.94^{***}	0.57	[0.82; 3.07]	2.25
Q4		3.98	1.26	[1.51; 6.44]	4.93		3.65	0.91	[1.87; 5.43]	3.56		$8.16^{***}_{\star\star\star}$	1.05	[6.10; 10.22]	4.11
Females															
Q1	281 (57.11%)	0.25	0.12	[0.01; 0.49]	0.48	361 (53.83%)	7.91^{**}	3.06	[1.90; 13.91]	12.01	689 (53.58%)	6.48^{***}	2.34	[1.89;11.06]	9.17
Q2		0.82	0.49	[−0.14;1.78]	1.92		9.76^{***}	1.85	[6.13;13.39]	7.26		$2.17^{**}_{\star\star\star}$	0.52	[1.15;3.18]	2.03
Q3		0.07	0.03	[0.02;0.12]	0.10		1.75^{***}	0.37	[1.03;2.47]	1.44		$0.39^{***}_{\star\star\star}$	0.17	[0.01;0.77]	0.76
Q4		1.86	0.83	[0.23;3.50]	3.26		3.25	0.68	[1.91;4.60]	2.68		$7.08^{***}_{\star\star\star}$	0.93	[5.26;8.91]	3.65

One-tailed t-test for differences in means: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$ for IQTs versus DQ, and $^{\star}p < 0.05$, $^{\star\star}p < 0.01$, $^{\star\star\star}p < 0.001$ for IST versus BarLev

Journal : Small 11135 | Dispatch : 3-8-2017 | Pages : 19
Article No. : 537 | ☐ LE | ☐ TYPESET
MS Code : QUQU-D-17-00222 | ☑ CP | ☑ DISK

BarLev method indicates that, on average, 2.12 times during the 90 days prior to the survey, students had difficulty in halting inappropriate sexual behaviour (2.73 times for the males and 1.75 times for the females). The IST estimates suggest a more frequent use of sex to escape from personal problems, on average 7.6 times in the 90 days prior to the survey (8.16 times for the males and 7.08 times for the females). Similar patterns were found regarding the consumption of cannabis. According to the IQTs, on average, during the last year, the students smoked around 14 cannabis cigarettes, much higher than the figure of roughly 3 cigarettes obtained by the DQ method. According to the BarLev method, male students smoked more cigarettes than female students (21.14 vs 7.91). Moreover, the students on average consumed cannabis on 9.33 days during the 90 days prior to the survey (8.85 days for the males and 9.76 days for the females).

Unfortunately, directly comparable benchmark data are not available for the phenomena investigated in this study. Nonetheless, there are very appreciable differences between the traditional DQ survey method and the IQTs. From the recent Informe 2016 survey[2] conducted in Spain during 2014 among secondary school students (aged 14–18 years) we know that the mean number of days of cannabis consumption in the last month before the survey is roughly 1 for the entire target population, 1.32 for males and 0.69 for females. It is worth noting that these estimates, obtained using an anonymous self-administered questionnaire, are very close to those obtained in the present study with the DQ method. We suggest, therefore, that they may underestimate the real values.

5 Optimal IST allocation

We conclude this article by acknowledging a recent advance in the IST which is of interest for practical purposes and that, when our research was being planned, had not been known. In general, a key problem in conducting ICT/IST surveys is how to determine the size of the LL-sample and SL-sample. The LL-sample is generally larger than the SL-sample in order to compensate for the variability introduced in the estimates by the nonsensitive variable(s). This problem was recently investigated by Perri et al. (2017), who proposed for the IST a rule for optimally allocating the sample units between the LL-sample and SL-sample.

