Title: Measuring inappropriate sexual behavior among university students: using the randomized response technique to enhance self-reporting

Abstract

This paper analyzes the efficacy of the randomized response technique (RRT) in achieving honest self-reporting about sexual behavior, compared to traditional survey techniques. A complex survey was conducted of 1246 university students in Spain, who were asked sensitive quantitative questions about their sexual behavior, either via the RRT (n= 754) or by direct questioning (DQ) (n=492). The RRT estimates of the number of times that the students were unable to restrain their inappropriate sexual behavior were significantly higher than the DQ estimates, among both male and female students. The results obtained suggest that the RRT method elicits higher values of self-stigmatizing reports of sexual experiences by increasing privacy in the data collection process. The RRT is shown to be a useful method for investigating sexual behavior.

Key words: Sexual behavior*,* Social desirability, Randomized response techniques, Complex surveys.

Introduction

Sexual behavior is difficult to study empirically because of its sensitive nature. The prevalence and frequency of sexual behavior are difficult to estimate by standard survey techniques because respondents tend to withhold information in such settings. Social desirability bias (the wish to make a favorable impression) poses a significant threat to the validity of self-reports, particularly when they concern sensitive information related to sexual activities (Tan and Grace, 2008).

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, generally known as indirect questioning techniques (IQT), 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.

The randomized response technique (RRT) is one of the more commonly used methods of indirect questioning. Since Warner's pioneering work on indirect questioning in 1965, the RRT has maintained a prominent position in this field of research. Fundamentally, the RRT employs (at least in its original formulation) a physical randomization device (decks of cards, colored numbered balls, dice, coins, spinners, random number generators, etc.) which determines whether respondents should answer the sensitive question or another, neutral question or even provide a pre-specified response (e.g. "yes") irrespective of how they would answer concerning the sensitive behavior.

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 the researcher and, hence respondents' privacy is preserved. Although the individual information provided by the respondents in accordance with the RRT cannot be used to discover their true status regarding the sensitive issues, the data compiled from all the survey participants can be profitably employed to draw inferences on parameters of interest regarding the study population, after making certain transformations in the original variables (see e.g. Arcos et al., 2015)

In recent years, indirect questioning techniques have become popular as an effective means of eliciting more honest responses to sensitive questions. The RRT was first developed by Warner (1965) to allow researchers to obtain sensitive information while guaranteeing privacy to respondents. This method encourages greater cooperation from respondents and reduces their motivation to report falsely. The most important claim made for RRT is that it yields more valid point estimates of sensitive behavior. Compared with standard direct questioning, areas of sensitive behavior are more frequently reported if questions are posed via the RRT (Williams and Suen, 1994). Warner's study generated a large body of research literature on alternative RRTs for eliciting sensitive information (Arnab, 2002; 2004; Bouza, 2009; Clark and Desharnais, 1998; Diana and Perri, 2010; 2012; Santiago, Bouza and Al-Omari, 2016; Singh and Sedory, 2011; Ulrich, Schröter, Striegel and Simon, 2012) in studies of socially undesirable behavior. The potential generalization of the technique to other cultures has been demonstrated by studies related to China (Geng, Gao, Ruan, Yu, and Zhou, 2016), Iran (Vakilian, Mousavi, and Keramat, 2014), the USA (Walsh and Braithwaite, 2008), Europe (Perri, Pelle, and Stranges, 2015) and Spain (Cobo, Rueda, and López-Torrecillas, 2017), all of which observe that surveys based on indirect questioning are commonly used when the questions relate to sensitive issues.

The RRT and its variants have been applied to examine a great variety of subjects, including the use of recreational drugs or of athletic or cognitive performance-enhancing substances (Goodstadt and Gruson, 1975; Kerkvliet, 1994; Shamsipour et al., 2014; Simon, Striegel, Aust, Dietz, and Ulrich, 2006; Striegel, Ulrich, and Simon, 2010), the impact of HIV/AIDS infection in Botswana (Arnab and Singh, 2010), the prevalence of induced abortion in the United States, Mexico, Botswana, Taiwan, and Turkey (Lara, García, Ellerton, Camlin, and Suarez, 2006; Oliveras and Letamo, 2010)

and the prevalence of induced abortion and of irregular immigrant status among foreign women in Calabria, Italy (Perri, Pelle, and Stranges, 2015).

With respect to studies of sexual behavior, LaBrie and Earleywine (2000) and Walsh and Braithwaite (2008) used indirect questioning methods to investigate risky sexual activity, and De Jong, Pieters, and Stremersch (2012) compared non-student samples from two countries in an analysis of permissive sexual attitudes and risky sexual behavior, using a RRT. These studies highlight the benefits of using the RRT to investigate sensitive issues like sexual behavior, when it might otherwise be very difficult to obtain accurate responses.

