University of Granada

Doctorate Program in Mathematical and Applied Statistics

Doctoral Thesis

Using auxiliary information in indirect questioning techniques

Beatriz Cobo Rodríguez

Thesis supervised by

Prof. María del Mar Rueda García

Prof. Pier Francesco Perri

Granada, April, 2018

Editor: Universidad de Granada. Tesis Doctorales
Autor: Beatriz Cobo Rodríguez
ISBN: 978-84-9163-957-2
URI: http://hdl.handle.net/10481/52559

Contents

Agradecimientos

En primer lugar me gustaría dar las gracias a mis directores de tesis, a María del Mar Rueda García por introducirme en el mundo de la investigación y hacerlo parte de mi vida, por compartir conmigo sus conocimientos, y por su apoyo y ánimo incondicionales, en definitiva por ser un ejemplo a seguir de mujer trabajadora que se ha ganado a pulso todo lo que es hoy en día, a Pier Francesco Perri por acogerme en su Universidad y hacerme sentir como en casa, y por demostrarme que el idioma y la distancia no es una barrera para investigar juntos, sin olvidar a Antonio Arcos Cebrián, que a pesar de no ser mi director, siempre ha estado dispuesto a ayudarme cuando lo he necesitado.
Al Departamento de Estadística e Investigación Operativa por brindarme la oportunidad de comenzar la investigación y la docencia, y a mis compañeros, por hacerme los días más amenos.
Por último, pero no menos importante, a mi familia y amigos, por animarme y apoyarme en todo momento y creer en mí durante este largo camino.

Nota: este trabajo de investigación se ha desarrollado al amparo de una beca para la Formación del Profesorado Universitario (FPU) concedida por el Ministerio de Educación, Cultura y Deporte así como de los proyectos de investigación liderados por María del Mar Rueda García y Antonio Arcos Cebrián en los cuales he sido partícipe.

Summary

A survey is a research method that is based on questioning a sample of individuals. The interest in sample surveys studies often focuses on sensitive or confidential aspects of the interviewees. Because of this, the typical problem that arises is social desirability, which is defined as the tendency of respondents to answer based on what is socially acceptable. For this reason, many respondents refuse to participate in the survey or provide false or conditioned answers, altering the accuracy and reliability of the estimations in a major way.

Randomized response (RR) technique (RRT) introduced by Warner is a possible solution for protecting the anonymity of the respondent and is used to reduce the risk of escape or no response to sensitive questions. Warner's study generated a rapidly-expanding body of research literature on alternative techniques for eliciting suitable RR schemes in order to estimate a population proportion. Standard RR methods are used primarily in surveys which require a binary response to a sensitive question, and seek to estimate the proportion of people presenting a given sensitive characteristic. On the other hand, some studies have addressed situations in which the response to a sensitive question results in a quantitative variable.

The methodology of RR has advanced considerably in recent years, but the most research in this area concerns only simple random sampling and the real studies are based on complex surveys.
Data from complex survey designs require special consideration with regard to estimation for parameters and corresponding variance estimation. Recently some authors have developed R-packages for estimation with RR surveys under the assumption on simple random sampling. In order to estimate parameters for sensitive characteristics, no existing software covers the estimation of these procedures from complex surveys. This gap is now filled by RRTCS package. The package includes the estimators for means and totals with several RR techniques and also provides confidence interval estimation.

Most research into RRT deals exclusively with the interest variable and does not make explicit use of auxiliary variables in the construction of estimators. We introduce auxiliary variables for a general class of estimators to improve sampling design and to achieve higher precision in population parameter estimates.

Warner's work originated a huge literature and has been used in many areas, but these techniques have difficulties and limitations. Due to this, other indirect techniques emerged as an alternative to RRT, among them we find the item count technique (ICT). This technique was conceived for surveys which require the study of a qualitative variable, but many practical situations may deal with sensitive variables which are quantitative in nature. So, the item sum technique (IST) was proposed as a generalization of ICT.

To contribute to the development of the IST in real-world studies, we suggest some methodological advances, as IST estimation under a generic sampling design, the use of auxiliary information to improve the efficiency of the estimates and we extend this calibration approach to the estimation for domain. We also investigate the impact on the estimates of including an increased number of innocuous questions in the list of items.

Traditionally, indirect questioning techniques (IQTs) deal with one sensitive variable. However, in real surveys, the researcher may be interested in investigating more than one sensitive variable. We discuss some estimation methods for multiple sensitive questions under different approaches.

A key design decision in an IST survey is how to split the total sample into the long list sample and short list sample. A simple solution is to allocate the same number of units to each sample irrespective of the variability of the items in the two lists. Clearly, this intuitive and basic solution is not efficient because responses in the long list sample are tendentially affected by high variability due to the presence of innocuous items. We achieve the optimal sample size allocation by minimizing the variance of the IST estimates under a budget constraint. Optimal allocation results are finally extended to the multiple sensitive estimation setting.

Finally, we use the IQTs for investigating some sensitive variables, drug addiction, sexual behavious and support for female genital cutting, in real studies and we compare these results with those get by direct question, obtaining in all cases higher estimates of the sensitive characteristics when we use IQTs.

Note: This thesis is presented as a compendium of seven publications in relation with the contents of the thesis. The full version of the papers is included in Appendices A1 - A7.

Resumen

Una encuesta es un método de investigación que se basa en cuestionar una muestra de individuos. El interés en los estudios de encuestas por muestreo a menudo se centra en aspectos sensibles o confidenciales para los entrevistados. Debido a esto, el problema típico que surge es la deseabilidad social, que se define como la tendencia de los encuestados a responder en función de lo que es socialmente aceptable. Por esta razón, muchos encuestados se niegan a participar en la encuesta o proporcionan respuestas falsas o condicionadas, lo que altera la precisión y fiabilidad de las estimaciones de manera importante.

La técnica de respuesta aleatoria (RRT) introducida por Warner es una posible solución para proteger el anonimato del encuestado y se usa para reducir el riesgo de evasión o la falta de respuesta ante preguntas delicadas. El estudio de Warner se extendió rapidamente y generó un cuerpo de literatura de investigación sobre técnicas alternativas para obtener esquemas de RR adecuados a fin de estimar una proporción de la población. Los métodos de RR estándar se utilizan principalmente en encuestas que requieren una respuesta binaria a una pregunta delicada, y buscan estimar la proporción de personas que presentan una característica sensible dada. Por otra parte, algunos estudios han abordado situaciones en las que la respuesta a una pregunta sensible da como resultado una variable cuantitativa.

La metodología de RR ha avanzado considerablemente en los últimos años, pero la mayoría de las investigaciones en este área se refieren únicamente al muestreo aleatorio simple y los estudios reales se basan en encuestas complejas.

Los datos de los diseños de encuestas complejas requieren una consideración especial con respecto a la estimación de los parámetros y la estimación de la varianza correspondiente. Recientemente, algunos autores han desarrollado paquetes de R para la estimación con encuestas de RR bajo el supuesto de muestreo aleatorio simple. Con el fin de estimar parámetros para características sensibles, ningún software existente cubre la estimación de estos procedimientos a partir de encuestas complejas. Esta brecha ahora se llena con el paquete RRTCS. El paquete incluye los estimadores de medias y totales con varias técnicas de RR y también proporciona una estimación del intervalo de confianza.

La mayoría de las investigaciones sobre la RRT tratan exclusivamente con la variable de interés y no hacen un uso explícito de variables auxiliares en la construcción de estimadores. Nosotros introducimos variables auxiliares para una clase general de estimadores para mejorar el diseño de muestreo y lograr una mayor precisión en las estimaciones de los parámetros de la población.

El trabajo de Warner originó una gran cantidad de literatura y se ha utilizado en muchas áreas, pero estas técnicas tienen dificultades y limitaciones. Debido a esto, surgieron otras técnicas indirectas como alternativa a la RRT, entre ellas encontramos la técnica de conteo de ítems (ICT). Esta técnica fue concebida para encuestas que requieren el estudio de una variable cualitativa, pero muchas situaciones prácticas pueden tratar con variables sensibles que son de naturaleza cuantitativa. Para ello se propuso la técnica de suma de ítems (IST) como una generalización de las ICT.

Para contribuir al desarrollo de la IST en estudios reales, sugerimos algunos avances metodológicos, como la estimación de la IST bajo un diseño de muestreo genérico, el uso de información auxiliar para mejorar la eficiencia de las estimaciones y ampliamos este enfoque de calibración a la estimación por dominios. También investigamos el impacto en las estimaciones al incluir un número mayor de preguntas inocuas en la lista de ítems.

Tradicionalmente, las técnicas de interrogación indirectas (IQTs) tratan con una variable sensible. Sin embargo, en encuestas reales, el investigador puede estar interesado en estudiar más de una variable sensible. Discutimos algunos métodos de estimación para preguntas sensibles múltiples bajo diferentes enfoques.

Una decisión clave de diseño en una encuesta de IST es cómo dividir la muestra total en la muestra de lista larga y la muestra de lista corta. Una solución simple es asignar el mismo número de unidades a cada muestra, independientemente de la variabilidad de los elementos en las dos listas. Claramente, esta solución intuitiva y básica no es eficiente porque las respuestas en la muestra larga se ven afectadas tendencialmente por una alta variabilidad debido a la presencia de elementos inocuos. Nosotros logramos la asignación óptima del tamaño de la muestra al minimizar la varianza de las estimaciones de la IST bajo una restricción presupuestaria. Los resultados óptimos de asignación finalmente se extienden a la estimación sensible múltiple.

Finalmente, utilizamos las IQTs para investigar algunas variables sensibles, drogadicción, comportamiento sexual y apoyo para el corte genital femenino, en estudios reales y comparamos estos resultados con los alcanzados mediante pregunta directa, obteniendo en todos los casos mayores estimaciones de las características sensibles cuando utilizamos las IQTs.

Nota: Esta tesis se presenta como un compendio de 7 publicaciones relacionadas con los contenidos de la tesis. La versión íntegra de los artículos se incluye en los Apéndices A1 - A7.

Part I

Chapter 1

Introduction

In many fields of applied research, and particularly in sociological, economic, demographic, ecological and medical studies, the investigator very often has to gather information concerning highly personal, sensitive, stigmatizing and perhaps incriminating issues such as abortion, drug addiction, HIV/AIDS infection status, duration of suffering from a disease, sexual behaviour, domestic violence, racial prejudice or non-compliance with laws and regulations. In these situations, collecting data by means of survey modes based on direct questioning (DQ) methods of interview is likely to encounter serious problems, participants in the survey may deliberately release untruthful or misleading answers, or participants may refuse to respond due to the social stigma. In particular, social desirability bias, i.e. the desire to make a favourable impression on others, poses a significant threat to the validity of self-reports in "sensitive research" as well described in Dickson-Swift et al. (2008). This type of bias generally produces an 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, sexual violence, drug and enhancing substances, tax evasion).

Refusal to answer and false answers constitute non-sampling errors that are difficult to deal with and can seriously flaw the quality of the collected data, thus jeopardizing the usefulness of subsequent analyses including statistical inference of unknown characteristics of the population under study. Although these errors cannot be totally avoided, they may be mitigated by enhancing respondents' cooperation. Since the decision to cooperate fully and honestly greatly depends on how survey participants perceive their privacy being disclosed, survey modes which ensure respondents' anonymity or, at least, a high degree of confidentiality, may go some way to improving cooperation and, consequently, ensure more reliable information on sensitive topics than that derived from DQ.

In this respect, survey statisticians and practitioners have developed many different strategies to ensure interviewees' anonymity and to reduce the incidence of evasive answers and underreporting of social taboos when direct questions are posed on sensitive issues. One possibility is to limit the influence of the interviewer, by providing self-administered questionnaires (SAQs) 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.

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 non-randomized response technique (NRRT). All the approaches have produced a considerable literature and attracted the interest of health, cognitive and behavioural psychologists, epidemiologists, health-care operators, researchers engaged in organizing, managing and conduction sensitive studies, as well as policy-makers committed in formulating effective diseases and mental disorders control measures and promoting public intervention programs to gauge progress toward improving the behavioural health of a state.

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. In its original version, this non-standard survey approach adopts a randomization device such as a deck of cards, dice, coins, coloured numbered balls, spinners or even a computer to conceal the true answer, in the sense that respondents reply to one of two or more selected questions depending on the result of the device.

The Warner procedure is as follows, to estimate for a community the proportion of people

bearing a stigmatizing characteristic (denoted by the symbol A) a sampled person is offered a box of a considerable number of identical cards with a proportion $p(0 < p < 1, p \neq 0.5)$ of them marked A and the rest marked A^c. The person on request is required to draw a "random" card and respond by answering "yes" for a "match" between the card type and the person's own real characteristic or a "no" for a "non-match" before returning the card to the box. The randomization is performed by the interviewee, and the interviewer is not permitted to observe the outcome of the randomization. The interviewee responds to the question selected by the randomization device, and the interviewer knows only the response provided. The respondent's privacy or anonymity is fully protected because no one but the respondent knows which question was answered. But it is possible statistically to derive a plausible estimate, on analyzing the bunch of randomized responses thus collectively gathered, for the required proportion bearing A. It is hoped that the privacy of the person responding is securely protected.
The randomization device generates a probabilistic relation between the sensitive question and a given answer which is used to draw inference about unknown parameters of interest. The rationale of the RRT is that interviewees 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 both to the interviewer and to the researcher and, consequently, respondents' privacy is preserved.

Contextually, many studies have assessed the validity of RR methods, showing that they can produce more reliable answers than conventional data collection methods (e.g., DQ in face-to-face interviews, self-administered questionnaires with paper and pencil and computer-assisted self interviews). In this respect, see van der Heijden et al. (2000), Lara et al. (2004) and Lensvelt-Mulders et al. (2005), to name just a few.

The RRT has been applied in surveys covering a variety of sensitive topics including, for instance, racism (Ostapczuk et al., 2009; Krumpal, 2012), drug use (Kerkvliet, 1994; Striegel et al., 2010; Stubbe et al., 2013; Petróczi et al., 2011; Dietz et al., 2013; Franke et al., 2013; James et al., 2013; Nakhaee et al., 2013; Shamsipour et al., 2014; Khosravi et al., 2015; Goodstadt and Gruson, 1975; Simon et al., 2006), smoking behaviour validation studies (Fox et al., 2013), abortion (Lara et al., 2004, 2006; Perri et al., 2016; Oliveras and Letamo, 2010; Moseson et al., 2015), sexual victimization (Krebs et al., 2011), HIV/AIDS infection and high-risk sexual behaviours (Arnab and Singh, 2010; Geng et al., 2016; Arentoft et al., 2016; Jong et al., 2012; Starosta and Earleywine, 2014; Kazemzadeh et al., 2016), animal diseases (Cross et al., 2010; Gunarathne et al., 2016; Randrianantoandro et al., 2015), illegal fishing (Blank and Gavin, 2009; Arias and Sutton, 2013) and hunting (Nuno et al., 2013; Conteh et al., 2015), the illegal use of natural resources (Chaloupka, 1985; Schill and Kline, 1995; Solomon et al., 2007; Blank and Gavin, 2009; Harrison et al., 2015), academic cheating and plagiarism (Fox and Meijer, 2008;

Jann et al., 2012), fraud in the area of disability benefits (van der Heijden et al., 2000; Lensvelt-Mulders et al., 2006), tax evasion (Houston and Tran, 2001; Korndörfer et al., 2014), voting turnout (Holbrook and Krosnick, 2010).

Warner's study generated a rapidly-expanding body of research literature on alternative techniques for eliciting suitable RR schemes in order to estimate a population proportion (see, e.g., Arnab, 1996; Barabesi and Marcheselli, 2006; Barabesi, 2008; Gjestvang and Singh, 2006; Lee et al., 2013; Liu and Tian, 2013; Perri, 2008).

Standard RR methods have been basically conceived to be used in surveys which require a binary response ("yes" or "no") to a sensitive question, and seek to estimate the proportion of people presenting a given sensitive attribute. Nevertheless, empirical studies may address situations in which the response to a sensitive question results in a quantitative variable and the interest of the researcher relies, in the easiest case, on the estimation of the mean or the total of the sensitive variable under study. To deal with such situations, Warner's idea has been promptly extended to sensitive quantitative variables by Greenberg et al. (1971); Eriksson (1973) and Pollock and Bek (1976). Since then a lot of mechanisms have been proposed in the literature to scramble the response and, thus, protect respondents' privacy (see, e.g., Eichhorn and Hayre, 1983; Bar-Lev et al., 2004; Saha, 2007a; Diana and Perri, 2010; Gupta et al., 2010; Odumade and Singh, 2010; Arcos et al., 2015, and the contributions collected in Chaudhuri et al., 2016). When dealing with quantitative sensitive variables, the idea is to ask respondents to not disclose the true value of the sensitive variable but rather to release a scrambled value obtained by algebraically perturb the true response making use of one or more scrambling random variables, independent each other and of the sensitive one, which distributions are completely known to the researcher.

Usually, RR methods, both for qualitative and quantitative variables, have been theoretically developed assuming that the observed responses are collected on sampled units selected according to simple random sampling. Indeed, most of the real studies are based on complex surveys involving, for instance, stratification, clustering and unequal probability sampling designs. Therefore, the RRT has been extended to more complex sampling design, as stratified sampling (Mahajan and Singh, 2005; Kim and Elam, 2007; Saha, 2007a; Singh and Tarray, 2015), or unequal probability sampling (Chaudhuri, 2001, 2004; Arnab and Dorffer, 2006; Saha, 2007b; Pal, 2008; Quatember, 2012).

For a comprehensive review of the topic, interested readers are referred to Fox and Tracy (1986); Chaudhuri and Mukerjee (1988); Chaudhuri (2011); Chaudhuri and Christofides (2013); Chaudhuri et al. (2016); Tian and Tang (2014). Useful and detailed studies on recent methodological advances, more complex estimation problems and new challenges may be found, among others, in Arcos et al. (2015); Barabesi et al. (2013, 2015); Diana and Perri (2011); Fox (2016);

Fox et al. (2014); Glynn (2013); Groenitz (2014); Hoffmann and Musch (2016); Hoffmann et al. (2016); Hussain et al. (2017); Ibrahim (2016); Imai (2011); Imai et al. (2015); Liu and Tian (2013); Moshagen et al. (2014); Nepusz et al. (2014); Perri and van der Heijden (2012); Petróczi et al. (2011); Tsuchiya (2005); Ulrich et al. (2012); Wu and Tang (2016).

Despite the good reputation that the RRT has acquired over time as a tool to obtain reliable data while protecting respondents' confidentiality, avoiding unacceptable rates of non-response and reducing social desirability response bias, the approach, at least in its basic idea, suffers from some inadequacies that have limited its complete acceptance among survey statisticians and practitioners. The main limitations may be summarized in the following points: (i) RRT surveys are, in general, more time-demanding and costly than other types of survey modes; (ii) RRT estimates are subject to greater sampling variance (i.e., lower efficiency) than DQ estimates. This loss of efficiency represents the cost of obtaining more reliable information by reducing response bias. Consequently, achieving estimates which are comparably efficient with those obtained under DQ may require a considerably larger sample with the consequent increase in cost, an aspect which is rarely acceptable; (iii) RRT surveys lack reproducibility, in the sense that the same respondent may give different information if asked to repeat the survey. This is because his/her answer depends on the outcome of the randomization device. Hence, conditioned to a selected sample of respondents, the estimation process may yield different estimates according to the outcome of the device; (iv) lack of understanding and trust among respondents. Chaudhuri and Christofides (2007) observed that the RRT basically asks respondents to provide information that may seem useless or even deceitful. When the respondent does not understand the mathematical logic underlying the technique, then the entire procedure may be suspect, leading the respondent to believe there might be a way for the interviewer to determine his/her exact status regarding the sensitive characteristic by processing the response provided. Moreover, respondents may not understand the instructions for using the RR device and/or not trust the privacy protection offered. Hence, they might intentionally refuse to participate in the survey or break the rules of the RR design; (v) RR procedures require a randomization device to drive the answer. Using physical devices limits the application of the RRT exclusively to face-to-face personal interviews and may also be more time consuming (the procedure must be explained to each survey participant) and costly (the devices must be obtained) than DQ. Other means of survey administration, such as telephone interview, self-administered mail questionnaire and internet-delivered interviews, seem to be precluded. In addition, respondents could find it difficult to use a physical device, for instance due to reduced motor capacity, or be suspicious of using something provided by the interviewer.

Mindful of these drawbacks, alternative IQTs have been proposed which overcome some of the limitations affecting the RRT and enable sensitive information to be acquired while preserving respondents' confidentiality. Such alternative methods are encompassed in different approaches which include the nominative technique (Miller, 1985), the three card method (Droitcour and Larson, 2002), the non-randomized response technique (Tian and Tang, 2014) and the item count technique, also known as the unmatched count technique (UCT), the block total response or the list experiment, (Raghavarao and Federer, 1979; Miller, 1984; Droitcour et al., 1991). All of these alternatives were originally conceived for surveys requiring a "yes" or "no" response to a sensitive question, or a choice of responses from a set of nominal categories, and do not address quantitative sensitive characteristics.

In ICT, respondents are asked directly about their own sensitive behaviour and, at the same time, about a number of innocuous behaviours. In the standard setting, the method requires the selection of two samples: a reference sample which receives a short list (SL) of items on questions only about innocuous behaviours, and a treatment sample which receives a long list (LL) containing the innocuous items in the SL-sample and a sensitive question. Units selected in the two samples are asked to report the total number of items that apply to them without revealing which item applies individually. The ICT is used in surveys which require the study of a qualitative variable. Nonetheless, many practical situations may deal with sensitive variables which are quantitative in nature. To address this situation, Chaudhuri and Christofides (2013) proposed a generalization of the ICT that can be used to estimate the mean or the total of a quantitative variable. Trappmann et al. (2014) called this variant the item sum technique (IST). The IST works in a similar way to the ICT. Two independent simple random samples are drawn from the population. Units belonging to one of the two samples are presented with the LL of items containing the sensitive question and a number of non-sensitive questions; units in the other sample receive only the SL of items consisting of the non-sensitive questions. All of the items refer to quantitative variables, possibly measured on the same scale as that of the sensitive variable. Respondents are then asked to report the total score of their answers to all of the questions in their list, without revealing the individual score for each question. Like the ICT, the mean difference of the answers between the LL-sample and the SL-sample is then used as an unbiased estimator of the population mean of the sensitive variable.

To the best of our knowledge, there is only the paper by Trappmann et al. (2014) who used IST in a CATI survey on undeclared work in Germany and outlined a procedure to estimate regression models for the IST, and Hussain et al. (2017) proposed a one-sample variant of the IST, in which each of the units in the simple random sample is provided with a list of items and just one of these items contains queries about stigmatizing and non-stigmatizing variables. These authors also considered ratio, product and regression estimators to incorporate auxiliary information into the IST estimation procedure. The one-sample approach to the IST has also been considered by Shaw (2015).

One of the main disadvantages of RRTs is the use of a randomization device. In the NRRT, no physical device is adopted, and neither are respondents asked to conduct a randomizing procedure (Tian and Tang, 2014). However, that does not mean that no randomization takes place. In some cases, an implicit randomization is performed. For example in Christofides (2009) a randomization is done based on the sensitive characteristic and the participant provides a “yes” or “no” answer if he/she has or does not have a non-stigmatizing characteristic. In other cases, such as the technique of Tian and J.W. Yu (2007) a respondent provides a response about his/her status as related to the sensitive characteristic and a non-sensitive one, in such a way that the response provided is not enough to infer whether he/she belongs to the sensitive group.

The aim of this thesis is to develop methodological advances and software for indirect questioning techniques, specifically, in randomized response technique and item sum technique.
In Appendix 1, we obtain estimators in the presence of auxiliary information for a general class of estimators for the total in RRT. In Appendix 2, we use a RRT for investigating cannabis use by Spanish university students and we compare these results with those obtained by DQ. We develop RRTCS package to compute the point and interval estimation of linear parameters using data obtained from RR surveys under complex sampling designs in Appendix 3. In Appendix 4 we get the estimation for IST under a generic sampling design and in the presence of auxiliary information. We also extend this calibration approach to the estimation for domain. In Appendix 5, we study under IST the problem of the muliple sensitive estimation and the optimal allocation of the total sample size into the LL-sample and the SL-sample under a generic sampling design. We also extend the allocation in multiple estimation. We use a RRT and IST in a real study conducted in Spain to investigate the frecuency of drug addiction and sexual behaviour among university students, and both are comparated with the corresponding estimates obtained by DQ method in Appendix 6. In Appendix 7, we use ICT to reveal hidden support for female genital cutting in Shouth Central Ethiopia.

Chapter 2

Objectives

2.1 An improved class of estimators in RR surveys

Most research into RRT deals exclusively with the interest variable and does not make explicit use of auxiliary variables in the construction of estimators. Diana and Perri (2010) pointed out that in sampling practice, direct techniques for collecting information about non-sensitive characteristics make massive use of auxiliary variables to improve sampling design and to achieve higher precision in population parameter estimates. Nevertheless, very few procedures have been suggested to improve randomization technique performance using the supplementary information. From a mathematical point of view, a process of seeking an optimal estimator in a class of estimators for the total of sensitive characteristic arises, under a general model for the scrambling response and in presence of additional information.
We propose a general class of estimators for the population total. Proposed estimators are based upon auxiliary variables and assume that observations on the variable of interest are obtained using a RRT. We present particular estimators of the proposed class of estimators, and we derive the asymptotic properties of these estimators. Also, we study some asymptotic properties under simple random sampling and stratified sampling.