In this section, by simulating some scenarios from the previous real data-based study, we explore the effectiveness of the optimal allocation. Following the notation set out in Sect. 3.2, the idea of the optimal allocation is first to consider a sample s of size n and then to optimally split it into two sub-samples, s_1 and s_2, in such a way as to maximize the efficiency of $\widehat{\bar{Y}}_{\text{IST}}$ or, equivalently, to minimize the variance of the estimator given in (6). According to this criterion, after some algebra, optimal sample size allocation in simple random sampling is given by

$$n_1^{\text{opt}} = n\frac{S_z}{S_z + S_t}, \quad n_2^{\text{opt}} = n\frac{S_t}{S_z + S_t}, \tag{8}$$

with $n_1^{\text{opt}} + n_2^{\text{opt}} = n$ while $S_{\cdot}$ denotes the population standard error of the variables $\mathcal{Z}$ and $\mathcal{T}$ which is unknown and has to be estimated, for instance, on the basis of a training sample or a pilot survey.

[2] Informe 2016. Encuesta sobre uso de drogas en enseñanzas secundarias en España (ESTUDES). 1994–2014. Observatorio español de la droga y las toxicomanías. Ministerio de Sanidad, Servicios Sociales e Igualdad. Available at: http://www.pnsd.msssi.gob.es/profesionales/sistemasInformacion/informesEstadisticas.

Journal : Small 11135 | Dispatch : 3-8-2017 | Pages : 19
Article No. : 537 | LE | TYPESET
MS Code : QUQU-D-17-00222 | CP | DISK

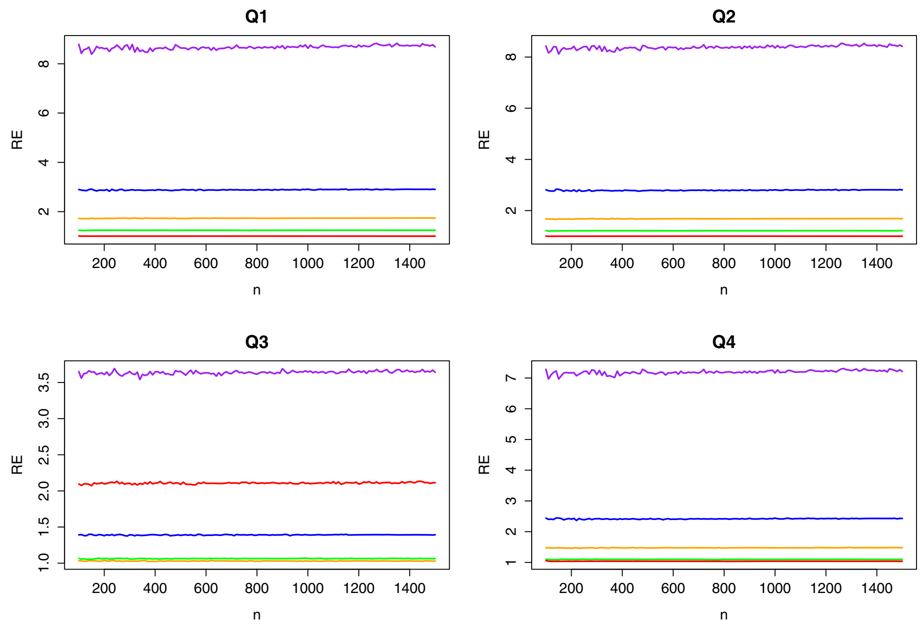

Fig. 1 Relative efficiency of optimal allocation w.r.t. arbitrary allocation for different values of α under simulated populations for questions Q1–Q4; $\alpha = 0.1, 0.3, 0.5, 0.7, 0.9$ (*violet*, *blue*, *orange*, *green*, *red*). (Color figure online)

5.1 Simulation study

We investigated optimal allocation under the IST by means of a simulation study with the aim to show the efficiency gain upon the estimates that can derive from wisely choose the size of the LL-sample and SL-sample. The first step in this study was to generate four artificial populations on the basis of the surveyed variables discussed in Sect. 3. Then, the estimated variances of the optimal IST estimates were compared with those stemming from an arbitrary allocation.