Specifically, LaBrie and Earleywine (2000) used indirect questioning to estimate base rates for risky sexual behavior after drinking, and compared their findings with those obtained by conventional methods. The indirect questioning approach revealed significantly higher base rates than a conventional self-report survey with respect to having had sex, having sex without a condom, and having sex without a condom after drinking. Subsequently, Walsh and Braithwaite (2008) examined the same issues of alcohol consumption and sexual behavior, in a study based on a sample of 842 students at a university in the American Midwest, and concluded that indirect questioning techniques produced higher rates of honest self-reporting than traditional survey techniques.

Geng, Gao, Ruan, Yu, and Zhou (2016) described the behavioral risk profile of men who have sex with men (MSM) in Beijing, using three RRTs. According to the responses made, and in comparison with statistics obtained by surveys based on anonymous questionnaires or direct interviews, the MSM population in the RRT-based surveys started their first sexual encounter at a younger age, had more male partners, and reported a lower rate of consistent condom use during anal sex with male partners.

These results suggest that the RRT might be a useful tool to obtain more honest feedback from respondents on sensitive information such as sexual behavior.

Krebs et al. (2011) hypothesized that the validity of self-reported data on sexual assault might be open to doubt, if victims were reluctant to disclose what had happened to them. In this study, using an anonymous, web-based survey, a sample of undergraduate women were asked, by direct and indirect questioning, about their experiences of physically forced sexual assault. The results obtained via indirect questioning reflected a slightly higher prevalence in this respect than those produced by direct questioning. However, the difference was not statistically significant. These results suggest that either direct questioning yields reasonably valid estimates of the prevalence of sexual assault or that the item count technique does not produce estimates that are any more valid. Miner (2008) used the RRT in a study of 424 men who were imprisoned for child sexual abuse or rape, and reported that the RRT estimates of prior offending (2.20 prior offenses) were significantly higher than officially recorded prior offenses (0.51 prior offenses). The numbers of prior sex offenses were obtained from participants' prison records at the time of recruitment screening. Miner found preliminary evidence that RRT is a useful method for generating data of a sensitive nature, when official records might be inaccurate.

In the present study, direct and indirect questioning methods are used to explore reports of inappropriate sexual behavior by university students. We assume that the study's focus on stigmatizing behavior will generate a social desirability bias, and so higher rates will be obtained by RRT than by direct questioning. To our knowledge, such an analysis of social desirability bias in a survey of sexual behavior has not previously been performed in Spain.

Method

To determine whether the RRT elicits better estimates of problematic sexual behavior than methods based on direct questioning, we conducted a survey of university students in this respect, using both approaches. The students were all informed about the aims of the study and provided signed informed consent. Ethical approval was obtained from the Research Ethics Committee of the University of Granada. In the following sections, we report how the sample size was determined, describe the procedures applied, and provide full details of the measures considered. All the analyses performed (whether they yielded significant or non-significant results) are presented.

Participants and sampling method

The sample population for this survey was composed of students from two universities in the regions of Andalusia and Murcia (southern Spain). A stratified sample of 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. For the purposes of this study, a cluster was assumed to be approximately equal to the size of a class, in each of the universities. Clusters were randomly chosen from the university classes, and all members of the class, both female and male, were included in every case.

In our study, we determinate the sample size to estimate the population mean with a coefficient of variation of 0.25, we decided that 500 students would be asked to a respond to a survey based on direct questioning (DQ) and another 625, to one based on the RRT. The numbers were greater in the latter case because RRT inherently has less statistical power.

The data collection and the field work were conducted by the FQM365 research group, part of the Andalusian Research Plan. The interviews were carried out during 2015. Students in 16 classes were contacted and randomly assigned to one of the two survey modes: RRT (subsample 1, n=754, 60.2% female and 79.3% aged 21 years or younger, given no missing data) and direct response (subsample 2, n=492, 57.1% female and 87.4% aged 21 years or younger, given no missing data). The dataset is available on the Open Science Framework platform at https://osf.io/823nw/?view_only=56833a3d255941e783d9d2922fafd5f0.

Procedure and measures

All questionnaires were administered on paper during the class break time. All students were invited to participate in the study, and provided signed informed consent. The classroom setting facilitated cooperation, no objection to the survey was raised and no empty questionnaires were returned.