2.2 Application of randomized response techniques for investigating cannabis use by Spanish university students

Cannabis (or marijuana) is the illicit drug that is most commonly used by young adults in Spain. 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 disturbances and numerous other conditions (Lamarine, 2012). Health care and social problems related to the use of cannabis have led researchers to investigate screening procedures aimed at detecting persons at risk. These screening instruments are capable of detecting (probable) cannabis dependence or problematic use and have been used in Spain in surveys for the National Plan on Drugs in schools and among the general population (Cuenca-Royo et al., 2012). The application of short screening scales to assess dependence and other problems related to the use of cannabis presents a time and cost-saving means of estimating the overall prevalence of cannabis use and of related negative consequences (Bastiani et al., 2013; Gyepesi et al., 2014; Hides et al., 2007; Legleye et al., 2013). Nevertheless, there is a need to formally evaluate the validity of the data gathered (Piontek et al., 2008). Studies by Harrison (1997) and Ramo et al. (2012) have evaluated the reliability and validity of anonymous studies of cannabis use, but these reports do not examine the other side of validity, namely the fact that respondents may lie, when faced with a question that they find embarrassing, or refuse to answer, or choose a response that prevents them from having to continue and, clearly, this situation may arise in questionnaires related to the use of illegal drugs. Other potential threats to survey accuracy are non-response and reporting error (Tourangeau and Yan, 2007).

The aim of this study is using indirect questioning techniques, specifically a RRT, to investigate in a university population the mean number of cannabis cigarettes consumed in the last year and the mean number of days that the students had consumed cannabis on the previous 90 days.

Surveys based on the RRTs are widely used when the questions are sensitive, and especially when the variable of interest is a qualitative one. RRTs also exist for quantitative variables, but these are not used as commonly, so it should be noted that in our study, we took into account quantitative variables in order to make the scope of the study as complete as possible.

2.3 RRTCS: An R package for randomized response techniques in complex surveys

Usually, RR methods are developed assuming that sample is obtained using simple random sampling. Most of the surveys in practice are complex surveys involving stratification, clustering and unequal probability of selection of sample. Data from complex survey designs require special consideration with regard to estimation for finite population parameters and corresponding variance estimation procedures, as a consequence of significant departures from simple random sampling assumption. In such a complex survey design, unbiased variance estimation is not easy to calculate because of clustering and involvement of second-order inclusion probabilities which are generally complex. Several software packages have been developed to facilitate the analysis of complex survey data and implement some of these estimators as SAS, SPSS, Systa, Stata, SUDAAN or PCCarp. CRAN contains several R packages that include these design-based methods typically used in survey methodology to treat samples obtained from a sampling frame, e.g., survey, sampling, laeken or TeachingSampling among others (see Templ 2015, for a detailed list of packages that include methods to analyze complex surveys).

Standard software packages for complex surveys cannot be used directly when the sample is obtained from RRTs. The analyses with standard statistical software, with certain modifications in the randomized variables, can yield correct point estimates of population parameters but still yield incorrect results for estimated standard errors. Recently some authors have developed R-packages for estimation with RR surveys, RRreg: Correlation and Regression Analyses for Randomized Response Data, (Heck and Moshagen, 2015) and rr: Statistical Methods for the Randomized Response Technique, (Blair et al., 2015).

The methods implemented in these packages are used under the assumption on simple random sampling and do not explore various theoretical and practical issues that may arise when adopting different survey sampling methods. In order to fill this gap, we have developed a R package named RRTCS. This package provides functions for point and interval estimation from RR surveys under complex sampling designs.

2.4 Advances in estimation by the item sum technique using auxiliary information in complex surveys

This study has two objectives, the first of these aims is motivated by the fact that real surveys are customarily conducted by using complex sampling designs such as stratified and/or cluster sampling, with units selected according to a specific varying probability scheme, so we provide a general framework for the IST by extending the results of Chaudhuri and Christofides (2013) and Trappmann et al. (2014) from simple random sampling to a generic complex sampling design.

And the second concerns the fact that, in sampling practice, DQ techniques for collecting information about non-sensitive characteristics make use of auxiliary variables to improve sampling designs and to achieve higher precision in the estimates of unknown population parameters, so we investigate the effectiveness of employing auxiliary information to improve, without incurring additional costs or increasing the sample size, the efficiency of estimates when the IST is used to obtain data from a complex survey and then extend this calibration approach to the estimation for domains.

In addition, we discuss variance estimation and the impact on the estimates of including an increased number of innocuous questions in the list of items.

2.5 Multiple sensitive estimation and optimal sample size allocation in the item sum technique

This study has two objectives, the first pertains the reduction of the statistical burden when multiple sensitive items are investigated and estimates of certain characteristics are required. This situation occurs frequently in real studies where researchers must incorporate $Q \geq 2$ sensitive questions in their surveys. Three different approaches are considered. The first consists of performing Q separate IST surveys, one for each sensitive item. This approach requires for each item the selection of one LL-sample and one SL-sample, for a total of $2Q$ samples. In practice, however, this solution does not appear to be feasible, because it is both time-consuming and costly, and also because possible associations between variables would be lost since each IST survey is independently executed on different subjects. To overcome these problems, a single IST survey could be performed. In this case, just one LL-sample and one SL-sample are selected and respondents are asked to participate in Q separate IST experiments, one for each sensitive item. As can be readily imagined, this procedure imposes a heavy statistical burden on the respondents, since they must provide the required information on the single sensitive items by separately implementing the IST Q times. A third viable alternative, which requires the selection of $Q + 1$ samples and acts as a trade-off between the first two approaches, is therefore proposed and its performance investigated.
The second problem we consider is how to split the total sample size into the LL-sample and the SL-sample. A simple solution would be to allocate the same number of units to each sample, irrespective of the variability of the items in the two lists. Although intuitive and easy to implement, this basic solution is inefficient because estimates may be affected by high variability. A possible alternative would be to achieve optimal sample size allocation by minimizing the variance of the IST estimates under a budget constraint. This possibility is first formalized and discussed under a generic sample design and, then, results are particularized to the simple random sampling and the stratified sampling designs. Optimal allocation results are finally extended to the multiple sensitive estimation setting.

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

In this study, 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 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 sensitive behaviours, and both compared with the DQ method.

2.7 Indirect questioning methods reveal hidden support for female genital cutting in South Central Ethiopia

Female genital cutting (FGC) has major implications for women's physical, sexual and psychological health, and eliminating the practice is a key target for public health policy-makers. To date one of the main barriers to achieving this has been an inability to infer privately-held views on FGC within communities where it is prevalent. As a sensitive, and often illegal, topic, people are anticipated to hide their true support for the practice when questioned directly.
To date most studies exploring FGC behaviour have relied on self-report data derived from DQ methods, with many indicating that rates of (and interest in) FGC are broadly in decline (Koski and Heymann, 2017). FGC status obtained through physical examination rarely exists to substantiate these claims, and where it does, has revealed discordance between the two measures (Elmusharaf et al., 2006; Klouman and R. Manongi, 2005). This disparity, between clinical and self-report data, confirms that people may be inclined to conceal FGC behaviour, and their support for it, in surveys. Yet, physical examination is intrusive and expensive, since it requires a health professional, and thus is infeasible as a tool to guide research and policy.

Here we use an IQT, specifically UCT, to identify hidden support for FGC in a rural South Central Ethiopian community where the practice is common, but thought to be in decline. Employing a socio-demographic household survey of Arsi Oromo adults, which incorporated both direct and indirect response techniques, we compare directly-stated versus privately-held views in support of FGC, and individual variation in responses by age, gender and education and target female, concretely daughters versus daughters-in-law.

Chapter 3

Methodology

3.1 Randomized Response Techniques

In literature there are a lot of RRTs, here we have included only some of them. Concretely some of the simplest to carry out in practice in real studies, including those that we used in our studies. Among the RRTs we find a large number of both qualitative and quantitative models. Among the qualitative models, it is worth mentioning the Warner, Horvitz and Forced response models. Among the quantitative, Eichhorn and Hayre, Bar-Lev, Bobovitch and Boukai and Eriksson models.

Consider a finite population $U = \{1, \ldots, i, \ldots, N\}$, consisting of N different elements. Let y_i be the value of the sensitive aspect under study for the ith population element. Our aim is to estimate the finite population total $Y = \sum_{i=1}^{N} y_i$ of the variable of interest y or the population mean $\bar{Y} = \frac{1}{N}\sum_{i=1}^{N} y_i$. In case the study variable y is qualitative, we define $y_i = 1$ if the ith unit of the population possesses certain attribute A (say) and $y_i = 0$ if the unit does not possess the attribute A. In this case the population mean $\bar{Y}$ is equal to the population proportion π.
Assume that a sample s of individuals is chosen according to a general 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 $i, j \in U$.

3.1.1 Description of some RRTs

Qualitative models

Warner model

In Warner RRT (Warner, 1965), a sampled person labelled i is offered a box of a considerable number of identical cards with a proportion $p, (0 < p < 1, p \neq 0.5)$ of them marked A and the rest marked A^c. The person is requested, randomly, to draw one of them, to observe the mark on the card, and to give the response

$$z_i = \begin{cases} 1 & \text{if card type "matches" the trait } A \text{ or } A^c \\ 0 & \text{if a "no match" results} \end{cases}$$

Horvitz model

Horvitz et al. (1967) and Greenberg et al. (1969) modified Warner's method by incorporating a sensitive question (character y) along with a non-sensitive (unrelated) question (character x). The RR device presents to the sampled person labelled i a box containing a large number of identical cards, with a proportion $p, (0 < p < 1)$ bearing the mark A and the rest marked B, an innocuous attribute whose population proportion α is known. The response solicited denoted by z_i takes the value y_i if i bears A and the card drawn is marked A or if i bears B and the card drawn is marked B. Otherwise z_i takes the value 0.

Forced Response model

The model proposed by Boruch (1972) is an alternative to the Warner scheme and provides greater protection for the interviewee. In the Forced Response scheme, the sampled person i is offered a box with cards: some are marked "Yes" with a proportion p_1, some are marked "No" with a proportion p_2 and the rest are marked "Sensitive", in the remaining proportion $p_3 = 1 - p_1 - p_2$, where $0 < p_1, p_2 < 1, p_1 \neq p_2, p_1 + p_2 < 1$. The person is requested to randomly draw one of them, to observe the mark on the card, and to respond

$$z_i = \begin{cases} 1 & \text{if the card is type "Yes"} \\ 0 & \text{if the card is type "No"} \\ y_i & \text{if the card is type "Sensitive"} \end{cases}$$

Quantitative models

Eichhorn and Hayre model

In Eichhorn and Hayre model (1983) each sampled respondent is to report $z_i = y_i S$ where S is a scramble variable whose distribution is assumed to be known.

Bar-Lev, Bobovitch and Boukai model

In the model considered in Bar-Lev et al. (2004) the RR given by the person i is

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

where S is a scramble variable, whose mean μ and standard deviation σ are known.

Eriksson model

In Eriksson model (1973) the RR given by the person labelled i is

$$z_i = \begin{cases} y_i & \text{with probability } p \\ S & \text{with probability } q_1, q_2, ..., q_j \text{ verifying } q_j > 0, \sum_j q_j = 1-c \end{cases}$$

where S is a discrete uniform variable.

3.1.2 Estimation and Variance

For the sake of notation, let $d_i = \pi_i^{-1}$, $d_{ij} = \pi_{ij}^{-1}$, $\Delta_{ij} = \pi_{ij} - \pi_i \pi_j$. Under a DQ survey mode, let $\hat{\bar{Y}}_{HT}$ denote the well-known Horvitz-Thompson estimator (hereafter HT-estimator; Horvitz and Thompson, 1952) of $\bar{Y}$

$$\hat{\bar{Y}}_{HT} = \frac{1}{N} \sum_{i \in s} d_i y_i$$

The estimator is unbiased and has variance

$$V(\hat{\bar{Y}}_{HT}) = \frac{1}{N^2} \sum_{i \in U} \sum_{j \in U} \Delta_{ij} d_i y_i d_j y_j$$

which can be unbiasedly estimated by

$$\hat{V}(\hat{\bar{Y}}_{HT}) = \frac{1}{N^2} \sum_{i \in s} \sum_{j \in s} d_{ij} \Delta_{ij} d_i y_i d_j y_j$$

In this case, y is a sensitive variable that cannot be observed directly. In order to consider a wide variety of RR procedures, we consider the unified approach given by Arnab (2004). The interviews of individuals in the sample s are conducted in accordance with a RR model. Because y_i is not directly available from the respondent, y_i is estimated through the RR obtained from the ith respondent. Suppose that the ith respondent has to conduct a RR trial independently and z_i is the RR (or scrambled response) for the trial. For each $i \in s$, the RR induces a revised RR r_i such as $E_R(r_i) = y_i$ and $V_R(r_i) = \phi_i$ where the operators E_R and V_R denote expectation and variance with respect to randomization procedure RR. As usual in the design-based approach to RRTs, it is assumed that the sampling design and the randomization stage are independent of each other (e.g., Barabesi et al. 2013) and that the randomization stage is performed on each selected individual independently. In this general setup, the HT-type estimator for the population total of the sensitive characteristic y is given by

$$\hat{Y}_{HT}(r) = \sum_{i \in s} d_i r_i$$

The variance of $\hat{Y}_{HT}(r)$ is given by

$$V(\hat{Y}_{HT}(r)) = \sum_{i \in U} \sum_{j \in U} \Delta_{ij} d_i y_i d_j y_j + \sum_{i \in U} d_i V_R(r_i) = V(\hat{Y}_{HT}) + \sum_{i \in U} d_i V_R(r_i)$$

and an unbiased estimator of $V(\hat{Y}_{HT}(r))$ is

$$\hat{V}(\hat{Y}_{HT}(r)) = \sum_{i \in s} \sum_{j \in s} d_{ij} \Delta_{ij} d_i r_i d_j r_j + \sum_{i \in s} d_i \hat{V}_R(r_i) = \hat{V}(\hat{Y}_{HT}) + \sum_{i \in s} d_i \hat{V}_R(r_i)$$

This estimator is an unbiased estimator of $V(\hat{Y}_{HT}(r))$ if $\hat{V}(\hat{Y}_{HT})$ is an RR-unbiased for $V(\hat{Y}_{HT})$. Similarly, an unbiased estimator for the population mean $\bar{Y}$ for the RR survey is given by

$$\hat{\bar{Y}}_{HT}(r) = \frac{1}{N} \sum_{i \in s} d_i r_i$$

and an unbiased estimator for its variance is calculated as:

$$\hat{V}(\hat{\bar{Y}}_{HT}(r)) = \frac{1}{N^2} \left(\sum_{i \in s} \sum_{j \in s} d_{ij} \Delta_{ij} d_i r_i d_j r_j + \sum_{i \in s} d_i \hat{V}_R(r_i) \right) = \frac{1}{N^2} \left(\hat{V}(\hat{\bar{Y}}_{HT}) + \sum_{i \in s} d_i \hat{V}_R(r_i) \right)$$

In qualitative models, the values r_i and $V_R(r_i)$ are obtained for each model. Then, the models explained previously are considered:

Warner model

The transformed variable is
$$r_i = \frac{z_i - (1-p)}{(2p-1)}$$
The variance of r_i is
$$V_R(r_i) = \frac{p(1-p)}{(2p-1)^2}$$
Now noting $V_R(r_i) = E_R(r_i^2) - (E_R(r_i))^2 = E_R(r_i^2) - y_i^2 = E_R(r_i^2) - y_i = E_R(r_i^2) - E_R(r_i) = E_R(r_i(r_i-1))$, we set an unbiased estimator of $V_R(r_i)$ as

$$\hat{V}_R(r_i) = r_i(r_i-1)$$

Horvitz model

The transformed variable is $r_i = \frac{z_i-(1-p)\alpha}{p}$ and the estimated variance is $\hat{V}_R(r_i) = r_i(r_i-1)$.

Forced Response model

The transformed variable is $r_i = \frac{z_i-p_1}{1-p_1-p_2}$ and the estimated variance is $\hat{V}_R(r_i) = r_i(r_i-1)$.

In some quantitative models, the values r_i and $V_R(r_i)$ are calculated in a general form (Arcos et al., 2015) as follows:
The randomized response given by the person i is

$$z_i = \begin{cases} y_i & \text{with probability } p_1 \\ y_iS_1 + S_2 & \text{with probability } p_2 \\ S_3 & \text{with probability } p_3 \end{cases}$$

with $p_1 + p_2 + p_3 = 1$ and where S_1, S_2 and S_3 are scramble variables whose distributions are assumed to be known. We denote by μ_i and σ_i respectively the mean and standard deviation of the variable $S_i, (i = 1, 2, 3)$.
The transformed variable is
$$r_i = \frac{z_i - p_2\mu_2 - p_3\mu_3}{p_1 + p_2\mu_1}$$
its variance is
$$V_R(r_i) = \frac{1}{(p_1 + p_2\mu_1)^2}(y_i^2A + y_iB + C)$$

where

$$A = p_1(1-p_1) + \sigma_1^2 p_2 + \mu_1^2 p_2 - \mu_1^2 p_2^2 - 2p_1p_2\mu_1$$

$$B = 2p_2\mu_1\mu_2 - 2\mu_1\mu_2p_2^2 - 2p_1p_2\mu_2 - 2\mu_3p_1p_3 - 2\mu_1\mu_3p_2p_3$$

$$C = (\sigma_2^2 + \mu_2^2)p_2 + (\sigma_3^2 + \mu_3^2)p_3 - (\mu_2p_2 + \mu_3p_3)^2$$

and the estimated variance is

$$\hat{V}_R(r_i) = \frac{1}{(p_1 + p_2\mu_1)^2}(r_i^2A + r_iB + C)$$

For example, if $p_1 = p_3 = 0, p_2 = 1, S_2 = 0$ then the proposed model becomes the Eichhorn and Hayre model (1983); if $p_1 = p, p_2 = 1-p, p_3 = 0, S_2 = 0$, then the proposed model becomes the BarLev model (Bar-Lev et al., 2004) or if $p_1 = p, p_2 = 0$ and S_3 is a discrete uniform variable with probabilities $q_1, q_2, ..., q_j$ verifying $q_1 + q_2 + ... + q_j = 1-p$, then the proposed technique becomes Eriksson method (1973).

To calculate the estimator variance we need to ascertain the second-order inclusion probabilities of each pair of units of sample s. In some complex sampling designs, this is a complex matter. A simpler alternative is to use resampling techniques.

3.2 An improved class of estimators in RR surveys

One considers k auxiliary variables $x_1, \ldots, x_k$, for which the population totals $X_1, \ldots, X_k$, are known. We assume that the values of auxiliary variables can be observed directly in the sample. Our goal is to estimate the population parameter Y by using observations of the variables $r, x_1, \ldots, x_k$ in the sample s, and the known population values $X_1, \ldots, X_k$ associated with the auxiliary variables. We note by $\widehat{X}_h$ the HT estimator of the total X_h $(h = 1, ..., k)$.
Motivated by Srivastava and Jhajj (1981), we suggest the class of estimators of Y

$$\widehat{Y}_g^{(r)} = \{G(\hat{Y}(r), u_1, ..., u_k)\},$$

where $G(\cdot)$ is a function of $u_h = \widehat{X}_h/X_h$, continuous in a closed convex sub-space, $P \subset \mathbb{R}^{k+1}$, containing the point $(Y, 1, ..., 1) = (Y, \mathbf{1})$, and such that

- $G(Y, \mathbf{1}) = Y$

- $G_0'(Y, \mathbf{1}) = 1$ where $G_0'(Y, \mathbf{1})$ denoting the first partial derivative of $G(\cdot)$ with respect to $\hat{Y}(r)$.

- The first and second-order partial derivatives of $G(\cdot)$ exist and are also continuous and bounded in P.

Theoretically, we study some asymptotic properties of $\widehat{Y}_g^{(r)}$, also under simple random and stratified sampling are derived. Empirically, we evaluate their behaviour and they are compared with alternative estimators through simulation studies.

3.3 Application of randomized response techniques for investigating cannabis use by Spanish university students

To investigate cannabis use in the Spanish universities, we conduct a survey of university students. The target population for this survey includes students at the University of Granada and the University of Murcia. Subjects are selected using probabilistic sampling stratified by university. Respondents are randomly selected to use the RRT (sub-sample 1) and to be asked directly about illicit drug use (sub-sample 2). All students are invited to participate in a study and provided informed signed consent.

The questionnaire is the same in two sub-samples, and the sensitive questions are:

- How many cannabis cigarettes did you consume last year?
- Over the past 90 days, how many days did you consume cannabis?

In sub-sample 1, the responses are randomized using a generalization of the model proposed by Bar-Lev et al. (2004) for simple random sampling and later extended by Arcos et al. (2015) to use with complex samples.

The randomizing device used is the app "Baraja Española", that it is a deck composed of forty cards, divided into four families or suits, each numbered one to seven and three figures. If the obtained card is a face card, the sensitive question should be answered; otherwise, the real number should be given, multiplied by the number shown on the card. This randomizing device must be appropriate to protect the confidentiality of the respondent.

Inference is used in survey sampling to estimate the parameters of interest. The HT estimator (Singh, 2003) is used to estimate the mean values for the DQ. We use the unified method of estimating population surveys characteristic in RR proposed by Arnab (1994) and Arcos et al. (2015).

We use the package RRTCS (Cobo et al., 2015), which is only one that incorporates estimation procedures for handling RR data obtained from complex surveys. Specifically, the BarLev function, that implements the BarLev model.

3.4 RRTCS: An R package for randomized response techniques in complex surveys

RRTCS (Cobo et al., 2015). Randomized Response Techniques for Complex Surveys.
RRTCS is a new R package to perform point and interval estimation of linear parameters using data obtained from RR surveys under complex sampling designs. The package works with a wide range of sampling designs, including simple random sampling with and without replacement (SRSWR and SRSWOR), stratified sampling, cluster sampling, unequal probabilities sampling and any combination of these.

The package consists of 21 main functions, each of which implements one of the following RR procedures for complex surveys:

- Randomized response procedures to estimate parameters of a qualitative sensitive characteristic: Christofides model (Christofides, 2003), Devore model (Devore, 1977), Forced response model (Boruch, 1972), Horvitz model (Horvitz et al., 1967; Greenberg et al., 1969), Horvitz model with unknown B (Chaudhuri, 2011, page 42), Kuk model (Kuk, 1990), Mangat model (Mangat, 1992), Mangat model with unknown B (Chaudhuri, 2011, page 53), Mangat and Singh model (Mangat and Singh, 1990), Mangat, Singh and Singh model (Mangat et al., 1992), Mangat, Singh and Singh model with unknown B (Chaudhuri, 2011, page 54), Singh and Joarder model (Singh and Joarder, 1997), Soberanis model (Soberanis-Cruz et al., 2008) and Warner model (Warner, 1965).

- Randomized response procedures to estimate parameters of a quantitative sensitive characteristic: Eriksson model (Eriksson, 1973), BarLev model (Bar-Lev et al., 2004), Chaudhuri and Christofides model (Chaudhuri and Christofides, 2013, page 97), Diana and Perri 1 model (Diana and Perri, 2010, page 1877), Diana and Perri 2 model (Diana and Perri, 2010, page 1879), Eichhorn and Hayre model (Eichhorn and Hayre, 1983), and Saha model (Saha, 2007a).

The package also includes an additional function, called ResamplingVariance, which provides estimates variance of the RR estimators using some resampling methods (Wolter, 2007) under stratified, cluster, and unequal probabilities sampling. This includes the jackknife method (Quenouille, 1949), the Escobar-Berger method (Escobar and Berger, 2013), and the Campbell-Berger-Skinner method (Berger and Skinner, 2005).
Finally, the package includes 20 data sets with observations from different surveys conducted in real and simulated populations using different RRTs.

3.5 Item Count Technique

3.5.1 Estimation and Variance

Assume that the researcher wishes to use the ICT to determine the prevalence of a sensitive attribute A in a population. The ICT was originally conceived by Raghavarao and Federer (1979) and Miller (1984) and consists of drawing two independent samples from the target population. Without loss of generality, units belonging to long list (LL) sample, s_{ll}, are provided with a LL of items containing $(G+1)$ dichotomous questions, of which G are non-sensitive, while the remaining one refers to the sensitive attribute A. The sampled units are instructed to consider the LL, and to count and report the number of items that apply to them (i.e., the number of "yes" responses) without answering each question individually. Consequently, respondents' privacy is protected since their true sensitive status remains undisclosed unless they report that none or all of the items in the list apply to them. By contrast, units belonging to short list (SL) sample, s_{sl}, are asked to make a similar response to a SL of items, containing only the G innocuous questions which are identical to those present in the LL. The innocuous items should be chosen and worded in sufficient quantity as to ensure the necessary variability in their application to the units in the population.

Without loss of generality, let $\mathcal{T}$ be the variable denoting the total score applicable to the G non-sensitive questions, and $\mathcal{Z} = \mathcal{Y} + \mathcal{T}$ the total score applicable to the non-sensitive questions and the sensitive question. Hence, the answer given by the ith respondent will be

$$z_i = \begin{cases} y_i + t_i & \text{if } i \in s_{ll} \\ t_i & \text{if } i \in s_{sl} \end{cases}$$

The answers given by samples s_{ll} and s_{sl} are then pooled to obtain an estimate of the prevalence π of units bearing the sensitive attribute A. Under simple random sampling design an unbiased estimator of π is termed the difference-in-means estimator, and is obtained as the difference between the means of the answers in sample s_{ll} and in sample s_{sl}:

$$\hat{\pi} = \hat{\mu}_{ll} - \hat{\mu}_{sl}.$$

Since the two samples are independents, the variance can be obtained as the sum of the variance of the two groups means. In the same way, the estimated variance can be estimated.