The simulation design is summarized in the following steps:

1. Generate an artificial population U of $N = 50{,}000$ sensitive values y_i from a normal distribution with mean and variance μ_{DQ} and σ^2_{DQ} computed on the sample of students assigned to the DQ survey method;
2. Generate N nonsensitive values t_i from an independent normal distribution with mean and variance μ_{SL} and σ^2_{SL} computed on the SL-sample of students;
3. Compute the total scores $z_i = y_i + t_i$, $i = 1, \ldots, N$;
4. Select a simple random sample from U of size n and split it to obtain IST estimates according to: (i) optimal allocation as given in (8); and (ii) arbitrary allocation defined as $n_1 = \alpha n$ and $n_2 = (1-\alpha)n$, with $\alpha \in (0, 1)$;
5. Compute the estimated variance of the estimator $\widehat{\overline{Y}}_{\text{IST}}$ under optimal and arbitrary allocations, that is, $\widehat{\mathbb{V}}(\widehat{\overline{Y}}^{\text{opt}}_{\text{IST}})$ and $\widehat{\mathbb{V}}(\widehat{\overline{Y}}^{\alpha}_{\text{IST}})$;
6. Repeat $B = 1000$ times the previous two steps and compute the mean ($\mathbb{E}_B$) of the estimated variances over the B replications, and hence compute the Relative Efficiency

Journal : Small 11135 | Dispatch : 3-8-2017 | Pages : 19
Article No. : 537 | ☐ LE | ☐ TYPESET
MS Code : QUQU-D-17-00222 | ☑ CP | ☑ DISK

$$\mathrm{RE} = \frac{\mathbb{E}_B\left[\widehat{\mathbb{V}}\left(\widehat{\overline{Y}}{}^{\alpha}_{\mathrm{IST}}\right)\right]}{\mathbb{E}_B\left[\widehat{\mathbb{V}}\left(\widehat{\overline{Y}}{}^{\mathrm{opt}}_{\mathrm{IST}}\right)\right]};$$

7. Run the simulation for each of the four variables referred to by questions Q1–Q4 (see Sect. 4.1).

The outcomes of the simulation study are graphically summarized in Fig. 1, where the behaviour of the relative efficiency is shown for different sample sizes and different values of α. We observe that the efficiency gain derived from the optimal allocation may be considerable, for all the variables investigated. Accordingly, future applications of the IST could benefit from this methodological advance.

6 Conclusions

This article discusses the salient aspects of a mixed-mode survey conducted among Spanish university students to investigate the frequency of certain behaviours concerning cannabis consumption and sexual addiction. Given the sensitive nature of the topics investigated, and in order to reduce nonresponse rates and obtain more truthful responses, the traditional DQ method based on anonymous self-administered questionnaires was supported by two IQTs, namely the randomized response method proposed by Bar-Lev et al. (2004), and the IST (Chaudhuri and Christofides 2013; Trappmann et al. 2014). The three data-collection methods were compared and their effects evaluated in terms of the reduction in nonresponse rates, and improvements upon the estimates according to the "more-is-better" assumption.

As expected, the DQ survey mode produced nonresponse rates that were higher than the IQT ones. In turn, the IST nonresponse rates were lower than the BarLev ones. Moreover, the DQ method produced underreporting of the sensitive behaviours under study—cannabis use and sexual addiction—and the IST estimates appeared to be more accurate than the BarLev values.

When significant underreporting is produced by DQ, researchers and practitioners actively engaged in organizing, managing and conducting sensitive studies should be suspicious about the validity of results. At the same time, operators and policy makers should proceed cautiously in the implementation of intervention programmes because the social and health problems stemming from drug consumption and sexual behaviour may be much more significant than is apparent from DQ self-reporting. The use of IQTs, as shown by this research, may provide a better understanding of the problems and help to carefully evaluate the potential extent of the phenomena under study. Even if the two methods considered are not the *panacea* for all the problems encountered in sensitive research, and may provoke mistrust among respondents, they should nevertheless represent a wake-up call for researchers and government agencies engaged in sensitive surveys.

We conclude by remarking upon the strength of this research, which provided practical experience of the two IQTs and contributed to empirically evaluating their effectiveness. The results obtained seem to be promising and we hope that can contribute to a more widespread appreciation of the benefits offered by IQTs to the scientific community in general and to survey practitioners in particular.