The questionnaire was the same in the two subsamples. It began with some academic questions followed by a set of basic demographic questions and then a sensitive question referring to sexual behavior, taken from the "Sexual Dependency Inventory-Revised" (SDI-R; Carnes & Delmonico, 1996). This screening test is a broad and comprehensive assessment designed to help researchers in the study of sexual addiction. In our study, the sensitive question used was: *Over the past 90 days, how many times have you had trouble stopping your sexual behavior when you know it is inappropriate?* In subsample 1 (using the RRT), for the sensitive question, the interviewer explained how the survey was being conducted, and gave an example of its use. The response was randomized using a generalization of the model proposed by

Bar-Lev, Bobovitch, and Boukai (2004) for simple random sampling and later extended by Arcos, Rueda, and Singh (2015) for use with complex samples.

Instructions were given for the following procedure to be carried out: if the respondent was in subsample 2 (using DQ), the sensitive question should be answered directly. If the respondent was in subsample 1 (using the RRT), he/she was asked to install the Baraja Española ("*Deck of Cards*") app in their phone. This app (available from Google Play Store, 2015) presents one of 40 cards, divided into four suits, each numbered 1-7, or one of three figures, and was used as a randomization mechanism with which to answer the sensitive question. If the card presented was a figure (for example, a king), the respondent should answer the sensitive question correctly; if the card showed a number, the respondent should answer the sensitive question multiplied by the number shown on the card (for example, a 5). Figure 1 shows the response procedure for the two subsamples.

Figure 1 here

The researcher explained that this technique preserved the students' anonymity and that the aim was to avoid provoking mistrust. Following this explanation, all students completed the full questionnaire. On the contrary, in subsample 2, with a questionnaire based on direct questioning, the survey was not fully completed by all of the respondents: the total nonresponse rate was 13%. In the next section, we present and discuss the nonresponse rates.

Statistical analysis

In survey sampling, inference is used to estimate the parameters of interest. In the present case, because the card provided by the app is randomly selected, and as we know the probability of a figure and of each number appearing, we can estimate the

overall response of the RRT group (but not that of each individual) and compare it to the DQ response. The weights were computed from a stratified clustered random design, modified to adjust for coverage bias. All statistical analyses were performed using the sampling weights. The Horvitz-Thompson estimator (Singh, 2003) was used to estimate the mean values for the direct questions. In RR with the BarLev technique, the unbiased generalized estimator proposed by Arcos et al. (2015) was used to estimate the mean values of the study variable (see Appendix 1). In both cases, the values were calibrated (by sex and age) to the population totals. The calibration adjustments applied to the mean estimators enhanced the validity and accuracy of the process. All statistical analyses were performed using R software (version R-3.3.3), with standard packages for estimation in survey sampling (*Sampling;* Tillé and Matei, 2015) and a specific package for handling RR data obtained from complex surveys (*RRTCS;* Rueda, Cobo and Arcos, 2015). Specifically, in this package the BarLev() function was used. The mean values obtained by DQ and RRT, and those for each of the subgroups (both by DQ and by the RRT), were compared using the method proposed by Wolter and Preisendörfer (2013) (see Appendix 2).

Results

Because the samples do not reproduce the gender composition of the population, we have reweighted the sample weights by calibration on gender.

Due to the sensitive nature of the research topic, we expected that a considerable share of respondents would refuse to participate in the survey or would underreport the sensitive behavior. However, the RRT has been found to be easy to understand and trusted to ensure the anonymity of the respondents, so that we are interested in knowing whether the response rates obtained when applying RRT are lower than in DQ.

Tables 1 here

Table 1 includes the nonresponse rates to the sensitive question for the full sample and for the sample according to gender and age, by direct and by randomized response. The nonresponse rate for the question was significantly lower in the RRT group (p-value <.0001). Among the RRT group, the difference between male and female students was also significant (p-value <0.05). However, in the DQ group, the responses of male and female students presented no statistically significant differences (p-value>0.05). With respect to age, we also found significant differences in nonresponse rates between the RRT and DQ groups (p-value <.0001). In the latter, the nonresponse rate for respondents aged ≤21 years was significantly lower than that of older students (p-value <.0001). In the RRT group, there was no such significant difference according to the students' age (p-value >0.05).

The point estimates of the sensitive question and the 95% confidence intervals for each technique (DQ and RR) are summarized in Tables 2 and 3.