3.6 Advances in estimation by the item sum technique using auxiliary information in complex surveys.

3.6.1 Estimation under a generic sampling design: the Horvitz-Thompson-type estimator

Consider a finite population $U = \{1, \ldots, N\}$ consisting of N different and identifiable units. Let y_i be the value of the sensitive character under study, say $\mathcal{Y}$, for the ith population unit. Let us suppose that the population mean $\bar{Y} = N^{-1} \sum_{i \in U}^{N} y_i$ is unknown and has to be estimated in an IST setting.

Chaudhuri and Christofides (2013) introduced the IST in the following way: one of the samples, say s_{ll}, is confronted with a LL of items containing $(G+1)$ questions of which G refer to non-sensitive characteristics and one is related to the sensitive characteristic under study. The other sample, s_{sl}, receives a SL of items that only contains the G innocuous questions present in the LL-sample. All sensitive and non-sensitive items are quantitative in nature. Respondents in each sample are requested to report the total score of all the items applicable to them, without revealing the individual scores for the items.

Assume that s_{ll} and s_{sl}, are selected from U according to the generic sampling designs $p_{ll}(\cdot)$ and $p_{sl}(\cdot)$ with positive first- and second-order inclusion probabilities $\pi_{i(ll)} = \sum_{s_{ll} \ni i} p_{ll}(s_{ll})$, $\pi_{ij(ll)} = \sum_{s_{ll} \ni i,j} p_{ll}(s_{ll})$, $\pi_{i(sl)} = \sum_{s_{sl} \ni i} p_{sl}(s_{sl})$ and $\pi_{ij(sl)} = \sum_{s_{sl} \ni i,j} p_{sl}(s_{sl})$ with $i, j \in U$. Let $d_{i(ll)} = \pi^{-1}_{i(ll)}$, $d_{ij(ll)} = \pi^{-1}_{ij(ll)}$, $d_{i(sl)} = \pi^{-1}_{i(sl)}$ and $d_{ij(sl)} = \pi^{-1}_{ij(sl)}$ denote the known sampling design-basic weight for unit $i \in U$ in each sampling design.
Without loss of generality, let $\mathcal{T}$ be the variable denoting the total score applicable to the G non-sensitive questions, and $\mathcal{Z} = \mathcal{Y} + \mathcal{T}$ the total score applicable to the non-sensitive questions and the sensitive question. When $G = 1$, $\mathcal{T}$ denotes the innocuous variable and t_i its value on unit $i \in U$. Hence, the answer given by the ith respondent will be

$$z_i = \begin{cases} y_i + t_i & \text{if } i \in s_{ll} \\ t_i & \text{if } i \in s_{sl} \end{cases}$$

Under the sampling designs $p_{ll}(\cdot)$, $p_{sl}(\cdot)$ let:

$$\hat{\bar{Z}}_{HT} = \frac{1}{N} \sum_{i \in s_{ll}} d_{i(ll)} z_i \quad , \quad \hat{\bar{T}}_{HT} = \frac{1}{N} \sum_{i \in s_{sl}} d_{i(sl)} t_i$$

be the unbiased 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 readily obtained as:

$$\hat{\bar{Y}}_{HT} = \hat{\bar{Z}}_{HT} - \hat{\bar{T}}_{HT}$$

From the unbiasedness of $\hat{\bar{Z}}_{HT}$ and $\hat{\bar{T}}_{HT}$, it readily follows that the estimator $\hat{\bar{Y}}_{HT}$ is unbiased for $\bar{Y}$. Furthermore, as long as the two samples are independent, the variance of $\hat{\bar{Y}}_{HT}$ can be expressed as:

$$V(\hat{\bar{Y}}_{HT}) = V(\hat{\bar{Z}}_{HT}) + V(\hat{\bar{T}}_{HT}) = \frac{1}{N^2}\left(\sum\sum_{i,j\in U}\Delta_{ij(ll)}d_{i(ll)}z_i d_{j(ll)}z_j + \sum\sum_{i,j\in U}\Delta_{ij(sl)}d_{i(sl)}t_i d_{j(sl)}t_j\right),$$

where $\Delta_{ij(a)} = \pi_{ij(a)} - \pi_{i(a)}\pi_{j(a)}$ with $a = ll, sl$. An unbiased estimator of $V(\hat{\bar{Y}}_{HT})$ is given by:

$$\hat{V}(\hat{\bar{Y}}_{HT}) = \frac{1}{N^2}\left(\sum\sum_{i,j\in s_{ll}} d_{ij(ll)}\Delta_{ij(ll)}d_{i(ll)}z_i d_{j(ll)}z_j + \sum\sum_{i,j\in s_{sl}} d_{ij(sl)}\Delta_{ij(sl)}d_{i(sl)}t_i d_{j(sl)}t_j\right)$$

3.6.2 Estimation in the presence of auxiliary information: the calibration-type estimator

We assume that a vector x of Q auxiliary variables is available from different sources such that the vector of values $x_i = (x_{i1}, \ldots, x_{iQ})^t$ is known $\forall i \in U$. Additionally, let $\bar{X} = N^{-1}\sum_{i\in U} x_i$ denote the vector for the known population means of the Q auxiliary variables. In order to obtain a calibration estimator of $\bar{Y}$ in the IST setting, we follow Deville and Särndal (1992) to obtain a new system of weights ω_{ij} based on sample s_j, $j = ll, sl$, by minimizing the χ^2 distance function

$$\Phi_{s_j}(d_i, \omega_{ij}) = \sum_{i\in s_j} \frac{(\omega_{ij} - d_i)^2}{d_i q_i}, \quad j = ll, sl$$

subject to the calibration equations

$$\frac{1}{N}\sum_{i\in s_j} \omega_{ij} x_i = \bar{X},$$

where the q_i's are known positive constants unrelated to the d_i's.
According to the calibration weights, we define a calibration-type estimator of $\bar{Y}$ as:

$$\hat{\bar{Y}}_C = \hat{\bar{Z}}_C - \hat{\bar{T}}_C,$$

where

$$\hat{\bar{Z}}_C = \frac{1}{N} \sum_{i \in s_{ll}} \omega_{i(ll)} z_i$$

is the calibration estimator of $\bar{Z}$ obtained on the basis of the LL-sample s_{ll} and

$$\hat{\bar{T}}_{\mathrm{C}} = \frac{1}{N} \sum_{i \in s_{sl}} \omega_{i(sl)} t_i$$

is the calibration estimator of $\bar{T}$ obtained from the SL-sample s_{sl}.
Following Deville and Särndal (1992), it can be shown that the estimator $\hat{\bar{Y}}_C$ is asymptotically unbiased for $\bar{Y}$ and its asymptotic variance and an estimator for this variance are obtained.

3.6.3 Estimation for domains

Let $U_d \subset U$ denote a domain of interest of N_d units, δ_{di} the domain identifier taking the value 1 if $i \in U_d$, and s_{jd} the subset of s_j containing units from U_d, $s_{jd} = s_j \cap U_d$, with $j = ll, sl$.
In order to obtain an estimate of the domain mean $\bar{Y}_d = N_d^{-1} \sum_{i \in U_d} y_i$, let us consider, the HT-type estimator defined as:

$$\hat{\bar{Y}}_{\mathrm{HT},d} = \frac{1}{N_d} \sum_{i \in s_{(ll)d}} d_i z_i - \frac{1}{N_d} \sum_{i \in s_{(sl)d}} d_i t_i$$

This estimator is design-unbiased and we obtain the variance and the unbiased estimated variance.
The domain calibration-type estimator can be defined as:

$$\hat{\bar{Y}}_{C,d} = \frac{1}{N_d} \sum_{i \in s_{ll}} \omega_{i(ll)} z_i \delta_{di} - \frac{1}{N_d} \sum_{i \in s_{sl}} \omega_{i(sl)} t_i \delta_{di},$$

where weights ω_{ij}, $j = ll, sl$, are determined by minimizing the χ^2 distance function

$$\Phi_{s_{jd}}(d_i, \omega_{ij}) = \sum_{i \in s_{jd}} \frac{(\omega_{ij} - d_i)^2}{d_i q_i}, \quad j = ll, sl$$

subject to the conditions

$$\bar{X}_{U_d} = \frac{1}{N_d} \sum_{i \in U_d} x_i = \frac{1}{N_d} \sum_{i \in s_j} \omega_{ij} x_i \delta_{di}$$

and

$$N_d = \sum_{i \in s_j} \omega_{ij} \delta_{di}, \quad j = ll, sl.$$

The expressions of its variance, and of the variance estimator can easily be obtained.

3.7 Multiple sensitive estimation and optimal sample size allocation in the item sum technique

3.7.1 Multiple sensitive estimation under IST

To obtain a reliable estimation, three different approaches are considered:

- Separate approach: to perform Q separate IST surveys, one for each sensitive item. This approach requires for each item the selection of one LL-sample and one SL-sample, for a total of $2Q$ samples.

- All-in-one approach: a single IST survey could be performed. In this case, just one LL-sample and one SL-sample are selected and respondents are asked to participate in Q separate IST experiments, one for each sensitive item.

- Mixed approach: which requires the selection of $Q+1$ samples and acts as a trade-off between the first two approaches.

Mixed approach

Let us focus on Q quantitative sensitive variables, $\mathcal{Y}_1, \dots, \mathcal{Y}_Q$, and on one innocuous variable $\mathcal{T}$. We want to estimate the mean of the variables, say $\bar{Y}_1, \dots, \bar{Y}_Q$. Under this approach, $Q+1$ independent samples are selected. For ease of notation, let us suppose that the same sampling design $p(\cdot)$ is used. Hence, let:

- s_0 be a sample of size n_0. The respondents are given a SL containing only the innocuous variable. The ith respondent provides the score t_{i_0} with $i = 1, \dots, n_0$;

- s_1 be a sample of size n_1. The respondents are given a list containing one sensitive variable, for instance $\mathcal{Y}_1$, and the innocuous one. The ith respondent provides the total score $y_{1i_1} + t_{i_1}$ with $i = 1, \dots, n_1$;

- s_2 be a sample of size n_2. The respondents are given a list containing the two sensitive variables and the innocuous one. The ith respondent provides the total score $y_{1i_2} + y_{2i_2} + t_{i_2}$ with $i = 1, \dots, n_2$.

- ...

Let

$$\hat{\bar{Z}}_k = \frac{1}{N}\sum_{i_k \in s_k} \frac{z_{i_k}}{\pi_{i_k}} = \frac{1}{N}\sum_{i_k \in s_k} \frac{\sum_{j=1}^{Q} y_{ji_k} + t_{i_k}}{\pi_{i_k}},$$

with $k = 1, \ldots, Q$. Hence, the estimator

$$\hat{\bar{Y}}_k^* = \hat{\bar{Z}}_k - \hat{\bar{Z}}_{k-1}$$

is the HT-unbiased estimator of $\bar{Y}_k$, $k = 1, \ldots, Q$. The variance of this estimator and an unbiased estimator for this variance are obtained.

Similarly, $G > 1$ innocuous variables, say $\mathcal{T}_1, \ldots, \mathcal{T}_G$, can be considered. In this case, $\mathcal{T}$ denotes the total score of the values of the G innocuous variables and $t_{i_k} = \sum_{g=1}^{G} t_{gi_k}$ is the total score of the G innocuous variables for the i_kth respondent in the kth sample s_k.

3.7.2 Total sample size allocation in the IST estimation

We assume that the total sample size n is fixed beforehand. Hence, the problem of optimal sample allocation is formulated as one of determining the LL-sample and SL-sample sizes, n_{ll} and n_{sl}, in such a way as to minimize the variance of $\hat{\bar{Y}}_{HT}$ subject to a fixed cost C.

Allocation under a generic sampling design

We provide a solution to this allocation problem for the case in which the sampling designs $p_{ll}(\cdot)$ and $p_{sl}(\cdot)$ provide a variance of the estimator which can be formulated as:

$$V(\hat{\bar{Y}}_{HT}) = \frac{A_z}{n_{ll}} + \frac{A_t}{n_{sl}} + B,$$

where the terms A_z, A_t and B do not depend on n_{ll} and n_{sl}.
The simple random sampling and the stratified random sampling designs meet this requirement.

Allocation in multiple IST estimation

- Separate approach: optimal sample size allocation is obtained for each IST survey by minimizing the variance of the estimator of the sensitive mean corresponding to the variable referred to by the IST survey.

- All-in-one approach: just one sample is selected for the entire survey on the Q sensitive questions. This sample must then be optimally split into the LL-sample and SL-sample, and so the initial question is to decide how this optimality is to be achieved.

- One possibility is to focus on one of the Q sensitive variables, perhaps the most relevant variable - if any - for the survey, and then to minimize the variance of the estimator of its mean. Obviously, however, obtaining the optimal sample size allocation for the variable considered does not ensure variance reduction in estimating the mean of the remaining variables.
- To overcome this limitation, a more general solution that involves all the study variables might be considered. Since multiple estimation leads to Q estimators of the Q population means of the sensitive variables under investigation, we may opt to minimize the variance of a convex combination of the Q variances of the estimators.

- Mixed approach: to find the optimal sample size allocation by minimizing the variance of one estimator is unfeasible since this will allocate the entire total size n between two samples, leaving a zero size for the remaining $Q-1$ samples. The only solution to this problem is to minimize the convex combination of the Q variances of the estimators.

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

We carry out a survey in Spanish universities to investigate patterns of cannabis consumption and sexual addiction. 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.

In particular, we aim to evaluate the effectiveness of the IQTs, specifically BarLev model and IST, in comparison with the DQ survey mode.

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

The sensitive questions are:

- 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?

To collect sensitive information using the BarLev method, we use as a randomizing device the smartphone application of the "Baraja Española", a deck composed of 40 cards, divided into 4 families or suits, each numbered from 1 to 7, and 3 figures for the each suit. For each sensitive question, the students are asked to run the application and to give the true sensitive response if the card shown is a figure. If the screen does not show a figure, the students are asked to multiply the real sensitive value of the response by the number shown on the card.

For the IST, four different non-sensitive questions, each corresponding to one of the sensitive questions, are formulated. The innocuous questions are:

- 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?

Hence, the students who are assigned to the IST receive two different questionnaires, depending on whether they belong to the SL-sample or the LL-sample.

For both the BarLev method and the IST, when the questionnaires are distributed, the students are assured of the confidentiality of their responses. It is emphasized 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 is generated by the mobile application or the individual score to the LL-questions.

3.9 Indirect questioning methods reveal hidden support for female genital cutting in South Central Ethiopia

A socio-demographic household survey is undertaken with adults living in rural sub-districts of Arsi and East Shewa zones, Southern Oromia. Research and Ethical approval to undertake this study is granted by the Ethics Committees at the University of Addis Ababa and the University of Bristol.

To compare openly-declared and privately-held support for FGC, the survey employs DQ on the desirability of FGC, as well as the UCT. All respondents are asked about the desirability of FGC for both a hypothetical daughter, and a prospective daughter-in-law.

There are four different versions of the survey which are randomly assigned to respondents, these include direct and indirect questions, a control and treatment condition (lists with and without the sensitive item, FGC). 70% of the sample undertakes a survey with the indirect UCT question, with participants equally and randomly assigned to either a control or a treatment condition. The remaining 30% of the sample answers direct questions.
All four versions of the survey are randomly assigned across households, and within household by gender and marital status. Interviewers then travel house to house, administering surveys to alternate households selected from a village plan supplied by the local district administrators. Accordingly, a random sample of 50% the households in the community are surveyed.

A novel aspect of this UCT study design is the use of cards with pictures for each item included in the list, allowing randomized presentation of the list items and improved respondent comprehension.

Further, the items included in the UCT list are carefully chosen to so as to minimize the chance of floor and ceiling effects, that is, of participants preferring either all or none of the items.

To contrast the proportions between the DQ method and UCT, and for subgroups (in both DQ and UCT methods) we develop the contrast of equal proportions proposed in Wolter and Preisendörfer (2013). We also perform multivariate analyses using generalized regression models, with and without iterations of the covariates, studied by Blair and Imai (2010); Imai (2011), and Blair and Imai (2012).

Chapter 4

Results

Some important results have been derived from the research carried out in this thesis. The most notable ones are summarized below.

4.1 An improved class of estimators in RR surveys

It is proved that

- Any estimator into the class is asymptotically unbiased for Y.
- An approximation of the bias of the proposed class of estimators is obtained.
- The asymptotic variance of any estimator into the class is also defined.
- Assuming simple random sampling, the estimators are asymptotically unbiased and normally distributed.
- Under stratified sampling,
  - we can consider a general class of estimators in each stratum and by an estimator into the class is achieved the minimum asymptotic variance. The properties of this estimator can be easily obtained by using the independence of sampling en each stratum.
  - the formulae are based on the assumption that the sample size in each stratum is large. This, however, is not always true in practice. To get over this difficulty, we suggest a general class of combined estimators and the asymptotic variance of any estimator into the class is obtained.

- Some different estimators based on information of auxiliary variables are extended to our case of RR questioning.

We consider two studies with real and simulated populations and the main conclusions derived in these studies are:

- The superiority of estimators based on auxiliary information is clear, the suggested estimators belonging to the class $\widehat{Y}_g^{(r)}$ are always more efficient than the HT estimator, whatever the adopted scrambling procedure.
- More efficient estimators values are obtained if the correlations between the auxiliary variables and the principal are high.
- The relative bias (RB) of the estimators are all within a reasonable range for the different sample sizes considered.
- The values of RB and mean square error (MSE) decrease as the sampling size increases, for all estimators and all RRT.
- Difference estimator or exponentiation estimator are the most efficient estimators for using one or two auxiliary variables, and with two auxiliary variables perform better than the estimators with one auxiliary variable, as expected.
- The values of RB and MSE are very similar between difference and exponentiation estimators. The difference in RB and MSE between these estimators is smaller as the sample size increases. This is expectable because the two estimators are asymptotically equivalents.

4.2 Application of randomized response techniques for investigating cannabis use by Spanish university students

- The non-response rates for the questions are significantly lower in the RR than in the DQ condition (p-value < 0.001). However, the non-response rates between men and women are similar and therefore not significant statistically (p-value > 0.05).
- The result obtained in the first question is, by DQ, the mean number of cannabis cigarettes consumed in the previous year is approximately three (95% confidence interval (CI) [2.0181 - 4.2103]), but according to RR, seventeen units are consumed (95% CI [9.7903 - 24.2119]). And in the second, by DQ, the students had consumed cannabis on approximately one of the previous 90 days (95% CI [0.3902 - 0.9773]), and on seven according to RR (95% CI

[5.1819 - 8.8538]).

- The estimate of the number of cannabis cigarettes consumed and the estimate of the number of days when consumption took place for the RR group are significantly higher than the estimates for the DQ group (p-values < 0.001).
- For all questions, the standard deviation is higher for the RR than for the DQ survey. This result is as we expect because the randomization mechanism of RRT increases the variability of the estimate, so surveys conducted with RRT require large sample sizes.
- If we consider the results by gender, we get more units of cannabis consumed and more number of days of consuming in men than women. This difference is statistically significant by DQ (p-value $= 3.8 \times 10^{-5}$ and 0.002 respectively) but this difference is not statisitically significant for RR (p-value $= 0.105$ and 0.108 respectively).

4.3 RRTCS: An R package for randomized response techniques in complex surveys

- RRTCS is a new R package to perform point and interval estimation of linear parameters with data obtained from complex surveys when randomization techniques are used. Estimators and variances for 14 RR methods for qualitative variables and 7 RR methods for quantitative variables are also implemented. The package also includes an additional function which provides estimates variance of the RR estimators using some resampling methods. In addition, some data sets from surveys with these randomization methods are included in the package.
- The package is freely available at the CRAN repository following the URL https://CRAN.R-project.org/package=RRTCS.
- A reference manual including information about all the functions composing the package and a vignette illustrating how to use it in different contexts can be also found.

4.4 Advances in estimation by the item sum technique using auxiliary information in complex surveys.

- We obtain the HT-type estimator of $\bar{Y}$ in IST under a generic sampling designs and we prove its unbiasedness. We compute the unbiased estimator of the variance of this estimator.

- We define a calibration-type estimator of $\bar{Y}$ in IST and show that it is asymptotically unbiased for $\bar{Y}$. Also we compute its unbiased estimated variance.
- We obtain an estimate for the domain mean in IST through HT-type estimator and calibration-type estimator.

Two simulation studies to numerically investigate the performance of the HT and calibration-type estimators when sensitive quantitative data are obtained by the IST, are considered.

- The first study is designed to:
  - Compare the proposed IST estimators and a RRT estimator which uses two different scrambling variables.

    The results obtained on the simulation study are:
    * While the HT estimator based on the true values slightly outperforms compared to the HT-type estimator based on the IST; the calibration estimators are unexpectedly nearly equivalent, both in terms of absolute relative bias (RB) and of mean squared error (MSE). These findings highlight the successful use of auxiliary information at the IST estimation stage.
    * In general, the IST seems to outperform the RRT approach, at least for the scrambling models considered in the present study, both for HT and for calibration-type estimates.
    * For all the estimators, the absolute RB falls within a reasonable range.
    * The MSE of the estimators tendentially decreases as the sample size increases, which is an evident indication of the consistency of all the estimates produced.
    * For all the estimators considered, it is also evident that using auxiliary information at the design stage through stratification and sampling with varying probability can improve the efficiency of the estimates with respect to SRSWOR.
  - Evaluate, within the IST framework, the effects of using innocuous items with different correlations with the target sensitive variable.

    * We observe that if HT-type estimators employ innocuous variables highly correlated with target variable, the efficiency of the estimates decreases. Hence, the choice of which innocuous variable to use is a matter of some importance for the researcher.

    * On the contrary, no striking differences are apparent when the IST calibration estimators are considered, and the results appear to be robust to the choice of the innocuous variable.
  - Evaluate the performance of the IST for domain estimation.
    The results obtained are very similar to those of the first simulation study and confirm that the IST can also be profitably used in more complex survey situations.

- The second simulation study focuses on the calibration approach and explores:
  - The influence on the estimates of the length of the list.
    * The performance of the estimators strongly depends on the length of the list. As the number of innocuous items increases, both the absolute RB and the relative MSE (RMSE) increase, although the RB always remains within an acceptable range of values.
    * The best performance of the estimators is achieved when one or two innocuous variables are used to perturb the true sensitive response.
  - The accuracy of the variance estimation.
    * Overall, both the absolute RB and the RMSE of the variance estimator for the suggested IST calibration estimator produce very small values.
    * The RMSE decreases as the sample size increases.
    * The satisfactory behaviour of the variance estimator does not seem to be affected by the increased number of innocuous variables used to perform the IST.

4.5 Multiple sensitive estimation and optimal sample size allocation in the item sum technique

- We consider the problem of how to reduce the statistical burden on respondents when $Q \geq 2$ sensitive variables are surveyed and the population means need to be estimated. We discuss some estimation methods for multiple sensitive questions under differents approaches, named separate, all-in-one and mixed. In the case of mixed approach we obtain the variance and the unbiased estimated variance.
- The optimal allocation of the total sample size into the LL-sample and the SL-sample is discussed. First, we consider a method of allocation based on minimizing the variance

of the IST estimator of the mean of one sensitive variable which is valid under a budget constraint and for a general sampling design. Thus, explicit expressions for the sampling fractions have been worked out when SRSWOR and stratified sampling are used. The allocation method has been then extended to the case of Q sensitive variables under the all-in-one and mixed approaches.

We run a number of simulation studies to evaluate the performance of the optimal allocation in different situations.

- Optimal allocation IST estimates versus DQ
  The variance of the IST estimator with optimal sample size allocation is compared with that of the sample mean estimator by difference and ratio estimator.
  - The variance of the IST estimator is higher than that of the sample mean estimator under DQ.
  - The difference becomes negligible as the sample size increases, while the ratio highlights the fact that the loss of efficiency remains within acceptable limits especially when the correlation between the sensitive variable and the innocuos one (ρ) is low.
  - Moreover, for a fixed sample size, the difference and the ratio increase with ρ.
- Optimal versus arbitrary IST allocation
  We consider the ratio between the variance of the optimal allocation IST estimator and that of the IST estimator arbitrarily obtained assuming $n_{ll} = \lambda n$ and $n_{sl} = (1 - \lambda)n$, $\lambda = 0.5, 0.6$.
  - The improved efficiency is evident in both situations.
  - The correlation coefficient does not appear to significantly affect the variance of the IST estimators.
- Optimal IST allocation in stratified SRSWOR
  The variance of the estimates under optimal allocation is compared by ratio estimator using two different forms of allocation:
  - Arbitrary allocation: In stratified IST with two strata, four samples are considered. From each stratum a LL-sample and a SL-sample are selected. Hence, we trivially assume the sample size for each sample is $n/4$.

  - Naive two-step optimal allocation: Allocation is conducted in two steps, separately determining the optimal IST allocation in each stratum. In the first step, a stratified sample of each stratum is selected with proportional allocation. In the second step, each of the two first-step samples is optimally allocated in the LL-sample and SL-sample.

  It can be seen that:

  - arbitrary allocation is not at all efficient.
  - the results obtained with naive two-step optimal allocation are almost identical to those attainable with the theoretical optimal allocation.