Journal : Small 11135 Dispatch : 3-8-2017 Pages : 19
Article No. : 537 ☐ LE ☐ TYPESET
MS Code : QUQU-D-17-00222 ☑ CP ☑ DISK

Acknowledgements This work is partially supported by Ministerio de Economía y Competitividad (grant MTM2015-63609-R, Spain), Ministerio de Educación, Cultura y Deporte (grant FPU, Spain) and by the project PRIN-SURWEY (grant 2012F42NS8, Italy).

References

Aggleton, P., Ball, A., Mane, P. (eds.): Sex, Drugs and Young People: International Perspectives. Routledge, London (2006)

Bar-Lev, S.K., Bobovitch, E., Boukai, B.: A note on randomized response models for quantitative data. Metrika **60**, 255–260 (2004)

Carnes, P.: Contrary to Love: Helping the Sexual Addict. CompCare Publisher, Minnesota (1989)

Chaudhuri, A., Christofides, T.: Indirect Questioning in Sample Surveys. Springer, Berlin (2013)

Cobo, B., Rueda, M., Arcos, A.: RRTCS: Randomized Response Techniques for Complex Surveys. R package version 0.0.3. http://cran.r-project.org/web/packages/RRTCS/ (2015)

Cochran, W.: Sampling Techniques. Wiley, New York (1977)

De Jong, M.G., Pieters, R., Stremersch, S.: Analysis of sensitive questions across cultures: an application of multigroup item randomized response theory to sexual attitudes and behavior. J. Personal. Soc. Psychol. **103**, 543–64 (2012)

Dickson-Swift, V., James, E.L., Liamputtong, P.: Undertaking Sensitive Research in the Health and Social Sciences: Managing Boundaries, Emotions and Risks. Cambridge University Press, New York (2008)

Dietz, P., Striegel, H., Franke, A.G., Lieb, K., Simon, P., Ulrich, R.: Randomized response estimates for the 12-month prevalence of cognitive-enhancing drug use in university students. Pharmacotherapy **33**, 44–50 (2013)

European Monitoring Centre for Drugs and Drug Addiction: European Drug Report 2016: Trends and Developments. Retrieved from http://www.emcdda.europa.eu/publications/edr/trends-developments/2016 (2016)

Geng, G.Z., Gao, G., Ruan, Y.H., Yu, M.R., Zhou, Y.H.: Behavioral risk profile of men who have sex with men in Beijing, China: results from a cross-sectional survey with randomized response techniques. Chin. Med. J. **129**, 523–529 (2016)

Goodstadt, M.S., Gruson, V.: The randomized response technique: a test on drug use. J. Am. Stat. Assoc. **70**, 814–818 (1975)

Groves, R.M., Fowler, F.J., Couper, M.O., Lepkowski, J.M., Singer, E., Tourangeau, R.: Survey Methodology. Wiley, Hoboken (2004)

Harrison, L., Hughes, A.: The validity of self-reported drug use: improving the accuracy of survey estimates. NIDA Res. Monogr. **167**, 1–16 (1997)

Horvitz, D.G., Thompson, D.J.: A generalization of sampling without replacement from a finite universe. J. Am. Stat. Assoc. **47**, 663–685 (1952)

Johnson, R.M., Fairman, B., Gilreath, T., Xuan, Z., Rothman, E.F., Parnham, T., Furr-Holden, C.D.: Past 15-year trends in adolescent marijuana use: differences by race/ethnicity and sex. Drug Alcohol Depend. **1**, 155–158 (2015)

Kerkvliet, J.: Estimating a logit model with randomized data: the case of cocaine use. Aust. J. Stat. **36**, 9–20 (1994)

Krebs, C.P., Lindquist, C.H., Warner, T.D., Fisher, B.S., Martin, S.L., Childers, J.M.: Comparing sexual assault prevalence estimates obtained with direct and indirect questioning techniques. Violence Against Women **17**, 219–235 (2011)

LaBrie, J.W., Earleywine, M.: Sexual risk behaviors and alcohol: higher base rates revealed using the unmatched-count technique. J. Sex Res. **37**, 321–326 (2000)

Lamarine, R.J.: Marijuana: modern medical chimera. J. Drug Educ. **42**, 1–11 (2012)