Tables 2 and 3 here

For the DQ group, the estimated mean number of days involving difficulty controlling inappropriate sexual behavior was 0.23, in contrast to the 1.45 days for the RR group (p-value <.0001). Analysis by gender showed that the mean number of days was higher according to RR than by DQ, both for men (p-value <.0001) and for women (p-value <.0001). Focusing on DQ, the mean number of days was higher for men than for women (p-value =0.03443). This was also so with RR, but in this case the difference was not significant (p-value =0.2473). While RR is arguably less prone to bias than DQ, it is also more susceptible to sampling variability (since as well as the sampling variance we must also take into account the variance arising from the randomization process, see formula 2 in Appendix 1). When the results are considered by age, the

difference between DQ and RR is also statistically significant (see Table 2). Both genders reported a low number of days when questioned directly (0.48 for men and 0.0686 for women), while indirect methods revealed substantially higher mean values in this respect (2.0445 for men and 1.1558 for women).

Discussion

The present study describes a procedure for asking sensitive quantitative questions about sexual behavior, which was applied in a survey conducted among 1246 students at two Spanish universities, in Granada and Murcia. Respondents were randomly selected to be questioned about their sexual behavior, either directly or by the RR technique. The calibration adjustments applied to these estimators increased their validity and accuracy. The results of the direct survey were then compared with those of the randomized response survey. The mean number of days that subjects reported having difficulty controlling their inappropriate sexual behavior was much higher according to the RRT responses than with the DQ technique. 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 (although, unfortunately, we do not have real values obtained by a census of inappropriate sexual behavior among university students with which to compare our findings).

The nonresponse rate was significantly lower in the RR than in the DQ group, which corroborates previous research in this field, such as Goodstadt and Gruson (1975), Geng et al. (2016), and Cobo et al. (2017).

Tourangeau and Yan (2007) suggested that misreporting about sensitive topics is common and largely situational, its extent depending on whether the respondent has anything embarrassing to report, and on the design features of the survey. The RRT is designed to decrease social desirability bias and thus obtain more reliable estimates (Arnab and Singh, 2010; De Jong et al., 2012; Geng et al., 2016; Krebs et al., 2011; Lara et al., 2006; Oliveras and Letamo, 2010; Miner, 2008; Perri et al., 2015; Shamsipour et al., 2014; Simon et al., 2006; Striegel et al., 2010). Our results are consistent with those obtained by De Jong, Pieters, and Stremersch (2012), who compared nonstudent samples from two countries on permissive sexual attitudes and risky sexual behavior, using a RRT, and highlighted the advantages of using RRT in sensitive issues like sexual behavior, where it can be very difficult to obtain accurate responses.

In many contexts, negative social attitudes towards sensitive issues such as aggressive sexual behavior can result in false and invalid data being provided in self-reported surveys. In this respect, De Jong et al. (2012) and Walsh and Braithwaite (2008) indicated that men´s self-reported sexually aggressive behavior should be interpreted with care, as it is highly subject to social desirability bias and so tends to be underreported. Other researchers, too, have raised concerns about the distortions that can be introduced into standardized measures by the effects of social desirability bias (Tourangeau and Yan, 2007; Schlachter and Rolf, 2016).

Limitations

The present study has certain limitations that should be acknowledged. The sample population was drawn from university students in one specific country, and the results obtained may not be generalizable to other social or cultural contexts. Therefore, further research should be undertaken, seeking to replicate our results in alternative

settings. Another limitation of the present study is the size of the sample: we believe it was sufficient for the DQ estimation but possibly not for precise analysis with the RRT.

Our study aim was to estimate the mean values of the variable in question (responses to a sensitive question) and to determine whether they were affected by social desirability bias in the population, or in any of its subgroups. However, we did not address other, more complex statistical analyses, such as regression, which could also be applied to the RRT (see e.g. Blair, Imai, & Zhou, 2015).

Conclusions

Our findings support the view that the randomized response technique enables researchers to obtain sensitive information while guaranteeing privacy to respondents. It encourages cooperation from respondents and reduces their motivation to falsely report their attitudes. The most important claim made for RRT is that it yields more valid point estimates of sensitive behavior. Various types of sensitive behavior are more frequently reported if respondents are questioned via the RRT, rather than directly. In our study, conducted in Spain (where RR techniques are not commonly used for studies of sexual behavior), quantitative variables were taken into account in order to make the scope of the study as complete as possible.

Nonresponse is a common problem in surveys, especially when sensitive issues are investigated. It is generally agreed that use of the RRT can increase respondents' degree of cooperation and thus reduce the rate of nonresponse (Barabesi et al., 2014). The results of our study show that the RRT achieves a significant improvement in response rates, which is in line with previous research findings.