- Stratified versus SRSWOR
  We compare the efficiency of stratified and SRSWOR IST estimates under optimal allocation.
  The results reflect the considerable gain in efficiency achieved by stratifying the population.

- Optimal allocation in multiple IST estimation
  We compare the IST estimates under the separate, all-in-one and mixed approaches focusing on two sensitive variables:

  - Under the separate approach, the optimal sample allocation for n_{ll} and n_{sl} is separately considered for each of the two variables in such a way that the estimates for $\bar{Y}_1$ and $\bar{Y}_2$ both attain their minimum variance bound.
  - In all-in-one estimates, the optimal sample sizes n_{ll} and n_{sl}, which minimize the variance are used to obtain the estimates of $\bar{Y}_1$ and $\bar{Y}_2$.
  - In mixed approach, the three sample sizes n_0, n_1 and n_2 are optimally determined to minimize the variance and then used in the single estimators $\hat{\bar{Y}}_1^*$ and $\hat{\bar{Y}}_2^*$.

  The results obtained are:

  - It is immediately apparent that both the absolute RB and the relative variance decrease as the sample size increases, which is a clear indication of the consistency of the estimates under the three approaches.
  - In general, the three approaches produce equivalent results in estimating the mean of the sensitive variables. As the sample size increases, the difference between the methods decreases.

- For the situations considered in this analysis, the mixed approach seems to be competitive in terms of efficiency while clearly reducing the statistical burden on the respondents.

We compare the theoretical estimated variances of the estimators under the three approaches and the results obtained are in accordance with those obtained in the previous study.

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

- The first notable result is the significant reduction in the non-response rate in the case of the IQTs. As expected, the DQ non-response rate is higher for questions refer to sexual behaviour than for the questions refer to cannabis use. In general, the comparison of the two IQTs reveals that the IST non-response rate is statistically lower than that of the BarLev method.
- The normality of the estimates under the three survey methods is ascertained by investigating the sampling distribution of the estimators using a bootstrap resampling procedure.
- We assess whether the random assignment of the students to the three survey modes produces comparable groups of respondents by gender. The Chi-squared test of independence confirms the effectiveness of the random assignment.
- As expected, the DQ method produces an underestimation of the sensitive characteristics investigated. Thus, the DQ estimates are statistically lower than the IQT ones. The BarLev estimates are statistically higher than the IST ones for questions Q2 and Q3, and lower for question Q4, while no significant difference is 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.
- With respect to accuracy, in general, the IST estimates present lower standard errors and narrower confidence intervals than the BarLev method. As expected, the DQ estimates are more precise than the IQT ones.
- 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. Specifically,

in IST the number of times during the past 90 days that students had difficulty in halting inappropriate sexual behaviour is significantly greater for men than for women. The 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 are 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, 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). Moreover, in IST, male students smoked more cigarettes than female students (24.65 vs 6.48) and for more days (5.51 vs 2.17) .

4.7 Indirect questioning methods reveal hidden support for female genital cutting in South Central Ethiopia

The results obtained after the analysis show that:

- The UCT reveals that people privately have higher levels of acceptance of FGC behaviour (22.4%) than is admitted openly through DQ (7.7%), being this difference significant.
- Respondents report no difference in level of support for FGC for daughters than daughters-in-law both when asked directly, 7.3% and 8.2%, or indirectly using UCT 19.7% and 25%. There is, however, evidence of concealment of FGC support, i.e. greater difference between direct and indirect UCT estimates, for both categories of female relatives.
- Men and women report similar and low levels of support for FGC when asked directly (8.3% and 7.1%). Using UCT, women appear privately more supportive of the practice than men (men: 18.5%, women: 26.7%), but this difference is not significant.
- When asked directly, individuals in the two age-groups, <26, $>=26$, report similar, low support for FGC (7.2%, 8% respectively). Indirect estimates, however, indicate that private support is significantly higher among those aged over 26 (29.6%) than those aged under 26 (7.8%).

- When asked directly, uneducated respondents are more likely to admit support for FGC than those who have received formal education (ever attended school) (12.0% compared to 5.0%). UCT, however, suggests a reversal of this with uneducated respondents privately being less supportive of FGC than educated individuals (19.4% vs 23.8%); however this difference is not statistically significant.

- Our analyses reveal that high levels of private support for FGC are found among older, educated males, where estimated acceptance levels reach 45% (39.7% for daughters, and 50.4% for daughters-in-law). This category of individuals is also the least likely to openly admit a preference for FGC (1.4% and 4.1%); which is reflected in a significant difference between directly-expressed and privately-held views, the largest of any subgroup of individuals within this population.

Chapter 5

Conclusions

5.1 An improved class of estimators in RR surveys

A class of estimators of a population total under a general RR model is defined when the sample is obtained under a general sampling design. Estimators belonging to this class are proved to be asymptotically design unbiased, and their asymptotic variances are obtained. We provide the expression of an optimal estimator in the class, the difference estimator, that is the estimator that attains the asymptotic minimum variance bound. This estimator is studied for some elementary sampling design as simple random sampling and stratified sampling. We introduce other estimators in this class, some of them have asymptotically the same variance as the optimal difference estimator.

We have conducted a simulation study to check the performance of the proposed estimators. The results obtained from the simulation study show that, given a set of auxiliary variables, the method performs well under different scenarios in both real and artificial populations.

5.2 Application of randomized response techniques for investigating cannabis use by Spanish university students

We present a survey related to the use of cannabis, in which a RRT is used to determine population means. On comparing the results of the DQ survey and those of the RR survey we find that the number of cannabis cigarettes consumed during the past year (DQ = 3, RR = 17 approximately), and the number of days when consumption took place (DQ = 1, RR = 7 approximately) are much higher with RRT in these universities. The results obtained suggest that estimates derived from standard questionnaire forms underestimate the incidence of drug use by university students. It must be stressed, however, that RR has wide confidence intervals.

The randomization procedure introduces additional random error into the data and increases the standard errors of the parameters estimated, thus, larger sample sizes are needed in order to increase the statistical power.
Another important issue in RRT is the choice of an appropriate randomizing device, which should be implemented in such a way as to make the confidentiality protection offered very clear to the respondent. The new technologies currently available offer alternatives that are more attractive to users, such as mobile phones. Thanks to smartphones, we have access to many interesting applications that can help in the randomization of telephone and personal surveys, especially among young people.

5.3 RRTCS: An R package for randomized response techniques in complex surveys

The need for a software for analyzing the data from RRTs for complex surveys led to the development of the R package RRTCS.
The latest version of the package, as well as documentation and illustrative examples of its use may be freely accessed through the URL https://CRAN.R-project.org/package=RRTCS.

5.4 Advances in estimation by the item sum technique using auxiliary information in complex surveys.

We describe advances that may be achieved in the use of the IST under a generic sampling design, including and not including auxiliary information when it is available for the entire population, at no additional cost.
Our findings reveal that IST surveys can provide estimates which are nearly as efficient as those obtained from a DQ survey while, in general, outperforming RRT estimates. This is particularly true for the calibration-type estimators. If we consider the perfomance of the IST for the domain estimation, the results are maintained. In the calibration-type estimator the results seem to be robust to the choice of the correlation between the innocuous and the sensitive variable, while that in HT-type estimator the efficiency depends on this correlation.
Additionally, we further investigated the behaviour of these estimators by running additional simulations in order to assess variance estimation and the impact made on the estimates when the number of innocuous variables is increased, obtaining that the best performance of the estimators is achieved when one or two innocuous variables are used.

5.5 Multiple sensitive estimation and optimal sample size allocation in the item sum technique

An extensive simulation study is conducted to investigate the performance of the proposed techniques and the related estimators under different sampling designs and for different sample sizes. All the situations examined reflect the benefits of determining the optimal sample size, which can significantly increase the efficiency of the estimates with respect to any arbitrary allocation of the sample units and could provide estimates which are nearly as accurate as those obtained by DQ, and without jeopardizing respondents' confidentiality.
A very interesting result is achieved when optimal allocation is used for multiple IST estimation purposes under the mixed approach. In this case, in relation to the marked reduction obtained in the statistical burden placed on respondents and in survey costs, the loss of efficiency with respect to the all-in-one and separate approaches may be considered very modest or even negligible. Hence, from a theoretical standpoint, the mixed approach appears to be a viable alternative for the purposes of multiple IST estimation.
We conclude by observing that all the ideas, the methodological advances and the results presented regarding the IST may be easily extended to its forerunner, the ICT, which, although it is a more widespread and long-established technique, suffers from the same drawbacks that are discussed with respect to the IST.

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

The DQ survey mode produces non-response rates that are higher than the IQT ones. In turn, the IST non-response rates are lower than the BarLev ones.
Moreover, the DQ method produces underreporting of the sensitive behaviours under study (cannabis use and sexual addiction) if we compare them with IQTs, but DQ estimations are more precise. If we compare BarLev and IST there are no evidence of a uniform superiority of one method over the other, but in general, the IST estimates present more accuracy than the BarLev method.
If we consider the gender of the student, in IST, the number of cannabis cigarettes consumed during the past year and the number of days when consumption took place are higher in male students than in female and in sexual addiction, the male group had difficulties in halting innappropriate sexual behaviour for more days than 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. When significant underreporting is produced by DQ, researchers and practitioners actively engaged in organizing, managing and conducting sensitive studies should suspect 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.

5.7 Indirect questioning methods reveal hidden support for female genital cutting in South Central Ethiopia

Our results demonstrate the inadequacy of traditional, yet widely used, DQ methods, and the potential for IQT to improve understanding of culturally-sensitive topics, like FGC.
Comparing direct and indirect response methods in rural Oromia, South Central Ethiopia, we identify substantial underreporting of support for FGC using DQ methods. Across the community, privately-held views in favour of FGC are approximately three times higher than those admitted when asked directly by an interviewer. We identify that older individuals hold the strongest views in favour of FGC, but are also the most likely to hide their 'true' support for the practice when questioned directly. The lowest concealed support for FGC is among the youngest cohort ($<$26-year olds). The results also indicate that educated Arsi Oromo give more socially desirable answers than those individuals without schooling, hiding their 'true' FGC intentions when questioned directly. Besides our results reveal that both men and women are equally supportive of FGC in our sample, and attempt to conceal their support in front of interviewers. Otherwise we find no clear evidence for weaker support for FGC for daughters than for daughters-in-law, in line with an evolutionary prediction that parents will be more concerned with controlling the sexual behaviour of their daughters-in-law. Finally, our results suggest that it is elders, particularly educated men who hold some of the strongest views in favour of the practice, $>$45% privately endorse FGC, but these views are hidden when asked directly. This group represents around 12% of the total population, and hold positions of authority in the community. Concealed support and pressure to continue FGC from this powerful and influential group could explain the stubborn persistence of the practice in this and similar communities.

Chapter 6

Conclusiones

6.1 An improved class of estimators in RR surveys

Se define una clase de estimadores para el total poblacional bajo un modelo de respuesta aleatoria general cuando la muestra se obtiene bajo un diseño de muestreo general. Se demuestra que los estimadores que pertenecen a esta clase son asintóticamente insesgados, y se obtienen sus varianzas asintóticas. Proporcionamos la expresión de un estimador óptimo en la clase, el estimador de diferencia, que es el estimador que alcanza la varianza mínima asintótica. Este estimador se estudia para algunos diseños de muestreo elementales como muestreo aleatorio simple y muestreo estratificado. Presentamos otros estimadores en esta clase, algunos de ellos tienen asintóticamente la misma varianza que el estimador de diferencia óptimo.

Hemos llevado a cabo un estudio de simulación para verificar el rendimiento de los estimadores propuestos. Los resultados obtenidos del estudio de simulación muestran que, dado un conjunto de variables auxiliares, el método funciona bien en diferentes escenarios en poblaciones reales y artificiales.

6.2 Application of randomized response techniques for investigating cannabis use by Spanish university students

Presentamos una encuesta relacionada con el consumo de cannabis, en la que se utiliza una RRT para determinar medias poblacionales. Al comparar los resultados de la encuesta de DQ con los de la encuesta de RR encontramos que el número de cigarrillos de cannabis consumidos durante el año pasado (DQ = 3, RR = 17 aproximadamente) y el número de días en que se consumió (DQ = 1, RR = 7 aproximadamente) son mucho más altos con la RRT en estas universidades. Los resultados obtenidos sugieren que las estimaciones derivadas de los cuestionarios estándar

subestiman la frecuencia del consumo de drogas por parte de los estudiantes universitarios. Debe destacarse, sin embargo, que la RR tiene amplios intervalos de confianza. El procedimiento de aleatorización introduce un error aleatorio adicional en los datos e incrementa los errores estándar de los parámetros estimados, por lo tanto, se necesitan tamaños de muestra más grandes para aumentar la potencia estadística.
Otro tema importante en la RRT es la elección de un dispositivo de aleatorización adecuado, que debe implementarse de forma que haga la protección de la confidencialidad ofrecida muy clara para el encuestado. Las nuevas tecnologías disponibles actualmente ofrecen alternativas que son más atractivas para los usuarios, como los teléfonos móviles. Gracias a los smartphones, tenemos acceso a muchas aplicaciones interesantes que pueden ayudar en la aleatorización de encuestas telefónicas y personales, especialmente entre los jóvenes.

6.3 RRTCS: An R package for randomized response techniques in complex surveys

La necesidad de un software para analizar los datos procedentes de técnicas de respuesta aleatoria para encuestas complejas permitió el desarrollo del paquete de R RRTCS.
Se puede acceder de forma gratuita a la última versión del paquete, así como a la documentación y ejemplos ilustrativos de su uso a través de la URL https://CRAN.R-project.org/package=RRTCS.

6.4 Advances in estimation by the item sum technique using auxiliary information in complex surveys.

Describimos los avances que se pueden lograr en el uso de la IST bajo un diseño de muestreo general, que incluye y no incluye información auxiliar cuando esta está disponible para toda la población, sin costo adicional.
Nuestros hallazgos revelan que las encuestas de IST pueden proporcionar estimaciones que son casi tan eficientes como las obtenidas de una encuesta de DQ, mientras que, en general, superan las estimaciones de la RRT. Esto es particularmente cierto para los estimadores de tipo de calibración. Si consideramos la realización de la IST para la estimación por dominios, los resultados se mantienen. En el estimador de tipo de calibración, los resultados parecen ser robustos a la elección de la correlación entre la variable inocua y la sensible, mientras que en el estimador de tipo HT la eficiencia depende de esta correlacin.
Además, investigamos más a fondo el comportamiento de estos estimadores ejecutando simulaciones adicionales para evaluar la estimación de la varianza y el impacto que se produce

en las estimaciones cuando aumenta el número de variables inocuas, obteniendo que el mejor rendimiento de los estimadores se alcanza cuando se usan una o dos variables inocuas.

6.5 Multiple sensitive estimation and optimal sample size allocation in the item sum technique

Se realizó un extenso estudio de simulación para investigar el rendimiento de las técnicas propuestas y los estimadores relacionados bajo diferentes diseños de muestreo y para diferentes tamaños de muestra. Todas las situaciones examinadas reflejan los beneficios de determinar el tamaño de muestra óptimo, lo que puede incrementar significativamente la eficiencia de las estimaciones con respecto a cualquier afijación arbitraria de las unidades de muestra y podría proporcionar estimaciones que son casi tan precisas como las obtenidas mediante DQ, y sin poner en peligro la confidencialidad de los encuestados.

Se alcanza un resultado muy interesante cuando la asignación óptima se utiliza para estimación múltiple en la IST bajo el enfoque mixto. En este caso, en relación con la notable reducción obtenida en la carga estadística impuesta a los encuestados y en los costos de la encuesta, la pérdida de la eficiencia con respecto a los enfoques todo en uno y separado puede considerarse muy modesta o incluso insignificante. Por lo tanto, desde un punto de vista teórico, el enfoque mixto parece ser una alternativa viable para el propósito de la estimación múltiple en la IST.

Concluimos observando que todas las ideas, los avances metodológicos y los resultados presentados con respecto a la IST pueden extenderse fácilmente a su precursor, la ICT, que, aunque es una técnica más extendida y establecida desde hace mucho tiempo, sufre los mismos inconvenientes que se discuten con respecto a la IST.

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

El modo de encuesta de DQ produce tasas de falta de respuesta más altas que las de las IQTs. A su vez, las tasas de falta de respuesta en la IST son más bajas que las de BarLev.

Por otra parte, el método de DQ produce una subestimación de las conductas sensibles en estudio (consumo de cannabis y adicción sexual) si las comparamos con las IQTs, pero las estimaciones de DQ son más precisas. Si comparamos BarLev y la IST no hay evidencia de superioridad uniforme de un método sobre el otro, pero en general, las estimaciones de la IST presentan más precisión que mediante el método de BarLev.

Si consideramos el género del estudiante, en la IST, el número de cigarrillos de cannabis consumidos durante el año pasado y el número de días en que se realizó el consumo son más altos en los hombres que en las mujeres y en la adicción sexual, el grupo de los hombres tuvo dificultades para detener su inapropiado comportamiento sexual durante más días que el de las mujeres. Desafortunadamente, no están disponibles los datos comparativos directamente para los fenómenos investigados en este estudio. No obstante, hay diferencias muy apreciables entre el método de encuesta de DQ tradicional y las IQTs. Cuando el DQ produce una significativa subestimación, los investigadores y profesionales involucrados activamente en la organización, gestión y realización de estudios sensibles deberían sospechar sobre la validez de los resultados. Al mismo tiempo, los operadores y los responsables políticos deberían proceder con cautela en la implementación de los programas de intervención porque los problemas sociales y de salud derivados del consumo de drogas y el comportamiento sexual pueden ser mucho más importantes de lo que se desprende de los informes de DQ.

6.7 Indirect questioning methods reveal hidden support for female genital cutting in South Central Ethiopia

Nuestros resultados demuestran la inadecuación de los métodos de DQ tradicionales, aunque ampliamente utilizados, y el potencial de las IQTs para mejorar la comprensión de temas culturalmente sensibles, como el FGC.
Al comparar los métodos de respuesta directa e indirecta en la comunidad rural de Oromia, centro sur de Etiopía, identificamos una subestimación sustancial de apoyo para el FGC utilizando métodos de DQ. En toda la comunidad, los puntos de vista privados a favor del FGC son aproximadamente tres veces más altos que los admitidos cuando se les pregunta directamente con un entrevistador. Identificamos que las personas mayores tienen los puntos de vista más fuertes a favor del FGC, pero también son los más propensos a ocultar su 'verdadero' apoyo a la práctica cuando se les pregunta directamente. El apoyo oculto más bajo para el FGC se encuentra entre la cohorte más joven (<26 años de edad). Los resultados también indican que los Arsi Oromo educados dan respuestas más deseables socialmente que aquellos individuos sin escolaridad, ocultando sus 'verdaderas' ideas sobre el FGC cuando son preguntados directamente. Además nuestros resultados revelan que tanto los hombres como las mujeres son igualmente partidarios del FGC en nuestra muestra, y tratan de ocultar su apoyo frente a los entrevistadores. Por otra parte no encontramos evidencia clara de un apoyo más débil para el FGC para las hijas que para las nueras, de acuerdo con una predicción evolutiva de que los padres estarán más interesados en controlar el comportamiento sexual de sus nueras. Finalmente, nuestros resultados sugieren

que son los mayores, especialmente los hombres educados quienes tienen algunos de los puntos de vista más fuertes a favor de la práctica, >45 % respaldan en forma privada el FGC, pero estos puntos de vista se ocultan cuando se les pregunta directamente. Este grupo representa alrededor del 12% de la población total, y tienen puestos de autoridad en la comunidad. El apoyo oculto y la presión para continuar con el FGC de este poderoso e influyente grupo podría explicar la obstinada persistencia de la práctica en esta y otras comunidades similares.

Bibliography

A. Arcos, M. Rueda, and S. Singh. A generalized approach to randomised response for quantitative variables. *Quality & Quantity*, 49:1239–1256, 2015.

A. Arentoft, K. Van Dyk, A.D. Thames, P. Sayegh, N. Thaler, D. Schonfeld, J. LaBrie, and C.H. Hinkin. Comparing the unmatched count technique and direct self-report for sensitive health-risk behaviors in HIV+ adults. *AIDS Care*, 28:370–375, 2016.

A. Arias and S.G. Sutton. Understanding recreational fishers' compliance with no-take zones in the Great Barrier Reef Marine Park. *Ecology and Society*, 18, 2013. doi: 10.5751/ES-05872-180418.

R. Arnab. Non-negative variance estimator in randomized response surveys. *Communications in Statistics - Theory and Method*, 23(6):1743–1752, 1994.

R. Arnab. Randomized response trials: A unified approach for qualitative data. *Communications in Statistics - Theory and Methods*, 25:1173–1183, 1996.

R. Arnab. Optional randomized response techniques for complex survey designs. *Biometrical Journal*, 46(1):114–124, 2004.

R. Arnab and G. Dorffer. Randomized response techniques for complex survey designs. *Statistical Papers*, 48:131–141, 2006.

R. Arnab and S. Singh. Randomized response techniques: An application to the botswana AIDS impact survey. *Journal of Statistical Planning and Inference*, 140:941–953, 2010.

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

L. Barabesi. A design-based randomized response procedure for the estimation of population proportion and sensitivity level. *Journal of Statistical Planning and Inference*, 138:2398–2408, 2008.

L. Barabesi and M. Marcheselli. A practical implementation and Bayesian estimation in Franklin's randomized response procedure. *Communications in Statistics - Simulation and Computation*, 35:563–573, 2006.

L. Barabesi, G. Diana, and P.F. Perri. Design-based distribution function estimation for stigmatized populations. *Metrika*, 76:919–935, 2013.

L. Barabesi, G. Diana, and P.F. Perri. Gini index estimation in randomized response surveys. *Asta-Advances in Statistical Analysis*, 99:45–62, 2015.

L. Bastiani, V. Siciliano, O. Curzio, C. Luppi, M. Gori, M. Grassi, and S. Molinaro. Optimal scaling of the CAST and of SDS scale in a national sample of adolescents. *Addictive Behaviors*, 38:2060–2067, 2013. doi: 10.1016/j.addbeh.2012.12.016.

Y.G. Berger and C.J. Skinner. A jackknife variance estimator for unequal probability sampling. *Journal of the Royal Statistical Society: Series B*, 67:79–89, 2005.

G. Blair and K. Imai. R package: statistical methods for item count technique and list experiment v.8.0., 2010.

G. Blair and K. Imai. Statistical analysis of list experiments. *Polit. Anal.*, 20:47–77, 2012.

G. Blair, K. Imai, and Y.Y. Zhou. Package rr: Statistical Methods for the Randomized Response, 2015. URL `http://CRAN.R-project.org/package=rr`.

S.G. Blank and M.C. Gavin. The randomized response technique as a tool for estimating non-compliance rates in fisheries: A case study of illegal red abalone (Haliotis rufescens) fishing in Northern California. *Environmental Conservation*, 36:112–119, 2009.

R.F. Boruch. Relations among statistical methods for assuring confidentiality of social research data. *Social Science Research*, 1:403–414, 1972.

M. Y. Chaloupka. Application of the randomized response technique to marine park management: An assessment of permit compliance. *Environmental Management*, 9:393–398, 1985.

A. Chaudhuri. Using randomized response from a complex survey to estimate a sensitive proportion in a dichotomous finite population. *Journal of Statistical Planning and Inference*, 94: 37–42, 2001.

A. Chaudhuri. Christofides' randomized response technique in complex sample survey. *Metrika*, 60:223–228, 2004.

A. Chaudhuri. *Randomized response and indirect questioning techniques in surveys.* Chapman & Hall, 2011.

A. Chaudhuri and T.C. Christofides. Item count technique in estimating the proportion of people with a sensitive feature. *Journal of Statistical Planning and Inference*, 137:589–593, 2007.

A. Chaudhuri and T.C. Christofides. *Indirect Questioning in Sample Surveys.* Springer-Verlag, 2013.

A. Chaudhuri and R. Mukerjee. *Randomized Response: Theory and Techniques.* Marcel Dekker, INC., 1988.

A. Chaudhuri, T.C. Christofides, and C.R. Rao. *Data Gathering, Analysis and Protection of Privacy through Randomized Response Techniques: Qualitative and Quantitative Human Traits*, volume 34. Elsevier, 2016.

T.C. Christofides. A generalized randomized response technique. *Metrika*, 57:195–200, 2003.

T.C. Christofides. Randomized response without a randomization device. *Advances and Applications in Statistics*, 11:15–28, 2009.

B. Cobo, M. Rueda, and A. Arcos. Package RRTCS: Randomized Response Techniques for Complex Surveys, 2015. URL `http://cran.r-project.org/web/packages/RRTCS/`.

A. Conteh, M.C. Gavin, and J. Solomon. Quantifying illegal hunting: A novel application of the randomised response technique. *Biological Conservation*, 189:16–23, 2015.

P. Cross, G. Edwards-Jones, H. Omed, and A.P. Williams. Use of a randomized response technique to obtain sensitive information on animal disease prevalence. *Preventive Veterinary Medicine*, 96:252–262, 2010.

A.M. Cuenca-Royo, A. Sánchez-Niubó, C. G. Forero, M. Torrens, J.M. Suelves, and A. Domingo-Salvany. Psychometric properties of the CAST and SDS scales in young adult cannabis users. *Addictive Behaviors*, 37:709–715, 2012. doi: 10.1016/j.addbeh.2012.02.012.

J.C. Deville and C.E. Särndal. Calibration estimators in survey sampling. *Journal of the American Statistical Association*, 87:376–382, 1992.