Lensvelt-Mulders, G.J.L.M., Hox, J.J., van der Heijden, P.G.M., Mass, C.J.M.: Meta-analysis of randomized response research: thirty-five years of validation. Sociol. Methods Res. **33**, 319–348 (2005)

Levine, S.B.: What is sexual addiction? J. Sex Marital Ther. **36**, 261–275 (2010)

Liamputtong, P.: Researching the Vulnerable: A Guide to Sensitive Research Methods. Sage Publications, London (2007)

Miller, J.D.: (1984). A New Survey Technique for Studying Deviant Behavior. Ph.D. Thesis, The George Washington University

Miller, W.R., Rollnick, S.: Motivational Interviewing: To Prepare for Change of Addictive Behaviors. The Guilford Press, London (2015)

Miner, M.H.: Improving the measurement of criminal sexual behavior. The application of randomized response technique. Sex. Abuse **20**, 88–101 (2008)

Journal : Small 11135 Dispatch : 3-8-2017 Pages : 19
Article No. : 537 □ LE □ TYPESET
MS Code : QUQU-D-17-00222 ✔ CP ✔ DISK

Perri, P.F., Rueda, M., Cobo, B.: Multiple sensitive estimation and optimal sample size allocation in the item sum technique. Biom. J. (in press) (2017). doi:10.1002/bimj.201700021

Pitsch, W., Emrich, E., Klein, M.: Doping in elite sports in Germany: Results of a www survey. Eur. J. Sport Soc. **4**, 89–102 (2007)

Simon, P., Striegel, H., Aust, F., Dietz, K., Ulrich, R.: Doping in fitness sports: estimated number of unreported cases and individual probability of doping. Addiction **101**, 1640–1644 (2006)

Shamsipour, M., Yunesian, M., Fotouhi, A., Jann, B., Rahimi-Movaghar, A., Asghari, F., Akhlaghi, A.A.: Estimating the prevalence of illicit drug use among students using the crosswise model. Subst. Use Misuse **49**, 1303–1310 (2014)

Srivastava, R., Nigam, A.K., Singh, N.: Application of randomized response techniques in estimation of prevalence of child sexual abuse. Stat. Appl. **13**, 37–45 (2015)

Striegel, H., Ulrich, R., Simon, P.: Randomized response estimates for doping and illicit drug use in elite athletes. Drug Alcohol Depend. **106**, 230–232 (2010)

Stubbe, J.H., Chorus, A.M.J., Frank, L.E., de Hon, O., van der Heijden, P.G.M.: Prevalence of use of performance enhancing drugs by fitness center members. Drug Test. Anal. **6**, 434–438 (2013)

Tian, G.-L., Tang, M.-L.: Incomplete Categorical Data Design: Non-Randomized Response Techniques for Sensitive Questions in Surveys. Chapman & Hall/CRC, Boca Raton (2014)

Tourangeau, R., Smith, T.: Asking sensitive questions: the impact of data collection, question format, and question context. Public Opin. Q. **60**, 275–304 (1996)

Tourangeau, R., Yan, T.: Sensitive questions in surveys. Psychol. Bull. **133**, 859–883 (2007)

Trappmann, M., Krumpal, I., Kirchner, A., Jann, B.: Item sum: a new technique for asking quantitative sensitive questions. J. Surv. Stat. Methodol. **2**, 58–77 (2014)

Walsh, J.A., Braithwaite, J.: Self-reported alcohol consumption and sexual behavior in males and females: Using the unmatched-count technique to examine reporting practices of socially sensitive subjects in a sample of university students. J. Alcohol Drug Educ. **52**, 49–72 (2008)

Warner, S.L.: Randomized response: a survey technique for eliminating evasive answer bias. J. Am. Stat. Assoc. **60**, 63–69 (1965)

Weissman, A.N., Steer, R.A., Lipton, D.S.: Estimating illicit drug use through telephone interviews and the randomized response technique. Drug Alcohol Depend. **18**, 225–233 (1986)

Wolter, K.: Introduction to Variance Estimation. Springer, New York (2007)