On the other hand, the RR approach also has certain drawbacks. Firstly, individual response patterns cannot be interpreted directly, due to the observation of

randomized responses, nor can individuals or groups of individuals be compared. Moreover, RR procedures require a randomization device to drive the answer. Using physical devices may be more time consuming and costly than direct questioning. The variance of the estimates is also increased by the randomization mechanism, although the use of auxiliary information at the estimation stage can help reduce this variance without additional costs and without infringing respondents' privacy. Recent contributions in this field have been proposed by Özgül and Cingi (2017) and by Rueda et al. (2018), among others.

As a final observation, the RRT could be used not only when the sensitive question is about engaging in sexually aggressive behavior, but also in treatment. Thus, treatment providers could use the RRT to evaluate whether participants are reporting sexual thoughts or fantasies about re-offending, this being an area of inquiry that is highly sensitive and likely to be minimized in direct questioning.

References

Arcos, A., Rueda, M. & Singh, S. A. (2015). Generalized approach to randomized response for quantitative variables. *Quality and Quantity*, 49, 1239-1256. doi:http://dx.doi:10.1007/s11135-014-0046-3.

Arnab, R. (2002). Optimum sampling strategies under randomized response surveys. *Biometrical Journal,* 44, 1–16. doi:http://dx.doi.org/10.1002/1521-4036(200206)44:4<490:AID-BIMJ490>3.0.CO;2-T

Arnab, R. (2004). Optional randomized response techniques for complex survey designs. *Biometrical Journal,* 46, 114–124. doi:http://dx.doi:10.1002/bimj.200210006.

Arnab, R. & Singh, S. (2010). Randomized response techniques: An application to the Botswana AIDS impact survey.*Journal of Statistical Planning and Inference*, 140, 941- 953. doi: http://dx.doi.org/10.1016/j.jspi.2009.09.019.

Bar-Lev S.K., Bobovitch, E. & Boukai, B. (2004). A note on randomized response models for quantitative data. *Metrika*, 60, 255-260. doi:http://dx.doi: 10.1007/s001840300308.

Blair, G., Imai, K. & Zhou, Y. (2015). Design and analysis of the randomized response technique. *Journal of the American Statistical Association*, 110, 1304-1319. doi:http://dx.doi:10.1080/01621459.2015.1050028

Bouza, C. (2009). Ranked set sampling and randomized response procedures for estimating the mean of a sensitive quantitative character. *Metrika,* 70, 267-277. doi:http://dx.doi: 10.1007/s00184-008-0191-6.

Carnes, P. & Delmonico, D. (1996). Sexual Dependency Inventory-Revised. Minneapolis, MN. Positive Living Press.

Clark, S. J. & Desharnais, R. A. (1998). Honest answers to embarrassing questions: Detecting cheating in the randomized response model. *Psychological Methods, 3*(2), 160-168. doi:http://dx.doi.org/10.1037/1082-989X.3.2.160.

Cobo, B., Rueda, M., & López -Torrecillas, F. (2017). Application of randomized response techniques for investigating cannabis use by Spanish university students. *International Journal of Methods in Psychiatric   Research.* 26:e1517.https://doi.org/10.1002/mpr.1517.

De Jong, M.G., Pieters, R. & Stremersch, S. (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*, 103,543-564. doi:http://dx. doi: 10.1037/a0029394.

Diana, G. & Perri, P.F. (2010). New scrambled response models for estimating the mean of a sensitive quantitative character. *Journal of Applied Statistics*, 37**,** 1875–1890.doi:http://dxdoi: 10.1080/02664760903186031.

Diana, G. & Perri, P.F. (2012). A calibration-based approach to sensitive data: a simulation study. *Journal of Applied Statistics*, 39, 53-65. doi:http://dx.*doi*.org/10.1080/02664763.2011.578615.

Geng, G.Z., Gao, G., Ruan, Y.H., Yu, M.R & Zhou, Y.H. (2016). Behavioral risk profile of men who have sex with men in Beijing, China: Results from a cross-sectional survey with randomized response techniques. *Chinese Medical Journal,* 129(5):523-529.doi:http://dx.doi: 10.4103/0366-6999.177003.

Goodstadt, M.S. & Gruson, V. (1975). The randomized response technique: A test on drug use. *Journal of the American Statistical Association*, 70, 814-818. doi:http://dx*doi*:10.1017/S037689290999004X.

Google Play Store (2015). Download the application "Baraja Española". https://play.google.com/store/apps/details?id=cartas.barajaes.