J.L. Devore. A note on the randomized response technique. *Communications in Statistics-Theory and Methods*, 6:1525–1529, 1977.

G. Diana and P.F. Perri. A new scrambled response models for estimating the mean of a sensitive quantitative character. *Journal of Applied Statistics*, 37:1875–1890, 2010.

G. Diana and P.F. Perri. A class of estimators for quantitative sensitive data. *Statistical Papers*, 52:633–650, 2011.

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

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

J.A. Droitcour and E.M. Larson. An innovative technique for asking sensitive questions: The three-card method. *Bulletin of Sociological Methodology*, 75:5–23, 2002.

J.A. Droitcour, R.A. Caspar, M.L. Hubbard, T.L. Parsley, W. Visseher, , and T.M. Ezzati. The item count technique as a method of indirect questioning: A review of its development and a case study application. In P.P. Biemer, R.M. Groves, L.E. Lyburg, N. Mathiowetz, and S. Sudmar, editors, *Measurement Error in Surveys.*, pages 187–209. Wiley, New York, 1991.

B. Eichhorn and L.S. Hayre. Scrambled randomized response methods for obtaining sensitive quantitative data. *Journal of Statistical Planning and Inference*, 7:307–316, 1983.

S. Elmusharaf, N. Elhadi, and L. Almroth. Reliability of self reported form of female genital mutilation and WHO classification: cross sectional study. *BMJ.*, 333(7559):124, 2006. doi: 10.1136/bmj.38873.649074.55.

S.A. Eriksson. A new model for randomized response. *International Statistical Review*, 41:40–43, 1973.

E. L. Escobar and Y. G. Berger. A new replicate variance estimator for unequal probability sampling without replacement. *Canadian Journal of Statistics*, 41:508–524, 2013.

J.A. Fox. *Randomized Response and Related Methods. Surveying Sensitive Data.* Sage Publication, Inc., California, 2016. ISBN 9781483381039.

J.A. Fox and P.E. Tracy. *Randomized Response: A Method for Sensitive Survey.* Sage Publication, Inc., 1986.

J.P. Fox and R.R. Meijer. Using item response theory to obtain individual information from randomized response data: An application using cheating data. *Applied Psychological Measurement*, 32:595–610, 2008.

J.P. Fox, M. Avetisyan, and J. van der Palen. Mixture randomized item-response modeling: A smoking behavior validation study. *Statistics in Medicine*, 32:4821–4837, 2013.

J.P. Fox, R.K. Entink, and M. Avetisyan. Compensatory and non-compensatory multidimensional randomized item response models. *British Journal of Mathematical and Statistical Psychology*, 67:133–152, 2014.

A.G. Franke, C. Bagusat, P. Dietz, I. Hoffmann, P. Simon, R. Ulrich, and K. Lieb. Use of illicit and prescription drugs for cognitive or mood enhancement among surgeons. *BMC Medicine*, 11:102, 2013.

G.Z. Geng, G. Gao, Y.H. Ruan, M.R. Yu, and Y.H. Zhou. 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*, 129:523–529, 2016.

R. Gjestvang and S. Singh. A new randomized response model. *Journal of the Royal Statistical Society B*, 68:523–530, 2006.

A.N. Glynn. What can we learn with statistical truth serum? Design and analysis of the list experiment. *Public Opinion Quarterly*, 77:159–172, 2013.

M.S. Goodstadt and V. Gruson. The randomized response technique: A test on drug use. *Journal of the American Statistical Association*, 70:814–818, 1975.

B.G. Greenberg, A.L. Abul-Ela, W.R. Simmons, and D.G. Horvitz. The unrelated question RR model: Theoretical framework. *Journal of the American Statistical Association*, 64:520–539, 1969.

B.G. Greenberg, R.R. Kuebler, J.R. Abernathy, and D.G. Horvitz. Application of the randomized response technique in obtaining quantitative data. *Journal of the American Statistical Association*, 66:243–250, 1971.

H. Groenitz. Improvements and extensions of the item count technique. *Electronic Journal of Statistics*, 8:2321–2351, 2014.

A. Gunarathne, S. Kubota, P. Kumarawadu, K. Karunagoda, and H. Kon. Is hiding foot and mouth disease sensitive behavior for farmers? A survey study in Sri Lanka. *Asian-Australasian Journal of Animal Sciences*, 29:280–287, 2016.

S. Gupta, J. Shabbir, and S. Sehra. Mean and sensitivity estimation in optional randomized response models. *Journal of Statistical Planning and Inference*, 140:2870–2874, 2010.

A. Gyepesi, R. Urbán, J. Farkas, L. Kraus, D. Piontek, B. Paksi, G. Horváth, A. Magi, A. Eisinger, J. Pilling, G. Kökönyei, B. Kun, and Z. Demetrovics. Psychometric properties of the Cannabis Abuse Screening Test in Hungarian samples of adolescents and young adults. *European Addiction Research*, 20:119–128, 2014. doi: 10.1159/000353238.

L. Harrison. The validity of self-reported drug use in survey research: an overview and critique of research methods. *NIDA Research Monograph*, 167:17–36, 1997.

M. Harrison, J. Baker, M. Twinamatsiko, and E.J. Milner-Gulland. Profiling unauthorized natural resource users for better targeting of conservation interventions. *Conservation Biology*, 29:1636–1646, 2015.

D.W. Heck and M. Moshagen. Package RReg Correlation and Regression Analyses for Randomized Response Data, 2015. URL `http://psycho3.uni-mannheim.de/Home/Research/Software/RReg/`.

L. Hides, S. Dawe, R.M. Young, and D.J. Kavanagh. The reliability and validity of the Severity of Dependence Scale for detecting cannabis dependence in psychosis. *Addiction*, 102:35–40, 2007. doi: 10.1111/j.1360-0443.2006.01669.x.

A. Hoffmann and J. Musch. Assessing the validity of two indirect questioning techniques: A stochastic lie detector versus the crosswise model. *Behavior Research Methods*, 48:1032–1046, 2016.

A. Hoffmann, B. Diedenhofen, B. Verschuere, and J. Musch. A strong validation of the crosswise model using experimentally-induced cheating behavior. *Experimental Psychology*, 62:403–414, 2016.

A.L. Holbrook and J.A. Krosnick. Measuring voter turnout by using the randomized response technique: Evidence calling into question the method's validity. *Public Opinion Quarterly*, 74:328–343, 2010.

D.G. Horvitz and D.J. Thompson. A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47:663–685, 1952.

D.G. Horvitz, B.V. Shah, and W.R. Simmons. The unrelated question RR model. *Proc. ASA Social Statist. Sec.*, pages 65–72, 1967.

J. Houston and A. Tran. A survey of tax evasion using the randomized response technique. *Advances in Taxation*, 13:69–94, 2001.

Z. Hussain, N. Shabbir, and J. Shabbir. An alternative item sum technique for improved estimators of population mean in sensitive surveys. *Hacettepe Journal of Mathematics and Statistics*, 46:907–934, 2017.

F. Ibrahim. An alternative modified item count technique in sampling survey. *International Journal of Statistics and Applications*, 6:177–187, 2016.

K. Imai. Multivariate regression analysis for the item count technique. *Journal of the American Statistical Association*, 106(494):407–416, 2011.

K. Imai, B. Park, and K.F. Greene. Using the predicted responses from list experiments as explanatory variables in regression models. *Political Analysis*, 23:180–196, 2015.

R.A. James, T. Nepusz, D.P. Naughton, and A. Petróczi. A potential inflating effect in estimation models: Cautionary evidence from comparing performance enhancing drug and herbal hormonal supplement use estimates. *Psychology of Sports and Exercise*, 14:84–96, 2013.

B. Jann, J. Jerke, and I. Krumpal. Asking sensitive questions using the crosswise model: An experimental survey measuring plagiarism. *Public Opinion Quarterly*, 76:32–49, 2012.

M.G. De Jong, R. Pieters, and S. Stremersch. 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, 2012.

Y. Kazemzadeh, M. Shokoohi, M.R. Baneshi, and A.A. Haghdoost. The frequency of high-risk behaviors among iranian college students using indirect methods: network scale-up and crosswise model. *International Journal of High Risk Behaviors & Addictions*, 2016. doi: 10.5812/ijhrba.25130.

J. Kerkvliet. Estimating a logit model with randomized data: The case of cocaine use. *Australian Journal of Statistics*, 36:9–20, 1994.

A. Khosravi, S.A. Mousavi, R. Chaman, F. Khosravi, M. Amiri, and M. Shamsipour. Crosswise model to assess sensitive issues: A study on prevalence of drug abuse among university students of Iran. *International Journal of High Risk Behaviors & Addictions*, 2015. doi: 10.5812/ijhrba.24388v2.

J.M. Kim and M.E. Elam. A stratified unrelated randomized response model. *Statistical Papers*, 48:215–233, 2007.

E. Klouman and K.I. Klepp R. Manongi. Self-reported and observed female genital cutting in rural Tanzania: associated demographic factors, HIV and sexually transmitted infections. *Trop Med Int Health.*, 10(1):105–115, 2005. doi: 10.1111/j.1365-5313156.2004.01350.x.

M. Korndörfer, I. Krumpal, and S.C. Schmukle. Measuring and explain tax evasion: Improving self-reports using the crosswise model. *Journal of Economic Psychology*, 45:18–32, 2014.

A. Koski and J. Heymann. Thirty-year trends in the prevalence and severity of female genital mutilation: a comparison of 22 countries. *BMJ Global Health*, 2(4), 2017. doi: 52110.1136/bmjgh-2017-000467.

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

I. Krumpal. Estimating the prevalence of xenophobia and anti-semitism in Germany: A comparison of the randomized response technique and direct questioning. *Social Science Research*, 41:1387–1403, 2012.

A.Y.C. Kuk. Asking sensitive questions indirectly. *Biometrika*, 77:436–438, 1990.

R.J. Lamarine. Marijuana: modern medical chimaera. *Journal of Drug Education*, 42:111, 2012. doi: 10.2190/DE.42.1.a.

D. Lara, J. Strickler, C.D. Olavarrieta, S.G. García, and C. Ellertson. Measuring induced abortion in Mexico: A comparison of four methodologies. *Sociological Methods & Research*, 32:529–558, 2004.

D. Lara, S.G. García, C. Ellertson, C. Camlin, and J. Suaréz. The measure of induced abortion in Mexico using random response technique. *Sociological Methods & Research*, 35:279–301, 2006.

C.S. Lee, S.A. Sedory, and S. Singh. Estimating at least seven measures of qualitative variables from a single sample using randomized response technique. *Statistics and Probability Letters*, 83:399–409, 2013.

S. Legleye, D. Piontek, L. Kraus, E. Morand, and B.A. Falissard. Validation of the Cannabis Abuse Screening Test (CAST) using a latent class analysis of the DSM-IV among adolescents. *International Journal of Methods in Psychiatric Research*, 22:16–26, 2013. doi: 10.1002/mpr.1378.

G.J.L.M. Lensvelt-Mulders, J.J. Hox, P.G.M. van der Heijden, and C.J.M. Mass. Meta-analysis of randomized response research: Thirty-five years of validation. *Sociological Methods & Research*, 33:319–348, 2005.

G.J.L.M. Lensvelt-Mulders, P.G.M. van der Heijden, O. Laudy, and van G. Gils. A validation of a computer-assisted randomized response survey to estimate the prevalence of fraud in social security. *Journal of the Royal Statistical Society-A*, 169:305–318, 2006.

Y. Liu and G.L. Tian. Multi-category parallel models in the design of surveys with sensitive questions. *Statistics and Its Interface*, 6:137–142, 2013.

P.K. Mahajan and R. Singh. Optimum stratification for scrambled response in pps sampling. *Metron*, LXIII:103–114, 2005.

N.S. Mangat. Two stage randomized response sampling procedure using unrelated question. *Journal of the Indian Society of Agricultural Statistics*, 44:82–87, 1992.

N.S. Mangat and R. Singh. An alternative randomized response procedure. *Biometrika*, 77: 439–442, 1990.

N.S. Mangat, R. Singh, and S. Singh. An improved unrelated question randomized response strategy. *Calcutta Statistical Association Bulletin*, 42:277–281, 1992.

J.D. Miller. *A New Survey Technique for Studying Deviant Behavior.* PhD thesis, The George Washington University, 1984.

J.D. Miller. The Nominative Technique: A New Method of Estimating Heroin Prevalence. *NIDA Research Monograph*, 57:104–124, 1985.

H. Moseson, M. Massaquoi, C. Dehlendorf, L. Bawo, B. Dahn, Y. Zolia, E. Vittinghoff, R.A. Hiatt, and C. Gerdts. Reducing under-reporting of stigmatized health events using the list experiment: Results from a randomized, population-based study of abortion in Liberia. *International Journal of Epidemiology*, 44:1951–1958, 2015.

M. Moshagen, B.E. Hilbig, E. Erdfelder, and A. Moritz. An experimental validation method for questioning techniques that assess sensitive issues. *Experimental Psychology*, 61:48–54, 2014.

M.R. Nakhaee, F. Pakravan, and N. Nakhaee. Prevalence of use of anabolic steroids by bodybuilders using three methods in a city of Iran. *Addiction & Health*, 5:77–82, 2013.

T. Nepusz, A. Petróczi, D.P. Naughton, T. Epton, and P. Norman. Estimating the prevalence of socially sensitive behaviors: attributing guilty and innocent noncompliance with the single sample count method. *Psychological Methods*, 19:334–355, 2014.

A. Nuno, N. Bunnefeld, L.C. Naiman, and E.J. Milner-Gulland. A novel approach to assessing the prevalence and drivers of illegal bushmeat hunting in the serengeti. *Conservation Biology*, 2:1355–1365, 2013.

O. Odumade and S. Singh. An alternative to the Bar-Lev, Bobovitch, and Boukai randomized response model. *Sociological Methods & Research*, 20:1–16, 2010.

E. Oliveras and G. Letamo. Examples of methods to address underreporting of induced abortion: preceding birth technique and randomized response technique. In S. Singh, L. Remez, and A. Tartaglione, editors, *Methodologies for Estimating Abortion Incidence and Abortion-Related Morbidity: A Review.*, chapter 5, pages 63–69. Guttmacher Institute - International Union for the Scientific Study of the Population, New York, Paris, 2010.

M. Ostapczuk, J. Musch, and M. Mashagen. A randomized-response investigation of the education effect in attitudes towards foreigners. *European Journal of Social Psychology*, 39:920–931, 2009.

S. Pal. Unbiasedly estimating of total of a stigmatizing variable from a complex survey on permitting options for direct or randomized responses. *Statistical Papers*, 49:157–164, 2008.

P.F. Perri. Modified randomized devices for Simmons' model. *Model Assisted Statistics and Applications*, 3:233–239, 2008.

P.F. Perri and P.G.M. van der Heijden. A property of the CHAID partitioning method for dichotomous randomized response data and categorical predictors. *Journal of Classification*, 29:76–90, 2012.

P.F. Perri, E. Pelle, and M. Stranges. Estimating induced abortion and foreign irregular presence using the randomized response crossed model. *Social Indicators Research*, 129:601–618, 2016.

A. Petróczi, T. Nepusz, P. Cross, H. Taft, S. Shah, N. Deshmukh, J. Schaffer, M. Shane, C. Adesanwo, J. Barker, and D.P. Naughton. New non-randomised model to assess the prevalence of discriminating behaviour: A pilot study on mephedrone. *Substance Abuse Treatment, Prevention, and Policy*, 6:20, 2011. doi: 10.1186/1747-597X-6-20.

D. Piontek, L. Kraus, and D. Klempova. Short scales to assess cannabis-related problems: a review of psychometric properties. *Substance Abuse Treatment, Prevention, and Policy*, 2: 25–35, 2008. doi: 10.1186/1747-597X-3-25.

K.H. Pollock and Y. Bek. A comparison of three randomized response models for quantitative data. *Journal of the American Statistical Association*, 71:884–886, 1976.

A. Quatember. An extension of the standardized randomized response technique to a multi-stage setup. *Statistical Methods & Applications*, 21:475–484, 2012.

M.H. Quenouille. Problems in plane sampling. *The Annals of Mathematical Statistics*, 20: 355–375, 1949.

D. Raghavarao and W.F. Federer. Block total response as an alternative to the randomized response method in survey. *Journal of the Royal Statistical Society-B*, 41:40–45, 1979.

D.E. Ramo, H. Liu, and J.J. Prochaska. Reliability and validity of young adults' anonymous online reports of marijuana use and thoughts about use. *Psychology of Addictive Behaviors*, 26:801–811, 2012. doi: 10.1037/a0026201.

T.N. Randrianantoandro, H. Kono, and S. Kubota. Knowledge and behavior in an animal disease outbreak: Evidence from the item count technique in a case of African swine fever in Madagascar. *Preventive Veterinary Medicine*, 118:483–487, 2015.

A. Saha. A simple randomized response technique in complex surveys. *Metron*, LXV:59–66, 2007a.

A. Saha. Optional randomized response in stratified unequal probability sampling: a simulation based numerical study with Kuk's method. *Test*, 16:346–354, 2007b.

D.J. Schill and P.A. Kline. Use of random response to estimate angler noncompliance with fishing regulations. *North American Journal of Fisheries Management*, 15:721–731, 1995.

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

P. Shaw. Estimating a finite population mean of a sensitive quantitative variable from a single probability sample by the Item Count Technique. *Model Assisted Statistics and Applications*, 10:411–419, 2015.

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

S. Singh. *Advanced Sampling Theory with Applications. In How Michael 'selected' Amy.* Springer Science & Business Media, Dordrecht, 2003. ISBN 9789400707894.

S. Singh and A.H. Joarder. Unknown repeated trials in randomized response sampling. *Journal of the Indian Statistical Association*, 30:109–122, 1997.

S. Singh and T.A. Tarray. An improvement over Kim and Elam stratified unrelated question randomized response Mode using Neyman allocation. *Sankhya B*, 77:91–107, 2015.

V. Soberanis-Cruz, G. Ramírez Valverde, S. Pérez Elizalde, and F. González Cossio. Randomized response sampling in finite populations: A unifying approach. *Agrociencia*, 42:537–549, 2008.

J. Solomon, S.K. Jacobson, K.D. Wald, and M. Gavin. Estimating illegal resources use at the Ugandan park with the randomized response technique. *Human Dimensions of Wildlife*, 12: 75–88, 2007.

S.K Srivastava and H.S. Jhajj. A class of estimators of the population mean in survey sampling using auxiliary information. *Biometrika*, 68:341–343, 1981.

A.J. Starosta and M. Earleywine. Assessing base rates of sexual behavior using the unmatched count technique. *Health Psychology and Behavioral Medicine*, 2:198–210, 2014.

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

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

M. Templ. CRAN Task View: Official Statistics and Survey Methodology, 2015. URL `https://cran.r-project.org/web/views/OfficialStatistics.html`.

G.L. Tian and Z. Geng J.W. Yu, M.L. Tang. A new nonrandomized moel for analyzing sensitive questions with binary outcomes. *Statistics in Medicine*, 26:4238–4252, 2007.

G.L. Tian and M.L. Tang. *Incomplete Categorical Data Design: Non-Randomized Response Techniques for Sensitive Questions in Surveys.* Chapman & Hall/CRC (Statistics in the Social and Behavioral Sciences), 2014.

R. Tourangeau and T. Yan. Sensitive questions in surveys. *Psychological Bulletin*, 133:859–863, 2007. doi: 10.1037/0033-2909.133.5.859.

M. Trappmann, I. Krumpal, A. Kirchner, and B. Jann. Item sum: A new technique for asking quantitative sensitive questions. *Journal of Survey Statistics and Methodology*, 2:58–77, 2014.

T. Tsuchiya. Domain estimators for ithe item count technique. *Survey Methodology*, 31:41–51, 2005.

R. Ulrich, H. Schörter, H. Striegel, and P. Simon. Asking sensitive questions: a statistical power analysis of randomized response models. *Psychological Methods*, 17:623–641, 2012.

P.G.M. van der Heijden, G. van Gils, J. Bouts, and J.J. Hox. A comparison of randomized response, computer-assisted self-interview, and face-to-face direct questioning. *Sociological Methods & Research*, 28:505–537, 2000.

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

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

K.M. Wolter. *Introduction to variance estimation.* Springer, 2007.

Q. Wu and M.L. Tang. Non-randomized response model for sensitive survey with noncompliance. *Statistical Methods in Medical Research*, 25:2827–2839, 2016.

Part II

Appendices

Appendix A1

An improved class of estimators in RR surveys

Rueda, María del Mar; Cobo, Beatriz; Arcos, Antonio (2016)
An improved class of estimators in RR surveys.
Mathematical Methods in the Applied Sciences, vol. 41, number 6, pp. 2307 - 2318
DOI: 10.1002/mma.4256

MATHEMATICS, APPLIED			
JCR Year	Impact factor	Rank	Quartile
2016	1.017	108/255	Q2

Abstract

This work proposes a general class of estimators for the population total of a sensitive variable using auxiliary information. Under a general randomized response model, the optimal estimator in this class is derived. Design-based properties of proposed estimators are obtained. A simulation study reflects the potential gains from the use of the proposed estimators instead of the customary estimators.

1 Introduction

Linear estimation parameters in a population is performed through surveys. An example is the number of voters to a particular party in an election poll.

In many surveys, it becomes necessary to probe into areas considered sensitive and potentially embarrassing. The validity of self-reports of sensitive attitudes and behaviors suffers from the tendency of individuals to distort their responses towards their perception of what is socially acceptable. As a consequence, studies self-report measures consistently underestimate the prevalence of undesirable attitudes or behaviors and overestimate the prevalence of desirable attitudes or behaviors. In an attempt to reduce this bias, Warner developed the randomized response technique (RRT) [39]. His idea spawned a vast volume of literature, (e.g., [4], [8], [15], [9], [32]).

The authors of [21] and [20] have extended Warner's model to the case where the responses to the sensitive question are quantitative rather than a simple yes or no. The respondent selects, by means of a randomization device, one of the two questions: one being the sensitive question, the other being unrelated. There are several difficulties that arise when using this unrelated question method [35]. These difficulties are no longer present in the scrambled randomized response method introduced by Eichhorn and Hayre [16]. In Eichhorn and Hayre model, each respondent scrambles their response y by multiplying it by a random variable S and then reveals only the scrambled result $z = yS$ to the interviewer; thus, the scrambled randomized response model maintains the privacy of the respondents. Saha [30] discussed the use of scrambled responses based on both multiplicative and additive model, which involve the respondent adding and multiplying the answer to the sensitive question by two random number. Bar-Lev, Bobovitch, and Boukai [7] proposed a method that generalizes the Eichhorn and Hayre model, which introduces a design parameter controlled by the researcher and used for randomizing the responses. Other important RR models are proposed by the authors of the literature [15], [17], and [19].

Most research into RRT techniques deals exclusively with the interest variable and does not make explicit use of auxiliary variables in the construction of estimators. Examples of these auxiliary variables in election polls could be sex, age, educational level, or taxes. Diana and

Perri [13] pointed out that in sampling practice, direct techniques for collecting information about non-sensitive characteristics make massive use of auxiliary variables to improve sampling design and to achieve higher precision in population parameter estimates. Nevertheless, very few procedures have been suggested to improve randomization technique performance using the supplementary information. Regression estimators for scrambled variables are defined in [33], [14], [27] and [36]. Tracy and Singh [38] introduced the calibration of scrambled responses and find the conditional bias and variance of the proposed estimator. Singh an kim [34] proposed an empirical log-likelihood estimator for estimating the population mean of a sensitive variable in the presence of an auxiliary variable. Diana and Perri [15] discussed the use of auxiliary information to estimate the population mean of a sensitive variable when data are perturbed by means of three scrambled response devices, namely, the additive, the multiplicative, and the mixed model. Koyuncu et al. [24] proposed exponential-type estimators using one and two auxiliary variables.

From a mathematical point of view, a process of seeking an optimal estimator in a class of estimators for the total of sensitive characteristic arises, under a general model for the scrambling response and in presence of additional information.

In this paper, we suggest a class of estimators for a finite population total when the population totals of the auxiliary variables are known. In Section 2, we introduce the problem of estimating the total of the target population when there are scrambled variables. In section 3, we propose a general class of estimators for the population total. Proposed estimators are based upon auxiliary variables and assume that observations on the variable of interest are obtained using a RRT. We present particular estimators of the proposed class of estimators, and we derive the asymptotic properties of these estimators. Using a real population, the proposed estimators are evaluated empirically in Section 4, and they are compared to alternative estimators. Finally, some conclusions are drawn.

2 Estimation of the population total in RRT

Consider a finite population U, consisting of N different individuals. Let y_i, $i = 1, ..., N$ be the value of the sensitive aspect under study for the ith population element. Our aim is to estimate the finite population total $Y = \sum_{i=1}^{N} y_i$ of the variable of interest y or the population mean $\bar{Y} = 1/N \sum_{i=1}^{N} y_i$.

Assume that a sample s of individuals is chosen according to a non-informative sampling design p with first-order inclusion probabilities $\pi_i = \sum_{s \ni i} p(s), i \in U$ and second-order inclusion probabilities $\pi_{ij} = \sum_{s \ni i,j} p(s), i, j \in U$. Let us assume that the operators E_d and V_d denote expectation and variance with respect to the sampling design [6] and that the first-order and second-order inclusion probabilities are positive.

If the value y_i is known exactly by observing the ith individual, then the standard Horvitz-Thompson (HT) estimator of the total Y can be used:

$$\hat{Y} = \sum_{i \in s} \frac{y_i}{\pi_i}$$

with variance

$$V_{HT}(\hat{Y}) = \frac{1}{2} \sum_{i \neq} \sum_{j \in U} (\pi_i \pi_j - \pi_{ij}) \left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j} \right)^2,$$

which can be unbiasedly estimated as

$$\hat{V}_{HT}(\hat{Y}) = \frac{1}{2} \sum_{i \neq} \sum_{j \in s} \frac{\pi_i \pi_j - \pi_{ij}}{\pi_{ij}} \left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j} \right)^2.$$

Let y be the variable under study, a sensitive variable that can not be observed directly. In order to consider a wide variety of RR procedures, we consider the unified approach given by [4]. The interviews of individuals in the sample s are conducted in accordance with a RR model. Because y_i is not directly available from the respondent, y_i is estimated through the randomized response obtained from the ith respondent. Suppose that the ith respondent has to conduct a RR trial independently and z_i is the randomized response (or scrambled response) for the trial. For each $i \in s$, the RR induces a revised randomized response r_i such as $E_R(r_i) = y_i$ and $V_R(r_i) = \phi_i$ where the operators E_R and V_R denote expectation and variance with respect to randomization procedure RR. As usual in the design-based approach to RR techniques, it is assumed that the sampling design and the randomization stage are independent of each other (e.g., [6]) and that the randomization stage is performed on each selected individual independently. In this general setup, the HT-type estimator for the population total of the sensitive characteristic y given by

$$\hat{Y}(r) = \sum_{i \in s} \frac{r_i}{\pi_i}$$

is an unbiased estimator because:

$$E(\hat{Y}(r)) = E_d \left(E_R \left(\sum_{i \in s} \frac{r_i}{\pi_i} \right) \right) = Y$$

The variance of $\hat{Y}(r)$ can be obtained from

$$V(\hat{Y}(r)) = V_d(E_R(\hat{Y}(r)) + V_R(E_d(\hat{Y}(r))) =$$

$$\left[\frac{1}{2}\sum_{i\neq}\sum_{j\in U}(\pi_i\pi_j-\pi_{ij})\left(\frac{y_i}{\pi_i}-\frac{y_j}{\pi_j}\right)^2+\sum_{i\in U}\frac{\phi_i}{\pi_i}\right]=V_{HT}+\sum_{i\in U}\frac{\phi_i}{\pi_i}$$

being V_{HT} the variance of the HT estimator based on the y_i's. An estimator of $V(\hat{Y}(r))$ is given by

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

This estimator is an unbiased estimator of $V(\hat{Y}(r))$ if $\hat{\phi}_i$ is an RR-unbiased for ϕ_i.

3 Estimators in the presence of auxiliary information

3.1 A general class of estimators for the total

The proposed estimators consider k auxiliary variables $x_1,\ldots,x_k$, for which the population totals $X_1,\ldots,X_k$, are known. We assume that the values of auxiliary variables can be observed directly in the sample. Our goal is to estimate the population parameter Y by using observations of the variables $r, x_1,\ldots,x_k$ in the sample s, and the known population values $X_1,\ldots,X_k$ associated with the auxiliary variables. We note by $\widehat{X}_h$ the HT estimator of the total X_h $(h=1,...,k)$.

Motivated by [37], we suggest the class of estimators of Y

$$\widehat{Y}_g^{(r)}=\{G(\hat{Y}(r),u_1,...,u_k)\},\tag{1}$$

where $G(\cdot)$ is a function of $u_h=\widehat{X}_h/X_h$, continuous in a closed convex sub-space, $P\subset\mathbb{R}^{k+1}$, containing the point $(Y,1,...,1)=(Y,\mathbf{1})$, and such that

(A1) $G(Y,\mathbf{1})=Y$

(A2) $G_0'(Y,\mathbf{1})=1$ where $G_0'(Y,\mathbf{1})$ denoting the first partial derivative of $G(\cdot)$ with respect to $\hat{Y}(r)$.

(A3) The first and second-order partial derivatives of $G(\cdot)$ exist and are also continuous and bounded in P.

Now we studies some asymptotic design-based properties of $\widehat{Y}_g^{(r)}$. We consider the asymptotic framework of [22], in which the finite population U and the sampling design p are embedded into a sequence of such populations and designs indexed by N, $\{U_N,p_N\}$, with $N\to\infty$. We assume that $N_N \to \infty$ and $n_N \to \infty$, $n_N/N_N \to f \in (0,1)$, as $N \to \infty$. Subscript N may be dropped for ease of notation, although all limiting processes are understood under the aforementioned conditions. Stochastic order $O_p(\cdot)$ is with respect to the aforementioned sequence of designs.

Theorem 1

Any estimator into the class (1) is asymptotically unbiased for Y.

Proof.

By expanding G about the point $(Y, \mathbf{1})$ in a first-order Taylor series, it is found that

$$\widehat{Y}_g^{(r)} = G(Y, \mathbf{1}) + (\hat{Y}(r) - Y)G'_0(Y, \mathbf{1}) + \sum_{h=1}^{k} G'_h|_{(Y,\mathbf{1})}(u_h - 1) + O_p(n^{-1}), \tag{2}$$

where G'_h denotes the first-order partial derivative with respect to u_h.

By taking expectations on both sides in (2) we obtain

$$E[\widehat{Y}_g^{(r)}] \simeq Y + E[\hat{Y}(r)] - Y + \sum_{h=1}^{k} E(\widehat{X}_h - X_h)\frac{G'_h|_{(Y,\mathbf{1})}}{X_h}.$$

We have $E[\hat{Y}(r)] = E_d E_R(\hat{Y}(r)) = Y$, $E[\widehat{X}_h] = E_d(\widehat{X}_h) = X_h$. Thus, $E[\widehat{Y}_g^{(r)}] = Y + O(n^{-1})$ so the bias is of order n^{-1}. □

Theorem 2

An approximation of the bias of the proposed class of estimators is given by:

$$B[\widehat{Y}_g^{(r)}] = \sum_{h<t} \frac{Cov(\widehat{X}_h, \widehat{X}_t)}{X_h X_t} G''_{ht}|_{(Y,\mathbf{1})} + \frac{1}{2}\sum_{h=1}^{k} \frac{V(\widehat{X}_h)}{X_h^2} G''_{hh}|_{(Y,\mathbf{1})}$$

$$+\frac{1}{2}\frac{V(\hat{Y}(r))}{Y} G''_{00}|_{(Y,\mathbf{1})} + \frac{1}{2}\sum_{h=1}^{k} \frac{Cov(\widehat{X}_h, \hat{Y}(r))}{X_h} G''_{0h}|_{(Y,\mathbf{1})},$$

where G''_{ht} denote the second-order partial derivative with respect to u_h and u_t, G''_{0h} is the second order partial derivative with respect to Y and u_h, and G''_{00} is second-order partial derivative respect to Y.

Proof.

By expanding G about the point $(Y, \mathbf{1})$ in a second-order Taylor series,

$$\widehat{Y}_g^{(r)} = Y + (\hat{Y}(r) - Y) + \sum_{h=1}^{k} G'_h|_{(Y,\mathbf{1})}(u_h - 1)$$

$$+\sum_{h<t}(u_h-1)(u_t-1)G''_{ht}|_{(Y,\mathbf{1})} + \frac{1}{2}\sum_{h=1}^{k}(u_h-1)^2 G''_{hh}|_{(Y,\mathbf{1})}$$

$$+\frac{1}{2}\sum_{h=1}^{k}(u_h-1)(\hat{Y}(r)-Y)G''_{0h}|_{(Y,\mathbf{1})} + \frac{1}{2}(\hat{Y}(r)-Y)^2 G''_{00}|_{(Y,\mathbf{1})} + O_p(n^{-2}).$$

Taking expectations in the aforementioned second degree approximation, we obtain the approximate bias (of order $O(n^{-2})$) of the proposed estimator. □

Note that, under a general sampling design, the variances and covariances in Theorem 2, can be computed as:

$$V(\widehat{X}_h) = \frac{1}{2}\sum_{i\neq}\sum_{j\in U}(\pi_i\pi_j - \pi_{ij})\left(\frac{x_{hi}}{\pi_i} - \frac{x_{hj}}{\pi_j}\right)^2$$

and

$$Cov(\widehat{X}_h, \widehat{X}_t) = \frac{1}{2}\sum_{i\neq}\sum_{j\in U}(\pi_i\pi_j - \pi_{ij})\left(\frac{x_{hi}}{\pi_i} - \frac{x_{tj}}{\pi_j}\right)^2,$$

$V(\hat{Y}(r))$ is given in section 2. Then, we only need to obtain the $Cov(\widehat{X}_h, \widehat{Y}(r))$. For this, using the covariance theorem, we have:

$$Cov(\widehat{X}_h, \hat{Y}(r)) = E_d(cov_R(\widehat{X}_h, \hat{Y}(r)) + cov_d(E_R(\widehat{X}_h), E_R(\hat{Y}(r))) =$$

$$E_d(0) + cov_d\left(\widehat{X}_h, \sum_{i\in s}\frac{y_i}{\pi_i}\right) = \frac{1}{2}\sum_{i\neq}\sum_{j\in U}(\pi_i\pi_j - \pi_{ij})\left(\frac{x_{hi}}{\pi_i} - \frac{y_j}{\pi_j}\right)^2.$$

Note 1. In deriving the expected value of $\widehat{Y}_g^{(r)}$, we assumed that the contribution of terms involving powers higher that the second is negligible. One can retain the terms up to and including degree third and four and proceed to obtain a better approximation to the expected value of $\widehat{Y}_g^{(r)}$. Unless n is small, the contribution of the third and fourth degree terms to the relative bias can be considered to be negligible. For appreciable large n, say 30 or larger, the approximation to $O(n^{-1})$ may be considered as adequate [3].

Theorem 3

The asymptotic variance of any estimator into the class (1) verifies:

$$AV(\widehat{Y}_g^{(r)}) \geq V(\hat{Y}(r)) - \sigma' \ \Sigma^{-1}\sigma,$$

where $\Sigma = (a_{ht})_{(k\times k)}$ with $a_{hh} = V(\widehat{X}_h)$, $a_{ht} = Cov(\widehat{X}_h, \widehat{X}_t)$ and $\sigma = (Cov(\widehat{X}_1, \hat{Y}(r)), \ldots, Cov(\widehat{X}_k, \hat{Y}(r)))'$.

Proof.

By squaring both sides in expression (2), taking expectations and neglecting higher-order terms we obtain the following approximation:

$$V(\widehat{Y}_g^{(r)}) = E[\widehat{Y}_g^{(r)} - Y]^2 \simeq E\left[\hat{Y}(r) + \sum_{h=1}^{k} G'_h|_{(Y,\mathbf{1})}(u_h - 1) - Y\right]^2. \quad (3)$$

On differentiating (3) and equating to zero, we obtain the optimum values of the parameters as

$$(G'_1|_{(Y,\mathbf{1})}, \ldots, G'_k|_{(Y,\mathbf{1})})' = D^{-1}b,$$

where $D = (d_{ht})$, $b = (b_1, ..., b_k)'$ and

$$d_{ht} = \frac{Y^2 Cov(\widehat{X}_h, \widehat{X}_t)}{X_h X_t} \quad ; \quad b_h = \frac{Y Cov(\widehat{X}_h, \hat{Y}(r))}{X_h}.$$

On substituting the optimum values into (3), we obtain the minimum first-order approximation for the variance

$$AV_{\min}(\widehat{Y}_g^{(r)}) = V(\hat{Y}(r)) - \sigma'\Sigma^{-1}\sigma = V(\hat{Y}(r))(1 - R^2_{\hat{Y}(r),\widehat{X}_1,\ldots,\widehat{X}_k}),$$

where $R^2_{\hat{Y}(r),\widehat{X}_1,\ldots,\widehat{X}_k}$ is the multiple correlation coefficient. This proofs the Theorem 3. □

Note 2. The aforementioned expression emphasizes the role of the auxiliary variables in improving the accuracy of the estimates. $(1 - R^2_{\hat{Y}(r),\widehat{X}_1,\ldots,\widehat{X}_k})$ denotes the reduction in the variance due to the use of auxiliary variables. We observe that the multiple correlation coefficient increases with the number of secondary variables and with the number of auxiliary parameters; hence, the variance of proposed estimators is a monotone decreasing function of the number of secondary variables.

Note 3. In practice, the value of $R^2_{\hat{Y}(r),\widehat{X}_1,\ldots,\widehat{X}_k}$ is unknown, and this fact is even more complicated in this case as y, being sensitive, makes difficult making some guess on the value of the $AV_{\min}$. If we consider the generalized randomized response procedure given in [2] the revised values are given by $r_i = \frac{z_i - a}{b}$ being a and b constants; thus an idea of the multiple correlation coefficient $R^2_{\hat{Y}(r),\widehat{X}_1,\ldots,\widehat{X}_k}$ can be obtained from the correspondent correlation coefficient using the scrambled responses z_i.

The proposed class of estimator can be used to obtain an optimal difference-type estimator using the idea proposed in [25] and [26]. Specifically, let us now consider a choice within the class G of the type

$$G(\hat{Y}(r), u_1, \ldots, u_k) = \hat{Y}(r) + \sum_{h=1}^{k} d_h(u_h - 1)X_h,$$

which yields to the difference estimator

$$\widehat{Y}_{gD} = \hat{Y}(r) + \sum_{h=1}^{k} d_h(X_h - \widehat{X}_h) \tag{4}$$

The optimum d_h values are $(d_1, ..., d_k)' = \Sigma^{-1}\sigma$ and

$$V(\widehat{Y}_{gD}) = V(\hat{Y}(r)) - \sigma'\Sigma^{-1}\sigma = AV_{\min}(\widehat{Y}_g^{(r)}).$$

It is interesting to note that the lower bound of the asymptotic variance of $\widehat{Y}_g^{(r)}$ is the variance of the difference estimator $\widehat{Y}_{gD}$ with the optimum d_h values. Thus, $\widehat{Y}_{gD}$ is, asymptotically, an optimal estimator into the class in the sense that it has a lower asymptotic variance, but is not unique. Any other estimator that attains the minimum variance bound is optimum as well; thus, to the first-order of approximation, that is, up to terms $O(n^{-1})$, these estimators will be equivalent to the optimal difference estimator $\widehat{Y}_{gD}$. For d_h, with $h = 1, \ldots, k$, known, this estimator has the advantage of providing exact results for the unbiasedness and the variance of the estimator of the total.

The optimum values d_h, with $h = 1, \ldots, k$, depend on population values, which are generally unknown in practice; hence, the optimal difference estimator $\widehat{Y}_{gD}$ cannot be used in general. Population values can be estimated by using sample values or using some replication methods.

After replacing Σ and σ by their estimators $\widehat{\Sigma}$ and $\widehat{\sigma}$, we obtain the difference-type estimator

$$\widehat{Y}_{gd} = \hat{Y}(r) + \left(\Theta - \widehat{\Theta}\right)' \widehat{\Sigma}^{-1}\widehat{\sigma}, \tag{5}$$

where $\widehat{\Theta} = (\widehat{X}_1, \ldots, \widehat{X}_k)'$ and $\Theta = (X_1, \ldots, X_k)'$.

3.2 Application to simple random sampling

Some asymptotic properties under simple random sampling are derived in this section.

Theorem 4
Assuming simple random sampling, the estimators $\widehat{Y}_g^{(r)}$, $\widehat{Y}_{gD}$, and $\widehat{Y}_{gd}$ are asymptotically unbiased and normally distributed.

Proof.

The asymptotic unbiasedness of $\widehat{Y}_{gD}$ and $\widehat{Y}_g^{(r)}$ is easily derived from its linear expression (1), and using the fact that $\hat{Y}(r)$ and $\widehat{X}_h$, with $h = 1, \ldots, k$, are unbiased of their respective parameters. Similarly, because $\hat{Y}(r)$ and $\widehat{X}_h$ are asymptotically normal, the estimators $\widehat{Y}_g^{(r)}$ and $\widehat{Y}_{gD}$ are also asymptotically normal.

Results derived from [28] can be used to show that $\widehat{Y}_{gd}$ has asymptotically the same distribution than

$$\widehat{Y}_{gD} = \hat{Y}(r) + \left(\Theta - \widehat{\Theta}\right)' \Sigma^{-1}\sigma.$$

Following [28], the proposed difference estimator can be expressed as $\widehat{Y}_{gd} = T_n(\widehat{\Sigma}^{-1}\widehat{\sigma})$, whereas $\widehat{Y}_{gD}$ can be expressed as $\widehat{Y}_{gD} = T_n(\Sigma^{-1}\sigma)$, where $T_n(\widehat{\Sigma}^{-1}\widehat{\sigma})$ is a function of the data and uses the estimator $\widehat{\Sigma}^{-1}\widehat{\sigma} = (\widehat{d}_1, ..., \widehat{d}_{kl})'$, which is also a function of the data, consistently estimating the vector parameter $\Sigma^{-1}\sigma$.

Let γ be a k dimensional vector of variables. By replacing the estimator $\widehat{\Sigma}^{-1}\widehat{\sigma}$ into $T_n(\cdot)$ by γ, which is denoted by $T_n(\gamma)$, the limiting mean of $T_n(\gamma)$ can be obtained when the actual parameter value is $\Sigma^{-1}\sigma$, that is,

$$\mu(\gamma) = \lim_{n \to +\infty} E_{\Sigma^{-1}\sigma}[T_n(\gamma)] = \tilde{Y}$$

where $\tilde{Y}$ is the limiting value of Y as $N \to \infty$. Therefore

$$\frac{\partial \mu(\gamma)}{\partial \gamma}\Big|_{\gamma = \mathbf{\Sigma}^{-1}\sigma} = \left(\frac{\partial \mu(\gamma)}{\partial \gamma_1}\Big|_{\gamma = \mathbf{\Sigma}^{-1}\sigma}, \ldots, \frac{\partial \mu(\gamma)}{\partial \gamma_k}\Big|_{\gamma = \mathbf{\Sigma}^{-1}\sigma} \right) = (0, \ldots, 0).$$

Assuming this condition, [28] showed that the limiting distribution of $T_n(\widehat{\Sigma}^{-1}\widehat{\sigma})$ is the same than the distribution of $T_n(\Sigma^{-1}\sigma)$, and hence, the estimator $\widehat{Y}_{gd}$ is asymptotically unbiased and has the same asymptotic variance than $\widehat{Y}_{gD}$. This completes the proof. $\square$

3.3 Application to stratified sampling

Stratified sampling designs form an interesting and useful subclass of sampling designs. To define a stratified sampling design, the population U of size N is divided into L non-overlapping subpopulations or strata U_l of size N_l. In each stratum U_l, we select a sample s_l by using a sampling design p_l independently of one another. For example, we consider that p_l is a simple random sampling without replacement of n_l size.

Let y_{il} be the value of the study variable y, and r_{il} the value of the randomized response for the ith population element of stratum l. If the values of auxiliary variables x_h are known for each population unit x_{hil} , in a similar way as [11], we can considered a general class of estimators in each stratum. The minimum asymptotic variance $AV_{minl} = V(\hat{Y}_l(r)) - \sigma_l' {\Sigma_l}^{-1} \sigma_l$ is achieved by the difference estimator in stratum l:

$$\widehat{Y}_{gDl} = \hat{Y}_l(r) + \sum_{h=1}^{k} d_{hl}(X_{hl} - \widehat{X}_{hl}),$$

where $\hat{Y}_l(r) = N_l \sum_{i \in s_l} \frac{r_{il}}{n_l}$, $X_{hl} = \sum_{i \in U_l} x_{hil}$, $\widehat{X}_{hl} = N_l \sum_{i \in s_l} \frac{x_{hil}}{n_l}$, $d_{hl} = {\Sigma_l}^{-1} \sigma_l$ $\Sigma_l = (a_{ht})_{(k \times k)}$ with $a_{hh} = V(\widehat{X}_{hl})$, $a_{ht} = Cov(\widehat{X}_{hl}, \widehat{X}_{tl})$ and $\sigma_l = (Cov(\widehat{X}_{1l}, \hat{Y}_l(r)), \ldots, Cov(\widehat{X}_{kl}, \hat{Y}_l(r)))'$.

Thus, the separate estimator:

$$\widehat{Y}_{gD}^{st} = \sum_{l=1}^{L} \widehat{Y}_{gDl}$$

can be used for estimating the total Y. The properties of this estimator can be easily obtained by using the independence of sampling en each stratum. For example, an expression for the variance is given by:

$$V(\widehat{Y}_{gD}^{s}) = \sum_{l=1}^{L} V(\hat{Y}_l(r)) - \sigma_l' {\Sigma_l}^{-1} \sigma_l.$$

The aforementioned formulae are based on the assumption that n_l is large $\forall l$. This, however, is not always true in practice. To get over this difficulty, we suggest a general class of combined estimators given by:

$$\widehat{Y}_{gc}^{(r)} = \{G(\hat{Y}_{st}(r), u_1, ..., u_k)\}, \tag{6}$$

where $G(\cdot)$ is a function of $u_h = \widehat{X}_{hst}/X_h$, being $\widehat{X}_{hst} = \sum_l \widehat{X}_{hl}$ and $\hat{Y}_{st}(r) = \sum_l \hat{Y}_l(r)$.

In a similar way of section 3.1, the asymptotic variance of any estimator into the class verifies

$$AV(\widehat{Y}_{gc}^{(r)}) \geq V\left(\sum_{l=1}^{L} \hat{Y}_l(r)\right) - \sigma_{st}' {\Sigma_{st}}^{-1} \sigma_{st}$$

where $\Sigma_{st} = (a_{ht})_{(k\times k)}$ with $a_{hh} = V(\sum_l \widehat{X}_{hl})$, $a_{ht} = Cov(\sum_l \widehat{X}_{hl}, \sum_l \widehat{X}_{tl})$ and $\sigma_{st} = (Cov(\sum_l \widehat{X}_{1l}, \sum_l \hat{Y}_l(r)), \ldots, Cov(\sum_l \widehat{X}_{kl}, \sum_l \hat{Y}_l(r)))'$.

An asymptotically optimal estimator is given by the combined difference estimator:

$$\widehat{Y}_{gcD} = \hat{Y}_{st}(r) + \left(\Theta - \widehat{\Theta}_{st}\right)' {\Sigma_{st}}^{-1} \sigma_{st}$$

being $\widehat{\Theta}_{st} = (\widehat{X}_{1st}, \ldots, \widehat{X}_{kst})'$.

The optimum sample allocation for the separate and for the combined difference estimators under a linear cost function can be obtained minimizing in n_l the aforementioned expressions given for its variances.

3.4 Other estimators in the class

For direct questioning, many different estimators based on information of auxiliary variables have been proposed following different approaches. Some of them can be extended to our case of RR questioning. To the first-order of approximation, some of these are equivalent to the difference estimator $\widehat{Y}_{gD}$, while others are less efficient. For space saving purpose, we do not show the plethora of estimator based on the information about parameters of auxiliary variables (see [12] for more examples of estimators in the class).

Some example of estimators which attain the minimum variance bound of the class are as follows:

- The exponentation estimator (based on the idea of [1]), which is given by

$$\widehat{Y}_g^{exp} = \hat{Y}(r) \prod_{h=1}^{k} \left(\frac{X_h}{\widehat{X}_h}\right)^{\alpha_h}. \tag{7}$$

- The exponentiation-difference estimator (based on the idea of [23]) given by

$$\widehat{Y}_g^{expD} = \hat{Y}(r) \prod_{h=1}^{k} \left(\frac{X_h}{\widehat{X}_h}\right)^{\alpha_h} + \sum_{h=1}^{k} b_h (X_h - \widehat{X}_h). \tag{8}$$

Some example of estimator that are not optimum in the class.

- The exponential ratio type estimator (based on the idea of [5])

$$\widehat{Y}_g^{expR} = \hat{Y}(r) \prod_{h=1}^{k} exp \frac{X_h - \widehat{X}_h}{X_h + \widehat{X}_h} \tag{9}$$

- The generalized regression-cum-exponential estimator (based on the idea of [24])

$$\widehat{Y}_g^{regcex} = \left(w_0 \hat{Y}(r) + \sum_h w_h (X_h - \widehat{X}_h) \right) exp \left(\frac{\sum_h (X_h - \widehat{X}_h)}{\sum_h (X_h + \widehat{X}_h)} \right). \tag{10}$$

4 Simulation study

We have tested the real performance of the proposed estimators through simulation studies. The free statistical software R ([31]) was used to perform this simulation study. The library RRTCS of R ([10]) was used and, where necessary, we have developed new R-code implementing the proposed estimators.

For this purpose, we consider two studies with real and simulated populations.

The first simulation has been performed using two simulated populations used previously by [24]. The populations of size N=1000 are generated from a multivariate normal distribution (y, x_1, x_2) with the same vector of means $(5, 5, 5)$ and with different covariance matrices. The correlations in population 1 are $\rho_{yx_1} = 0.6844426$ and $\rho_{yx_2} = 0.6458839$, and the correlations in population 2 are $\rho_{yx_1} = 0.8659185$ and $\rho_{yx_2} = 0.8279276$.

We calculate the mean estimation of de variable of interest, dividing the aforementioned proposed estimators by population size.