Kerkvliet, J. (1994). Estimating a logit model with randomized data: The case of cocaine use. *Australian Journal of Statistics*, 36, 9-20. doi:http://dx.doi: 10.1111/j.1467-842X.1994.tb00634.x

Krebs, C.P., Lindquist, C.H., Warner, T.D., Fisher, B.S., Martin, S.L. & Childers, J.M. (2011). Comparing sexual assault prevalence estimates obtained with direct and indirect questioning techniques. *Violence Against Women,* 17(2) 219–235. doi:http://dx.doi: 10.1177/1077801210397743.

LaBrie, J. W. & Earleywine, M. (2000). Sexual risk behaviors and alcohol: higher base rates revealed using the unmatched-count technique. *Journal of Sex Research*, 37(4), 321–326. doi:http://dx.doi:10.1080/00224490009552054.

Lara, D., García, S.G., Ellertson, C., Camlin, C. & Suárez, J. (2006). The measure of induced abortion in Mexico using random response technique. *Sociological Methods & Research* 35:279-301. doi:http://dx.doi:10.1177/0049124106290442.

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

Miner, M.H. (2008). Improving the measurement of criminal sexual behaviors. The application of randomized response technique. *Sexual Abuse*, 20 (1): 88-101. doi:http://dx.doi: 10.1177/1079063208314816.

Özgül, N. & Cingi, H. (2017). A new estimator based on auxiliary information through quantitative randomized response techniques, *Journal of Modern Applied Statistical Methods,* 16, 364-387.doi:http://dx.doi.10.22237/jmasm/1493598060

Oliveras, E. & Letamo, G. (2010). Examples of methods to address underreporting of induced abortion: preceding birth technique and randomized response technique. In

Methodologies for Estimating Abortion Incidence and Abortion-Related Morbidity: A Review, Singh, S. & Remez, L. & Tartaglione, A. (Eds.), The Guttmacher Institute – International Union for the Scientific Study of the Population.

Perri, P.F., Pelle, E. & Stranges, M. (2015). Estimating induced abortion and foreign irregular presence using the randomized response crossed model. *Social Indicators Research*, 1-18. Published online. doi:http://dx.doi: 10.1007/s11205-015-1136-x.

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

Rueda, M., Cobo, B. & Arcos, A. (2018). An improved class of estimators in RR surveys, *Mathematical Methods in the Applied Sciences*, 41, 2307–2318

Santiago, A., Bouza, C. & Al-Omari, I. (2016). Randomized response procedure for the estimation of the population ratio using ranked set sampling, *Journal of Mathematics and Statistics,* 12, 107-114. doi:http://dx.doi: 10.3844/jmssp.2016.107.114.

Schlachter, S. & Rolf, S. (2016). Using the IAT: how do individuals respond to their results? *International Journal of Social Research Methodology,* 8, 1-16. doi:http://dx.doi.org/10.1080/13645579.2015.1117799.

Shamsipour, M., Yunesian, M., Fotouhi, A., Jann, B., Rahimi-Movaghar, A., Asghari, F. & Akhlaghi, A.A. (2014). Estimating the prevalence of illicit drug use among students using the crosswise model. *Substance Use & Misuse,* 49, 1303-1310. doi:http://dx.doi: 10.3109/10826084.2014.897730.

Simon, P., Striegel, H., Aust, F., Dietz, K. & Ulrich R. (2006). Doping in fitness sports: Estimated number of unreported cases and individual probability of doping. *Addiction.* 101, 1640-1644. doi:http://dx.doi: 10.1111/j.1360-0443.2006.01568.x.

Singh, S. (2003). *Advanced Sampling Theory with Applications. How Michael 'selected' Amy*. Springer, Netherlands.

Singh, S. & Sedory, S.A. (2011). Cramer-Rao lower bound of variance in randomized response sampling. *Sociological Methods and Research,* 40(3) 536–546. doi:http://dx.doi: 10.1177/0049124111416596.

Striegel, H., Ulrich, R. & Simon, P. (2010). Randomized response estimates for doping and illicit drug use in elite athletes. *Drug and Alcohol Dependence*, 15, 230-232. doi:http://dx.doi: 10.1016/j.drugalcdep.2009.07.026.

Tan, L., & Grace, R.C. (2008). Social desirability and sexual offenders: a review. *Sexual Abuse: A Journal of Research and Treatment,* 20(1):61-87. doi:http://dx.doi: 10.1177/1079063208314820.

Tillé, Y. & Matei, A. (2015). R Package sampling: survey sampling. Available from: https://cran.r-project.org/web/packages/sampling/sampling.pdf.

Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin*, 133(5):859-883. doi:http://dx.doi.org/10.1037/0033-2909.133.5.859.

Ulrich, R., Schröter, H., Striegel, H. & Simon, P. (2012) Asking sensitive questions: A statistical power analysis of randomized response models. *Psychological Methods*, 17(4): 623-641.doi:http://dx.doi.org/10.1037/a002931.