For all populations, randomized response data were generated by using three different randomized response models. In recent years, many models of randomized response have been proposed; we have included in the simulations these three models because there are some kind of kernel of RR procedures families:

- Eichhorn and Hayre model: In Eichhorn and Hayre model ([16]) each sample respondent is to report $z_i = S * y_i$ where S is a random sample from a population with known mean θ and known variance γ^2. In this model, $E_R(z_i) = \theta, r_i = z_i/\theta, \phi = \gamma/\theta^2$

- Eriksson model: In Eriksson RR technique ([17]), it is assumed that the variable under study y can take any value in the known interval (a, b). M values $Q_1(= a), Q_2, ..., Q_M(= b)$ are chosen in the interval (a, b). The vector $Q = (Q_1, ..., Q_M)$ covers the range (a, b) and the value of M depends on the length of the interval. The respondent is supposed to report either the true value y_i with probability c or the Q_j value with probability $q_j(q_j > 0, \sum_j q_j = 1-c)$ as his/her RR response. In this case $E_R(z_i) = c * y_i + \sum_j q_j Q_j$, $E_R(z_i^2) = c * y_i^2 + \sum_j q_j Q_j^2$, $r_i = (z_i - \sum_j q_j Q_j)/c$, $\phi_i = \alpha * y_i^2 + \beta * y_i + \gamma$ and $\hat{\phi}_i = \frac{\alpha r_i^2 + \beta r_i + \gamma}{1+\alpha}$ where $\alpha = \frac{1-c}{c}$, $\beta = \frac{-2\sum_j q_j Q_j}{c}$ and $\gamma = \frac{\sum_j q_j Q_j^2 - (\sum_j q_j Q_j)^2}{c^2}$

- Bar-Lev, Bobovitch and Boukai model: This model considered in ([7]) is a special case of Eriksson model. Here, each of the sampled respondents is requested to rotate a spinner unobserved by the interviewer, and if the spinner stops in the shaded area, then the respondent is asked to disclose the true value y_i, otherwise, the respondent is asked to scramble their response y by multiplying it by a random variable S with known distribution. So in this method, $z_i = y_i$ with probability p_1, $z_i = S * y_i$ with probability $1 - p_1$ and $r_i = \frac{z_i}{p_1 + (1-p_1) * E(S)}$.

The parameters for these models are

- Eichhorn and Hayre $S \sim F(20, 20)$
- Eriksson $Q = (36656.0, 40200.5, 43698.0), c = 0.7, q_1 = q_2 = q_3 = 0.1$; and
- Bar-Lev, Bobovitch and Boukai $S \sim exp(1), p = 0.6$.

For comparison purposes, the HT estimator, $\hat{Y}(r)$; the difference estimator, $\widehat{Y}_{gd}$; the exponentiation estimator, $\widehat{Y}_g^{exp}$; and the exponential ratio type estimator, $\widehat{Y}_g^{expR}$, are computed.

In this context, simple random samples with different sizes ($n = 30, 50, 100, 200$, and 300) have been drawn. We have tested the performance of these estimators with respect to the criteria: relative bias and mean square error through simulation studies.

$$RB = \frac{(1/T) * \sum_{i=1}^{T} |\hat{\bar{Y}} - \bar{Y}|}{\bar{Y}}; MSE = (1/T) * \sum_{i=1}^{T} (\hat{\bar{Y}} - \bar{Y})^2$$

where $\hat{\bar{Y}}$ is a given estimator and $\bar{Y}$ the population mean and T is the number of replicates, in our case 1000.

Table 1 and Table 2 present the RB and MSE statistics for population 1 and population 2 for some sample sizes. The value NAV indicates the number of auxiliary variables used in the estimation process.

Table 1: Relative bias and mean square error for some estimators and some RRT devices. Population 1 ($\rho_{yx_1} = 0.6844426$ and $\rho_{yx_2} = 0.6458839$)

		Eichhorn and Hayre		Eriksson		Bar-Lev, Bobovitch and Boukai	
	NAV	Relative bias	Mean square error	Relative bias	Mean square error	Relative bias	Mean square error
n=30							
Horvitz-Thompson	0	0.122105	0.632723	0.119747	0.56197	0.129363	0.668025
Difference	1	0.106300	0.478182	0.100275	0.405386	0.117332	0.554959
Exponentiation	1	0.106098	0.478106	0.100739	0.412917	0.117523	0.555011
ExpRatio	1	0.113011	0.546582	0.110181	0.479325	0.122619	0.600449
Difference	2	0.102131	0.441093	0.095689	0.369376	0.113429	0.521364
Exponentiation	2	0.102416	0.442217	0.096668	0.383995	0.114218	0.528864
ExpRatio	2	0.106037	0.482795	0.103034	0.419880	0.116618	0.550594
n=50							
Horvitz-Thompson	0	0.095893	0.378586	0.093214	0.340943	0.102764	0.428592
Difference	1	0.082850	0.285976	0.077843	0.238776	0.092737	0.339595
Exponentiation	1	0.083071	0.287671	0.078479	0.243594	0.093055	0.342951
ExpRatio	1	0.088917	0.327752	0.085621	0.289689	0.097201	0.382186
Difference	2	0.079662	0.269144	0.074633	0.221482	0.091103	0.332474
Exponentiation	2	0.080185	0.272712	0.075299	0.226668	0.091618	0.336719
ExpRatio	2	0.083768	0.293867	0.080110	0.254902	0.093304	0.355199
n=100							
Horvitz-Thompson	0	0.066638	0.180880	0.062108	0.155636	0.072514	0.212142
Difference	1	0.055411	0.127707	0.050513	0.102501	0.067067	0.180812
Exponentiation	1	0.055564	0.127824	0.050817	0.104094	0.067127	0.181399
ExpRatio	1	0.061385	0.153358	0.056563	0.129093	0.069137	0.193490
Difference	2	0.055044	0.125252	0.049137	0.098345	0.066055	0.177752
Exponentiation	2	0.055488	0.126739	0.049578	0.100861	0.066543	0.180119
ExpRatio	2	0.057970	0.137688	0.052878	0.112785	0.066884	0.182271
n=200							
Horvitz-Thompson	0	0.044155	0.078191	0.042876	0.076481	0.055025	0.118622
Difference	1	0.037582	0.056182	0.035919	0.052029	0.050808	0.100988
Exponentiation	1	0.037572	0.056068	0.035998	0.052412	0.050901	0.101165
ExpRatio	1	0.040683	0.066405	0.039351	0.064241	0.052819	0.109047
Difference	2	0.036508	0.053085	0.034267	0.047843	0.049808	0.097363
Exponentiation	2	0.036613	0.053426	0.034395	0.048208	0.049967	0.097604
ExpRatio	2	0.038524	0.058488	0.036468	0.055366	0.051175	0.101879
n=300							
Horvitz-Thompson	0	0.035667	0.051763	0.033776	0.046499	0.046136	0.081825
Difference	1	0.031225	0.039008	0.027283	0.031188	0.043103	0.071165
Exponentiation	1	0.031199	0.038963	0.027358	0.031354	0.043133	0.071313
ExpRatio	1	0.033400	0.044901	0.030836	0.038968	0.044535	0.076105
Difference	2	0.030198	0.036942	0.026489	0.029207	0.042907	0.069589
Exponentiation	2	0.030232	0.037094	0.026584	0.029420	0.043034	0.069915
ExpRatio	2	0.031567	0.040153	0.028728	0.033884	0.043536	0.072219

Table 2: Relative bias and mean square error for some estimators and some RRT devices. Population 2 ($\rho_{yx_1} = 0.8659185$ and $\rho_{yx_2} = 0.8279276$)

		Eichhorn and Hayre		Eriksson		Bar-Lev, Bobovitch and Boukai	
	NAV	Relative bias	Mean square error	Relative bias	Mean square error	Relative bias	Mean square error
n=30							
Horvitz-Thompson	0	0.106670	0.486278	0.084411	0.285370	0.117151	0.567523
Difference	1	0.086009	0.306308	0.060335	0.149244	0.100215	0.413828
Exponentiation	1	0.085868	0.301077	0.060513	0.151140	0.100051	0.407549
ExpRatio	1	0.095796	0.391094	0.072407	0.209705	0.108193	0.482807
Difference	2	0.081512	0.270888	0.054502	0.121669	0.094466	0.367412
Exponentiation	2	0.081203	0.263149	0.054845	0.124365	0.094024	0.356536
ExpRatio	2	0.088474	0.327594	0.063047	0.160628	0.100685	0.417619
n=50							
Horvitz-Thompson	0	0.079028	0.253276	0.065659	0.171725	0.090236	0.323928
Difference	1	0.065999	0.174221	0.046760	0.089259	0.075676	0.228148
Exponentiation	1	0.066010	0.173262	0.046973	0.090071	0.076008	0.229806
ExpRatio	1	0.071924	0.208751	0.056170	0.126811	0.083042	0.275677
Difference	2	0.061896	0.150319	0.041894	0.072563	0.072017	0.208150
Exponentiation	2	0.061813	0.148364	0.042217	0.074100	0.072274	0.209616
ExpRatio	2	0.065983	0.173529	0.048747	0.096205	0.077515	0.241732
n=100							
Horvitz-Thompson	0	0.054629	0.121016	0.046083	0.085763	0.063392	0.155941
Difference	1	0.045091	0.082673	0.033176	0.043984	0.055671	0.119216
Exponentiation	1	0.045157	0.082898	0.033443	0.045007	0.055921	0.120383
ExpRatio	1	0.049644	0.100069	0.039878	0.064161	0.059505	0.136610
Difference	2	0.041880	0.071943	0.029607	0.035960	0.052620	0.105662
Exponentiation	2	0.041917	0.072086	0.029940	0.036881	0.052676	0.106071
ExpRatio	2	0.045223	0.083422	0.034589	0.048632	0.055817	0.119154
n=200							
Horvitz-Thompson	0	0.037040	0.055113	0.028886	0.034199	0.046213	0.083300
Difference	1	0.031334	0.039189	0.021923	0.019378	0.041857	0.067203
Exponentiation	1	0.031323	0.039088	0.021899	0.019386	0.041889	0.067447
ExpRatio	1	0.033881	0.046152	0.024981	0.025738	0.043891	0.074681
Difference	2	0.028849	0.033905	0.020133	0.016159	0.040219	0.061899
Exponentiation	2	0.028735	0.033661	0.020272	0.016369	0.040227	0.062008
ExpRatio	2	0.031060	0.038938	0.022115	0.019979	0.041923	0.067677
n=300							
Horvitz-Thompson	0	0.028446	0.032859	0.022610	0.021417	0.037813	0.055494
Difference	1	0.023195	0.022167	0.017047	0.011603	0.033501	0.043550
Exponentiation	1	0.023255	0.022241	0.017101	0.011683	0.033620	0.043845
ExpRatio	1	0.025895	0.027081	0.019720	0.015993	0.035844	0.049533
Difference	2	0.021505	0.019052	0.015480	0.009754	0.032609	0.041205
Exponentiation	2	0.021502	0.019026	0.015684	0.009968	0.032655	0.041401
ExpRatio	2	0.023445	0.022483	0.017163	0.012297	0.034257	0.045268

The main conclusions derived in this study are

- The relative absolute bias of the estimators are all within a reasonable range for the different sample sizes considered.

- More efficient estimators values are obtained if the correlations between the auxiliary variables and the principal are high.

- The values of relative bias and mean square error decrease as the sampling size increase, for all estimators and all RR techniques.

- The superiority of estimators based on auxiliary information is clear: the suggested estimators belonging to the class $\widehat{Y}_g^{(r)}$ are always more efficient than the HT estimator, whatever the adopted scrambling procedure.

- The values of relative bias and mean square error are very similar between $\widehat{Y}_{gd}$ and $\widehat{Y}_g^{exp}$. The difference in bias and MSE between these estimators is smaller as the sample size increases. This is expectable because the two estimators are asymptotically equivalents.

- Difference estimator $\widehat{Y}_{gd}$ or exponentiation estimator $\widehat{Y}_g^{exp}$ are the most efficient estimator for using one or two auxiliary variables.

- The suggested estimators $\widehat{Y}_{gd}$, and $\widehat{Y}_g^{exp}$ with two auxiliary variables perform better than the estimator with one auxiliary variable, as expected.

The second simulation study was carried out with a natural population called FAM1500 [18], [29]. The study considers this population of 1500 families living in an Andalusian province to investigate their income tax return. In these simulations, we use as auxiliary variable, food expenses. The total for this variable is known. The sample is drawn by stratified sampling by house ownership. We select $T = 1000$ stratified samples of different samples sizes $n = 30, 50, 100, 200, 300$ with proportional allocation.

Table 3 shows the RB and MSE statistics for the FAM1500 population.

Table 3: Relative bias and mean square error for some estimators and some RRT devices in FAM1500 population

	Eichhorn and Hayre		Eriksson		Bar-Lev, Bobovitch and Boukai	
	Relative bias	Mean square error	Relative bias	Mean square error	Relative bias	Mean square error
n=30						
Horvitz-Thompson	0.062969	9970422	0.023375	1398428.1	0.075021	15303213
Difference	0.061768	9540168	0.016944	749632.8	0.073091	14721819
Exponentiation	0.061737	9521919	0.016935	749648.3	0.073146	14728094
ExpRatio	0.061755	9553421	0.018058	843063.9	0.073385	14791268
n=50						
Horvitz-Thompson	0.047696	5868140	0.017066	729256.6	0.059368	9310484
Difference	0.047041	5692488	0.012862	421624.8	0.058332	9065942
Exponentiation	0.046981	5674113	0.012844	420192.0	0.058328	9065732
ExpRatio	0.046915	5668912	0.013339	451857.9	0.058413	9070390
n=100						
Horvitz-Thompson	0.039153	3729670	0.012654	407192.7	0.044445	5153787
Difference	0.038019	3524888	0.009457	224479.4	0.044083	5021198
Exponentiation	0.038023	3526251	0.009448	224163.8	0.044082	5017795
ExpRatio	0.038218	3556232	0.009901	246333.3	0.044052	5021400
n=200						
Horvitz-Thompson	0.028094	1951266	0.008686	194422.6	0.033014	2895354
Difference	0.027709	1874658	0.007003	123289.9	0.032623	2823758
Exponentiation	0.027701	1874357	0.007000	123240.1	0.032615	2822372
ExpRatio	0.027710	1883025	0.007120	129032.5	0.032629	2827938
n=300						
Horvitz-Thompson	0.024538	1425025	0.007346	136750.26	0.028466	1981292
Difference	0.024299	1378528	0.006107	92563.45	0.028095	1946481
Exponentiation	0.024299	1378814	0.006104	92480.66	0.028102	1947923
ExpRatio	0.024296	1383988	0.006228	96837.40	0.028172	1947810

Results of this simulation are in accordance with those obtained in the previous study: for all randomized response models used, there is a decrease in the relative bias and the mean square error if we compared the HT estimator with others estimators that use auxiliary information. The gain in efficiency is relevant for the Eriksson model. The values of relative bias and mean square error for all models are very similar between $\widehat{Y}_{gd}$ and $\widehat{Y}_g^{exp}$. Nevertheless, the proposed estimators $\widehat{Y}_{gd}$ and $\widehat{Y}_g^{exp}$ dominate the other, for any choice of the sample size and the randomized technique.

5 Conclusions

Privacy protection is a crucial objective for both data collection and statistical analyses in the study of sensitive variables as tax evasion, sexual behaviors, reckless driving, indiscriminate gambling, and abortion. Randomized response methods can be beneficially employed for collecting and analyzing information about sensitive topics. Many studies have assessed the validity of RR methods showing that they can produce more reliable answers than other conventional data collection methods, but RRT estimates are affected by higher sampling variance than direct questioning estimates. The loss of efficiency represents the cost to pay for obtaining more reliable information by reducing response bias. Consequently, achieving efficient estimates, which are comparable with those under direct questioning, may require considerable larger sample with an obvious increasing of the survey cost. A way to reduce the sampling variance of RRT estimators is the use of auxiliary information.

This paper makes an attempt to provide a general form of estimation of a total of a sensitive variable using auxiliary information of supplementary variables. This situation is very common in the sampling practice. Many estimators were proposed to deal with the problem of estimating a total of a non sensitive variable when supplementary information is available. Nevertheless, in spite of different ideas followed to construct the estimators, most of them show the same efficiency. The unawareness of this aspect caused a proliferation of several types of estimators. This situation could be extended to the case of sensitive variables.

A class of estimators of a finite total population under a general randomized response model has been defined when the sample is obtained under a general sampling design. Estimators belonging to this class have been proofed to be asymptotically design unbiased, and their asymptotic variances has been obtained. We provide also the expression of an optimal estimator in the class, the difference estimator, that is the estimator that attains the asymptotic minimum variance bound. This estimator is studied for some elementary sampling design as simple random sampling and stratified sampling. We introduce other estimators in this class, some of them have asymptotically the same variance as the optimal difference estimator.

We have conducted a simulation study to check the performance of the proposed estimators. The results obtained from the simulation study support the theoretical background and show that, given a set of auxiliary variables, the method performs well under different scenarios in both natural and artificial populations.

In short, this paper generalizes some existing results about the use of auxiliary information in RRT ([27]) and want to contribute stopping the tentative to spread in the RRT literature new estimators by extending non-optimum estimators conceived for direct questioning surveys.

Acknowledgement

This study was partially supported by Ministerio de Educación y Ciencia (grant MTM2015-63609-R and FPU grant program, Spain) and by Consejería de Economía, Innovación, Ciencia y Empleo (grant SEJ2954, Junta de Andalucía).

References

[1] Abu-Dayyeh WA, Ahmed MS, Ahmed RA, Muttlak HA. Some estimators of finite population mean using auxiliary information. *Applied Mathematics and Computation* 2003; **139**: 287–298.

[2] Arcos A, Rueda M, Singh S. Generalized approach to randomized response for quantitative variables. *Quality and Quantity* 2015; **49**: 1239–1256.

[3] Ayachit GR. 1953 *Some aspects of large-scale sample surveys with particular reference to the ratio method of estimation.* M.Sc. Thesis, Bombay University, Bombay.

[4] Arnab R. Optional randomized response techniques for complex survey designs. *Biom. J.* 2004; **46**(1): 114–124.

[5] Bahl S, Tuteja RK. Ratio and product type exponential estimator. *Information and Optimization Sciences* 1991; **XII**(I): 159–163

[6] Barabesi L, Diana G, Perri PF. Design-based distribution function estimation for stigmatized populations *Metrika* 2013; **76**: 919–935.

[7] Bar-Lev SK, Bobovitch E, Boukai B. A note on randomized response models for quantitative data. *Metrika* 2004; **60**: 255–260.

[8] Bouza CN, Herrera C, Mitra PG. A review of randomized responses procedures: the qualitative variable case. *Investigación Oper.* 2010; **31**(3): 240–247.

[9] Chaudhuri A, Mukherjee R. 1988. *Randomized response. Theory and techniques. Statistics: Textbooks and Monographs, 85* New York: Marcel Dekker, Inc. xvi, 162 p.

[10] Cobo B, Rueda M, Arcos A. 2015. *RRTCS: Randomized Response Techniques for Complex Surveys.* R package version 1.0.

[11] Dalabebara M, Sahoo LN. A class of estimators in stratified sampling with two auxiliary variables. *Jour. Inti. Soc. Ag. Statistics* 1997; **50**(2): 144–149.

[12] Diana G, Perri PF. Estimation of finite population mean using multi-auxiliary information. *Metron* 2007; **LXV**(I): 99–112.

[13] Diana G, Perri PF. New scrambled response models for estimating the mean of a sensitive quantitative character. *J. Appl. Stat.* 2010; **37**: 1875–1890.

[14] Diana G, Perri PF. A class of estimators for quantitative sensitive data. *Statistical Papers* 2011; **52**(3): 633–650.

[15] Diana G, Perri PF. A calibration-based approach to sensitive data: a simulation study. *Journal of Applied Statistics* 2012; **39**(1): 53–65.

[16] Eichhorn BH, Hayre LS. Scrambled randomized response methods for obtaining sensitive quantitative data. *Journal of Statistical Planning and Inference* 1983; **7**: 306–316.

[17] Eriksson SA. A new model for randomized response. *International Statistical Review* 1973; **41**: 40–43.

[18] Fernandez FR, Mayor JA. 1994 *Muestreo en Poblaciones Finitas: Curso Basico.* PPU, Barcelona.

[19] Gjestvang CR, Singh S. A new randomized response model. *J. Royal Statist. Soc Ser. B* 2006; **68**: 523–530

[20] Greenberg BG, Abul-Ela AL, Simmons WR, Horvitz DG. The unrelated question RR model: Theoretical framework. *J. Amer. Statist. Assoc.* 1969; **64**: 520–539.

[21] Horvitz DG, Shah BV, Simmons WR. 1967. The unrelated question RR model. *In: Proc. Social Statist. Sec.* ASA, 65–72.

[22] Isaki CT, Fuller WA. Survey design under the regression superpopulation model. *Journal of the American Statistical Association* 1982; **77**(377): 89–96.

[23] Kadilar C, Cingi H. A new estimator using two auxiliary variables. *Applied Mathematics and Computation* 2005; **162**: 901–908.

[24] Koyuncu N, Gupta S, Sousa R. Exponential-Type Estimators of the Mean of a Sensitive Variable in the Presence of Nonsensitive Auxiliary Information. *Communications in Statistics - Simulation and Computation* 2014; **43** (7): 1583–1594.

[25] Montanari GE. Post-sampling Efficient QR-prediction in Large-sample Surveys. *International Statistical Review* 1987; **55**: 191–202.

[26] Montanari GE. On regression estimation of finite population means. *Survey Methodology* 1998; **24**,1: 69–77.

[27] Perri PF, Diana G. Scrambled response models based on auxiliary variables. *In: Advances in Theoretical and Applied Statistics, Torelli, Nicola; Pesarin, Fortunato; Bar-Hen, Avner (Eds.)*, 2013; 281-291, Spriger-Verlag, Berlin.

[28] Randles RH. On the asymptotic normality of statistics with estimated parameters. *The Annals of Statistics* 1982; **10**: 462–474.

[29] Rueda M., Arcos, A. On estimating the median from survey data using multiple auxiliary information. *Metrika* 2001; **54** (1): 59–76.

[30] Saha A. A simple randomized response technique in complex surveys. *Metron* 2007; **LXV**: 59–66.

[31] R Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria, 2015, URL: https://www.R-project.org/.

[32] Santiago A, Bouza CN, Sautto JM, Al-Omari AI. Randomized Response Procedure for the Estimation of the Population Ratio using Ranked Set Sampling. *Journal of Mathematics and Statistics* 2016; **12** (2). DOI: 10.3844/jmssp.2016.107.114.

[33] Singh S, Joarder AH, King ML. Regression analysis using scrambled responses. *Aust. J. Stat.* 1996; **38**: 201–211.

[34] Singh S, Kim JM. A pseudo-empirical log-likelihood estimator using scrambled responses. *Statist. Probab. Lett.* 2011; **81**: 345–351.

[35] Singh S, Sedory S, Arnab R. Estimation of Finite Population Variance Using Scrambled Responses in the Presence of Auxiliary Information. *Communications in Statistics - Simulation and Computation* 2015; **44**(4): 1050–1065.

[36] Singh , Tracy DS. Ridge regression using scrambled responses. *Metron* 1999; **LVII**: 147–157.

[37] Srivastava SK, Jhajj HS. A class of estimators of the population mean in survey sampling using auxiliary information. *Biometrika* 1981; **68**: 341–343.

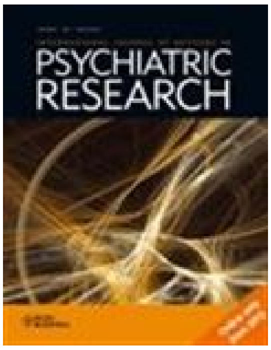
[38] Tracy D, Singh S. Calibration estimators in randomized response survey. *Metron* 1999; **LVII**: 47–68.

[39] Warner SL. Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association* 1965; **60**(309): 63–69.

Appendix A2

Application of randomized response techniques for investigating cannabis use by Spanish university students

Cobo, Beatriz; Rueda, María del Mar; Lopez-Torrecillas, Francisca (2016)
Application of randomized response techniques for investigating cannabis use by Spanish university students.
International Journal of Methods in Psychiatric Research, vol. 26, number 4
DOI: 10.1002/mpr.1517



PSYCHIATRY			
JCR Year	Impact factor	Rank	Quartile
2016	3.176	33/139	Q1

Abstract

Cannabis is the most widely used illicit drug in developed countries, and has a significant impact on mental and physical health in the general population. Although the evaluation of levels of substance use is difficult, a method such as the randomized response technique (RRT), which includes both a personal component and an assurance of confidentiality, provides a combination which can achieve a considerable degree of accuracy. Various RRT surveys have been conducted to measure the prevalence of drug use, but to date no studies have been made of the effectiveness of this approach in surveys with respect to quantitative variables related to drug use.

This paper describes a probabilistic, stratified sample of 1146 university students asking sensitive quantitative questions about cannabis use in Spanish universities, conducted using the RRT.

On comparing the results of the direct question (DQ) survey and those of the randomized response (RR) survey, we find that the number of cannabis cigarettes consumed during the past year (DQ = 3, RR = 17 approximately), and the number of days when consumption took place (DQ = 1, RR = 7) are much higher with RRT.

The advantages of RRT, reported previously and corroborated in our study, make it a useful method for investigating cannabis use.

1 Introduction

Cannabis (or marijuana) is the illicit drug that is most commonly used by young adults in Spain. On average, it is consumed by nearly 17 % of the European Union (EU) population aged 15–34 years (EMCDDA, 2015; PNSD, 2013). The prevalence of past 30-day use from 1999 to 2013 for all groups and both sexes was 22.5 % . Although young males have historically had a higher prevalence of marijuana use, current results indicate that male–female differences in marijuana 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 disturbances and numerous other conditions (Lamarine, 2012). However, the Spanish National Plan on Drugs (PNSD, 2013) has called for a change in the approach taken to understanding this phenomenon, especially as regards how young people, influenced by subcultural networks, become regular cannabis users. Patterns of consumption should be analysed so that appropriate intervention and prevention programmes can be designed.