Vakilian, K., Mousavi, S.V. & Keramat, A. (2014). Estimation of sexual behavior in the 18-to-24-years-old Iranian youth based on a crosswise model study. *BMC Research Notes,* 7-28.doi:http://dx.doi.org/10.1186/1756-0500-7-28.

Walsh, J.A. & Braithwaite, J. (2008). 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. *Journal of Alcohol and Drug Education,* 52(2), 49–72. Retrieved from https://www.ncjrs.gov/App/Publications/abstract.aspx?ID=245966.

Warner, S.L. (1965). Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association,* 60, 63-69.

Williams, R.I. & Suen, H. (1994). A methodological comparison of survey techniques in obtaining self-reports of condom-related behaviors. *Psychological Reports,* 75 (3), 1531-1537. doi:http://dx.doi: 10.2466/pr0.1994.75.3f.1531.

Wolter, F., Preisendörfer P. (2013) Asking sensitive questions: an evaluation of the randomized response technique versus direct questioning using individual validation data. *Sociological Methods & Research* 42 (3), 321-353.

Appendix 1

Let z_i denote the randomized response obtained from the i −th student. In RR using the BarLev technique, the RR given by person i is expressed as

$$z_i = \begin{cases} y_i, & \text{with probability } p \\ y_i S, & \text{with probability } 1 - p \end{cases}$$

where S is a scramble variable, with a mean $\mu = 4$ and a standard deviation $\sigma = 2$.

The transformed variables are defined as

$$r_i = \frac{z_i}{p + (1-p)\mu}$$

with $p = 0.3$ and with a variance that is expressed as

$$\hat{\phi}_i = \frac{1}{(p + (1-p)\mu)^2} (r_i^2 A)$$

where

$$A = p(1-p) + \sigma^2(1-p) + \mu^2(1-p) - \mu^2(1-p)^2 - 2p(1-p)\mu$$

The Horvitz-Thompson estimator of the mean for the RR survey is given by

$$\hat{\bar{Y}}_{ht}(r) = \frac{1}{N} \sum_{i \epsilon s} di\ ri \qquad (1)$$

and the unbiased estimator of the variance of this estimator is

$$\hat{V}_{ht}(r) = \frac{1}{N^2} \left[\frac{1}{2} \sum_{i \neq} \sum_{j \epsilon s} \frac{(\pi_i \pi_j - \pi_{ij})}{\pi_{ij}} \left(\frac{r_i}{\pi_i} - \frac{r_j}{\pi_j} \right)^2 + \sum_{i \epsilon s} \frac{\hat{\phi}_i}{\pi_i} \right] \qquad (2)$$

where π_i and π_{ij} are the inclusion probabilities (Singh, 2003) of the i -th unit and of the ith and jth units $(i \neq j)$ respectively and d_i is the sampling weight of the i -th unit. First and second-order inclusion probabilities are computed from this stratified cluster random sampling.

The confidence interval at $(1 - \alpha)\%$ is given by

$$ci = \left(\hat{\bar{Y}}_{ht}(r) - z_{1-\frac{\alpha}{2}}\sqrt{\hat{\bar{V}}_{ht}(r)}, \hat{\bar{Y}}_{ht}(r) + z_{1-\frac{\alpha}{2}}\sqrt{\hat{\bar{V}}_{ht}(r)} \right)$$

where $z_{1-\frac{\alpha}{2}}$ denotes the $(1 - \alpha)\%$ quantile of a standard normal distribution.

Appendix 2

To test the equality of means, given that the variance of the RRT includes the randomization term, we use the methodology proposed by Wolter and Preisendörfer (2013) and we calculate z scores using the equation:

$$z = \frac{\hat{\bar{Y}}_{ht}(r) - \hat{\bar{Y}}_{ht}(y)}{\sqrt{\hat{\bar{V}}_{ht}(r) + \hat{\bar{V}}_{ht}(y)}}$$

where

$\hat{\bar{Y}}_{ht}(y) = \frac{1}{N}\sum_{i\epsilon s} d_i\, y_i$ and $\quad \hat{\bar{V}}_{ht}(y) = \frac{1}{N^2}\left[\frac{1}{2}\sum_{i \neq}\sum_{j \in s} \frac{(\pi_i\pi_j - \pi_{ij})}{\pi_{ij}}\left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j} \right)^2 \right] \quad$ are the estimator and the variance based on the DQ response.
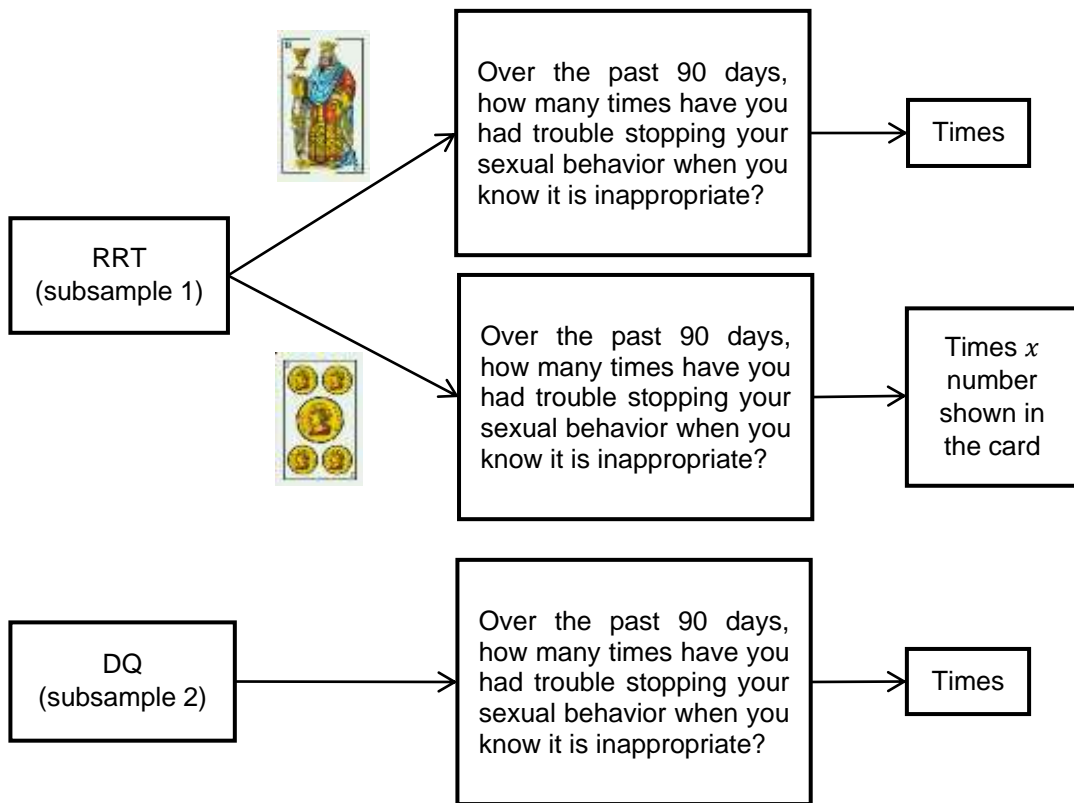
Figure 1: Procedure of response for the two samples

Table 1: Nonresponse rates of sample

	DQ (n=492)	RR (n=754)	Statistic	P_value DQ vs RR
Total	21.1%	3.7%	95.439	<.0001
Gender				
Male	38.9%	6.0%	84.301	<.0001
Female	34.9%	2.2%	147.15	<.0001
Statistic	0.82574	7.3221		
P_value male vs female	0.3635	0.00681		
Age				
<=21	19.6%	3.3%	71.401	<.0001
>21	58.5%	5.2%	66.257	<.0001
Statistic	31.858	1.2414		
P_value <=21 vs >21	<.0001	0.2652		

Table 2. Estimations refer to the number of times the participant had engaged in inappropriate sexual behavior over the past 90 days.

Study technique	DQ (n=492)				RR (n= 754)			
	Estimation	Standard deviation	Confidence Interval (95%)		Estimation	Standard deviation	Confidence Interval (95%)	
			Lower bound	*Upper bound*			*Lower bound*	*Upper bound*
Total	0.2319	0.0584	0.1174	0.3464	1.4537	0.1900	1.0812	1.8262
Gender								
Male	0.4800	0.1382	0.2091	0.7508	2.0445	0.4433	1.1755	2.9136
Female	0.0686	0.0220	0.0254	0.1118	1.1558	0.1527	0.8565	1.4551
Age								
<= 21	0.2457	0.0717	0.1050	0.3863	1.5660	0.2368	1.1018	2.0302
> 21	0.1052	0.0731	0.0381	0.2486	1.3261	0.2669	0.8030	1.8492

Table 3. *P*-values DQ vs RR

	P-values	Cohen's d
Total	<.0001	8.022
Male	<.0001	4.399
Female	<.0001	9.090
<= 21	<.0001	6.961
> 21	0.0011	5.741