Health care and social problems related to the use of cannabis have led researchers to investigate screening procedures aimed at detecting persons at risk. Two such procedures, which are now commonly used, are the Cannabis Abuse Screening Test (CAST; Cuenca-Royo et al., 2012), in which response options are based on a five-point Likert scale (never, rarely, occasionally, quite often and very often) and the cannabis Severity of Dependence Scale (SDS; Cuenca-Royo et al., 2012), with response options based on a four-point Likert scale (never, rarely, often and always). These screening instruments are capable of detecting (probable) cannabis dependence or problematic use and have been used in Spain in surveys for the National Plan on Drugs in schools and among the general population (Cuenca-Royo et al., 2012). The application of short screening scales to assess dependence and other problems related to the use of cannabis presents a time and cost-saving means of estimating the overall prevalence of cannabis use and of related negative consequences (Bastiani et al., 2013; Gyepesi et al., 2014; Hides et al., 2007; Legleye et al., 2013). These instruments can also help identify persons at risk, as an initial approach before using more extensive diagnostic instruments. Nevertheless, there is a need to formally evaluate the validity of the data gathered (Piontek et al., 2008). Studies by Harrison (1997) and Ramo et al. (2012) have evaluated the reliability and validity of anonymous studies of cannabis use, but these reports do not examine the other side of validity, namely the fact that respondents may lie, when faced with a question that they find embarrassing, or refuse to answer, or choose a response that prevents them from having to continue and, clearly, this situation may arise in questionnaires related to the use of illegal drugs. Other potential threats to survey accuracy are non-response and reporting error (Tourangeau and Yan, 2007).

The aim of the randomized response technique (RRT) is to decrease social desirability bias, thus guaranteeing confidentiality, improving respondent cooperation and procuring reliable estimates. This technique obtains stronger estimates of sensitive characteristics, compared to the direct question (DQ) survey, by reducing respondents' motivation to falsely report their attitudes.

The RRT was introduced by Warner in 1965. The procedure is as follows, to estimate for a community the proportion of people bearing a stigmatizing characteristic (denoted by the symbol A), like addiction to marijuana consumption, a sampled person is offered a box of a considerable number of identical cards with a proportion $p(0 < p < 1, p \neq 0.5)$ of them marked A and the rest marked A^c. The person on request is required to draw a "random" card and respond by answering "Yes" for a "match" between the card type and the person's own real characteristic or a "No" for a "non-match" before returning the card to the box.

The randomization is performed by the interviewee, and the interviewer is not permitted to observe the outcome of the randomization. The interviewee responds to the question selected by the randomization device, and the interviewer knows only the response provided. The respondent's privacy or anonymity is fully protected because no one but the respondent knows which question was answered. But it is possible statistically to derive a plausible estimate, on analysing the bunch of randomized responses thus collectively gathered, for the required proportion bearing A. It is hoped that the privacy of the person responding is securely protected.

It is assumed that respondents are more willing to provide honest answers with this technique because their answers do not reveal any information about themselves.

Some studies have addressed situations in which the response to a sensitive question results in a quantitative variable. Thus, Greenberg et al. (1971) extended RR to this case, rather than a simple *Yes* or

No. Other important work in this respect includes Bar-Lev et al. (2004); Bouza (2009); Diana and Perri (2010, 2012); Eichhorn and Hayre (1983); Saha (2007); Gjestvang and Singh (2006, 2007) and Odumade and Singh (2010). These authors concur that the RRT is an effective means of obtaining estimates with a relatively high degree of reliability. However, most studies concern only simple random sampling, while most of the surveys conducted in practice are complex, involving stratification, clustering and unequal probability in the sample selection.

The RRT was developed in an attempt to improve the quality of self-reported survey research, but it is not very often applied in the educational or psychological context (Dietz et al., 2013; Goodstadt and Gruson, 1975; Pitsch et al., 2007; Striegel et al., 2009; Weissman et al., 1986). Specifically, Goodstadt and Gruson (1975) compared 854 students' responses concerning drug use, derived from either traditional direct questioning or an indirect, more anonymous method of inquiry, the RR procedure. The results obtained in this study showed that the RR procedure produced significantly fewer response refusals and significantly higher drug use estimates, thus supporting the hypothesized greater sensitivity and validity of the RR procedure. Weissman et al. (1986) examined whether telephone interviewing could be a viable alternative to field interviewing as a method for eliciting drug use information. 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. (2009) 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 produce 0.81 % positive test results, but the RRT shows that the real prevalence is 6.8 % . In another study, Dietz et al. (2013) calculated the prevalence of students who take drugs in order to improve their cognitive performance, and reported that 20 % used cognitive-enhancing drugs. 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 may underestimate the use of these drugs. Kerkvliet (1994) combines RRTs with logit models. The academic performance of college students, their personal habits and socio-economic characteristics are used to estimate a logit model for predicting whether or not they have consumed cocaine.

Surveys based on the RRT are widely used when the questions are sensitive, and especially when the variable of interest is a qualitative one. Techniques also exist for quantitative variables, but these are not used as commonly. In our study, conducted in Spain (where RRTs are not generally used for studies of drug consumption), we took into account quantitative variables in order to make the scope of the study as complete as possible.

2 Methods

To investigate cannabis use in the Spanish universities, we conducted a survey of university students.

2.1 Participants and sampling method

The target population for this survey included students at the University of Granada and the University of Murcia. Subjects were selected using probabilistic sampling stratified by university. Respondents were randomly selected to use the RRT (sub-sample 1) and to be asked directly about illicit drug use (sub-sample 2). We determinate the sample size to estimate the population mean in stratified sampling with a coefficient of variation of 0.25. We used a pilot sample to estimate the unit relvariances.

One thousand one hundred and forty-six student participants voluntarily responded to a questionnaire. All questionnaires were administered during the class time break. All students were invited to participate in a study and provided informed signed consent.

2.2 Procedure and measure

The questionnaire is the same in two sub-samples. This questionnarie began with some academic questions followed by a set of basic demographic questions and then a block containing the sensitive questions, referring to drug use (taken from the questionnaire proposed by Miller and Rollnick, 2015).

The following sensitive questions were asked:

- P1: How many cannabis cigarettes did you consume last year?
- P2: Over the past 90 days, how many days did you consume cannabis?

In sub-sample 1 (using RRT), for the block of sensitive questions, the interviewer explained how the survey was being conducted, and gave an example of its use. The responses were randomized using a generalization of the model proposed by Bar-Lev et al. (2004) for simple random sampling and later extended by Arcos et al. (2015) for use with complex samples.

The randomizing device used was the app “Baraja Española” (“Baraja Española” is a deck composed of 40 cards, divided into four families or suits, each numbered one to seven and three figures) (Play Store, 2015), which had previously been installed on the student’s phone (see Figures 1 and 2). The application is very simple to use: for each sensitive question the user touches the screen and a card is shown. If it is a face card, the sensitive question (P1, P2) should be answered; otherwise, the real number should be given, multiplied by the number shown on the card.

Figure 3 shows the procedure of response for the two sub-samples.

The interviewer explained that this technique preserved the students’ anonymity with the aim not to provoke mistrust of them. The data collection and fieldwork was conducted by the research group FQM365 of the Andalusian Research Plan. The interviews were carried out during 2015, in Spain. Data were obtained from 654 students using the RRT and from 492 using DQ.

Figure 1: “Baraja Española” app

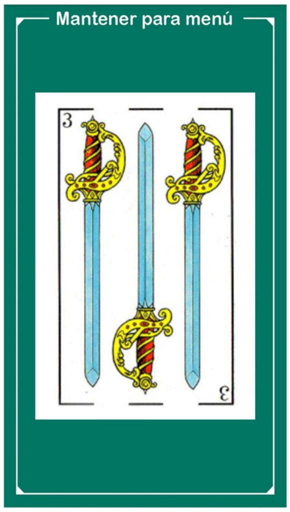


Figure 2: card from deck

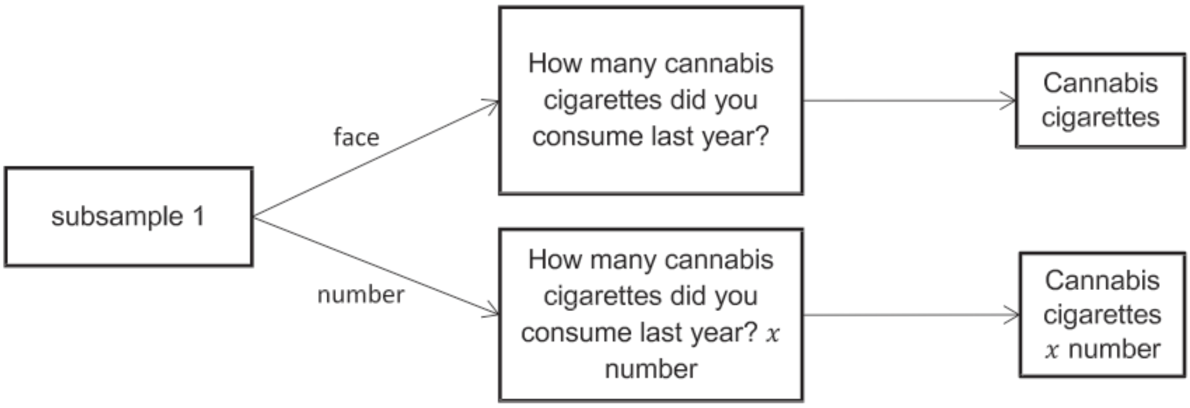


Figure 3: Procedure of response for the two sub-samples.

2.3 Response rates

Table 1 shows the non-response rates for the questions for the full sample and for the sample separated by gender considering DQ response and RR.

The non-response rate for the question was lower in the RR than in the DQ condition. These differences are statistically significant (p-value < 0.001). However, the non-response rate between men and women is similar and therefore not significant statistically (p-value > 0.05).

Table 1: Non-response rates

	DQ	RR
P1: units	0.10975610	0.05810398
P2: days	0.12601626	0.01834862
Men		
P1: units	0.30331754	0.08088235
P2: days	0.32227488	0.03676471
Women		
P1: units	0.24199288	0.04188482
P2: days	0.256227758	0.005235602

2.4 Statistical analysis

Inference is used in survey sampling to estimate the parameters of interest. The Horvitz–Thompson estimator (Singh, 2003) was used to estimate the mean values for the direct questions.

We use the unified method of estimating population surveys characteristic in RR proposed by Arnab (1994). For each unit in the sample the RR induces a random response (denoted scrambled response) and we can obtain a certain transformation of these scambled responses that is an unbiased estimation of the population mean [see Arnab (1994) or Arcos et al. (2015) for a detailed description].

2.5 Software

Standard software packages for complex surveys cannot be used directly when the sample is obtained using RRTs. Although analyses with standard statistical software, with certain modifications in the randomized variables, can yield correct point estimates of population parameters they could still yield incorrect results for the standard errors estimated.

R-packages have been developed for estimation with RR surveys, such as the RRreg package (Heck and Moshagen, 2014) and the rr package (Blair et al., 2015) but the methods implemented in these packages assume simple random sampling. Therefore, we used the package RRTCS (Rueda et al., 2015), which is the only one that incorporates estimation procedures for handling RR data obtained from complex surveys.

In this package, the function BarLev() (see Appendix) implements the BarLev model.

3 Results

The socio-demographic distribution of the samples is shown in Table 2.

The study was conducted for all students and also separating respondents by gender.

In DQ, the survey had a population of 492 individuals, of whom 42.89 % were men and 57.11 % were women. In RR, the study population was composed of 654 students, with 41.59 % men and 58.41 % women.

The point estimates of the sensitive variables and the corresponding 95 % confidence intervals for each technique (DQ and RR) are summarized in Table 3.

- P1: How many cannabis cigarettes did you consume last year?
- P2: Over the past 90 days, how many days did you consume cannabis?
- P1: By DQ, the mean number of cannabis cigarettes consumed in the previous year was approximately three, but according to RR, seventeen units were consumed.
- P2: By DQ, the students had consumed cannabis on approximately one of the previous 90 days, and on seven according to RR.

The estimate of the number of cannabis cigarettes consumed and the estimate of the number of days when consumption took place for the RR group were significantly higher than the estimates for the DQ group (p-values < 0.001).

For all questions, the standard deviation was higher for the RR than for the DQ survey. This result is as we expect because surveys conducted with RRT require large sample sizes.

If we consider the results by gender, we get more units of cannabis consumed and more number of days of consuming in men than women. This difference is statistically significant by DQ (p-value = 3.8×10^{-5} and 0.002 respectively) but this difference is not statisitically significant for RR (p-value 0.105 and 0.108 respectively) because the RR estimates have higher variances.

Table 2: Socio-demographic distribution of sample

	DQ		RR	
	Frequency	Percentage	Frequency	Percentage
Total	492	100%	654	100%
Male	211	42.8861%	272	41.5902%
Female	281	57.1138%	382	58.4098%

4 Discussion

In this paper, we present a survey related to the use of cannabis, in which a RRT is used to determine population means that are valid for any sampling design. On comparing the results of the DQ survey and those of the RR survey we find that the number of cannabis cigarettes consumed during the past year

Table 3: Estimation of the patterns of cannabis consumption

Study technique	DQ (n=492) Estimation	Standard deviation	Confidence Interval (95%) Lower bound	Upper bound	RR (n=654) Estimation	Standard deviation	Confidence Interval (95%) Lower bound	Upper bound
P1: units	3.1142	0.5592	2.0181	4.2103	17.0011	3.6790	9.7903	24.2119
P2: days	0.6837	0.1498	0.3902	0.9773	7.0179	0.9367	5.1819	8.8538
Men								
P1: units	6.3452	1.2522	3.8910	8.7995	24.4972	7.3536	10.0843	38.9100
P2: days	1.2805	0.3220	0.6494	1.9115	8.7713	1.6352	5.5664	11.9763
Women								
P1: units	0.2479	0.1090	0.0341	0.4616	11.6636	3.4850	4.8331	18.4941
P2: days	0.1304	0.06605	0.0010	0.2599	5.7693	1.0999	3.6136	7.9250

(DQ = 3, RR = 17 approximately), and the number of days when consumption took place (DQ = 1, RR = 7 approximately) are much higher with RRT in these universities.

These results are in line with those reported by Dietz et al. (2013); Goodstadt and Gruson (1975); Pitsch et al. (2007) and Striegel et al. (2009). All of these authors conclude that the RRT is an effective means of obtaining estimates with a relatively high degree of reliability. In the case of doping among professional athletes, this approach could be a promising means of evaluating the effectiveness of anti-doping programmes. The RRT has also highlighted the existence of high values for the 12-month prevalence of cognitive-enhancing drug use among university students, which suggests that other direct survey techniques underestimate this kind of drug use.

The results obtained suggest that estimates derived from standard questionnaire forms underestimate the incidence of drug use by university students. We believe that the advantages of RR revealed in this study and elsewhere make it a useful method to investigate sensitive behaviour related to alcohol and drug use. It must be stressed, however, that RR has wide confidence intervals. The randomization procedure introduces additional random error into the data and increases the standard errors of the parameters estimated: thus, larger sample sizes are needed in order to increase the statistical power. Another important issue in RRT is the choice of an appropriate randomizing device, which should be implemented in such a way as to make the confidentiality protection offered very clear to the respondent. The randomizing devices most commonly employed to date have been serial numbers on a banknote, the flip of a coin, a spinner, playing cards, numbers selected from the phone book or the respondent's month of birth. However, the new technologies currently available offer alternatives that are more attractive to users, such as mobile phones. Thanks to smartphones, we have access to many interesting applications that can help in the randomization of telephone and personal surveys (Rueda et al., 2016), especially among young people.

Acknowledgements

This study was partially supported by Ministerio de Educación, Cultura y Deporte (FPU grant program, Spain), by Ministerio de Economía y Competitividad (MINECO) and Fondo Europeo de Desarrollo Regional (FEDER) (grant MTM2015-63609-R) and by Consejería de Economía, Innovación, Ciencia y Empleo (grant SEJ2954, Junta de Andalucía, Spain).

Declaration of interest statement

The authors have no competing interests.

Appendix

Description of use of BarLev function in R

`BarLev(z,p,mu,sigma,pi,type=c("total","mean"),cl,N=NULL,pij=NULL)`

`z` vector of the observed variable; its length is equal to n (the sample size)
`p` probability of direct response
`mu` mean of the scramble variable S
`sigma` standard deviation of the scramble variable S
`pi` vector of the first-order inclusion probabilities
`type` the estimator type: total or mean
`cl` confidence level
`N` size of the population. By default it is NULL
`pij` matrix of the second-order inclusion probabilities. By default it is NULL

References

[1] Arcos A., Rueda M., Singh S.A. (2015) Generalized approach to randomized response for quantitative variables. Quality and Quantity, 49, 1239–1256. DOI:10.1007/s11135-014-0046-3.

[2] Arnab R. (1994) Non-negative variance estimator in randomized response surveys. Communications in Statistics – Theory and Method, 23(6), 1743–1752.

[3] Bar-Lev S.K., Bobovitch E., Boukai B. (2004) A note on randomized response models for quantitative data. Metrika, 60, 255–260. DOI:10.1007/s001840300308.

[4] Bastiani L., Siciliano V., Curzio O., Luppi C., Gori M., Grassi M., Molinaro S. (2013) Optimal scaling of the CAST and of SDS Scale in a national sample of adolescents. Addictive Behaviors, 38, 2060–2067. DOI: 10.1016/j.addbeh.2012.12.016.

[5] Blair G., Imai K., Zhou Y.Y. (2015) Package 'rr': Statistical methods for the randomized response technique. http://cran.r-project.org/web/packages/rr. [01 February 2016 and 06 July 2016].

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

[7] Cuenca-Royo A.M., Sánchez-Niubó A., Forero C. G., Torrens M., Suelves J.M., Domingo-Salvany A. (2012) Psychometric properties of the CAST and SDS scales in young adult cannabis users. Addictive Behaviors, 37, 709–715. DOI:10.1016/j.addbeh.2012.02.012.

[8] 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:10.1080/02664760903186031.

[9] Diana G., Perri P.F. (2012) A calibration-based approach to sensitive data: a simulation study. Journal of Applied Statistics, 39, 53–65. DOI:10.1080/02664763.2011.578615

[10] Dietz P., Striegel H., Franke A.G., Lieb K., Simon P., Ulrich R. (2013) Randomized response estimates for the 12-month prevalence of cognitive-enhancing drug use in university students. Pharmacotherapy, 33, 44–50. DOI:10.1002/phar.1166.

[11] Eichhorn B.H., Hayre L.S. (1983) Scrambled randomized response methods for obtaining sensitive quantitative data. Journal of Statistical Planning and Inference, 7, 306–316. DOI:10.1016/0378-3758(83)90002-2.

[12] EMCDDA. (2015) European Monitoring Centre for Drugs and Drug Addiction Annual Report 2015. The State of the Drug Problem in Europe, Luxembourg, Publications Office of the European Union. http://www.emcdda.europa.eu/publications/edr/trends-developments/2015 [01 February 2016 and 06 July 2016].

[13] Gjestvang C.R., Singh S. (2006) A new randomized response model. Journal of the Royal Statistical Society, Series, 68, 523–530. DOI:10.1111/j.1467-9868.2006.00554.x

[14] Gjestvang C.R., Singh S. (2007) Forced quantitative randomized response model: a new device. Metrika, 66, 243–257. DOI:10.1007/s00184-006-0108-1.

[15] Goodstadt M.S., Gruson V. (1975) The randomized response technique: a test on drug use. Journal of the Acoustical Society of America, 70 (352), 814–818.

[16] Greenberg B.G., Kuebler R.R., Abernathy J.R., Horvitz D.G. (1971) Application of the randomized response technique in obtaining quantitative data. Journal of the American Statistical Association, 66(334), 243–250.

[17] Gyepesi A., Urbán R., Farkas J., Kraus L., Piontek D., Paksi B., Horváth G., Magi A., Eisinger A., Pilling J., Kökönyei G., Kun B., Demetrovics Z. (2014) Psychometric properties of the Cannabis Abuse Screening Test in Hungarian samples of adolescents and young adults. European Addiction Research, 20, 119–128. DOI:10.1159/000353238.

[18] Harrison L. (1997) The validity of self-reported drug use in survey research: an overview and critique of research methods. NIDA Research Monograph No. 167. NIH Publication No. 97-4147. 17–36 Washington, D.C.: U.S. Government Printing Office.

[19] Heck D.W., Moshagen M. (2014) Package 'RReg': Correlation and Regression Analyses for Randomized Response Data. https://cran.r-project.org/web/packages/RRreg. [01 February 2016 and 06 July 2016].

[20] Hides L., Dawe S., Young R.M., Kavanagh D.J. (2007) The reliability and validity of the Severity of Dependence Scale for detecting cannabis dependence in psychosis. Addiction, 102, 35–40. DOI:10.1111/j.1360-0443.2006.01669.x

[21] Johnson R.M., Fairman B., Gilreath T., Xuan Z., Rothman E.F., Parnham T., Furr-Holden C.D. (2015) Past 15-year trends in adolescent marijuana use: differences by race/ethnicity and sex. Drug and Alcohol Dependence, 1, 155–158. DOI:10.1016/j.drugalcdep.2015.08.025.

[22] Kerkvliet J. (1994) Estimating a logit model with randomized data: the case of cocaine use. Australian Journal of Statistics, 36(1), 9–20.

[23] Lamarine R.J. (2012) Marijuana: modern medical chimaera. Journal of Drug Education, 42, 1–11. DOI:10.2190/DE.42.1.a.

[24] Legleye S., Piontek D., Kraus L., Morand E., Falissard B.A. (2013) Validation of the Cannabis Abuse Screening Test (CAST) using a latent class analysis of the DSM-IV among adolescents. International Journal of Methods in Psychiatric Research, 22, 16–26. DOI:10.1002/mpr.1378

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

[26] Odumade O., Singh S. (2010) An alternative to the Bar-Lev, Bobovitch and Boukai Randomized Response Model. Sociological Methods & Research, 39, 206–221. DOI: 10.1177/0049124110378094.

[27] Piontek D., Kraus L., Klempova D. (2008) Short scales to assess cannabis-related problems: a review of psychometric properties. Substance Abuse Treatment, Prevention, and Policy, 2, 25–35. DOI: 10.1186/1747-597X-3-25.

[28] Pitsch W., Emrich E., Klein M. (2007) Doping in elite sports in Germany: results of a www survey. European Journal for Sport and Society, 4, 89–102. DOI:10.1080/15298868.2015.1074101.

[29] Play Store. (2015) Download the application "Baraja Española". https://play.google.com/store/apps/details?id=cartas.barajaes [01 February 2016 and 06 July 2016].

[30] PNSD. (2013) Plan Nacional Sobre Drogas. Consumo problemático de cannabis en estudiantes españoles de 14–18 años: validación de escalas. Ministerio de Sanidad y Política Social. http://www.pnsd.msssi.gob.es/profesionales/publicaciones/catalogo/bibliotecaDigital/publicaciones/pdf/ConsProblematico_cannabis.pdf [01 February 2016 and 06 July 2016].

[31] Ramo D.E., Liu H., Prochaska J.J. (2012) Reliability and validity of young adults' anonymous online reports of marijuana use and thoughts about use. Psychology of Addictive Behaviors, 26, 801–811. DOI:10.1037/a0026201.

[32] Rueda M., Cobo B., Arcos A. (2015) Package 'RRTCS': Randomized Response Techniques for Complex Surveys. http://cran.r-project.org/web/packages/RRTCS/ [01 February 2016 and 06 July 2016].

[33] Rueda M., Cobo B., Arcos A., Arnab R. (2016) Software for randomized response techniques, in Handbook of Statistics. In Data Gathering, Analysis and Protection of Privacy through Randomized Response Techniques – Qualitative And Quantitative Human Traits, Amsterdam: Elsevier.

[34] Saha A.A. (2007) Simple randomized response technique in complex surveys. Metron LXV, 1, 59–66. DOI:10.1080/02664763.2013.846305.

[35] Singh S. (2003) Advanced Sampling Theory with Applications. In How Michael 'selected' Amy, Dordrecht: Springer.

[36] Striegel H., Ulrich R., Simon P. (2009) Randomized response estimates for doping and illicit drug use in elite athletes. Drug and Alcohol Dependence, 15, 230–232. DOI:10.1016/j.drugalcdep.2009.07.026.

[37] Tourangeau R., Yan T. (2007) Sensitive questions in surveys. Psychological Bulletin, 133, 859–863. DOI:10.1037/0033-2909.133.5.859.

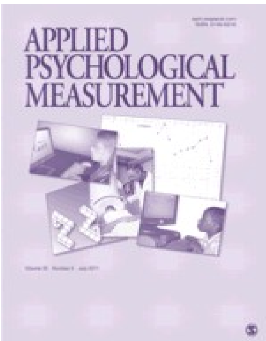
[38] Warner S.L. (1965) Randomized response: a survey technique for eliminating evasive answer bias. Journal of the American Statistical Association, 60(309), 63–69.

[39] Weissman A.N., Steer R.A., Lipton D.S. (1986) Estimating illicit drug use through telephone interviews and the randomized response technique. Drug and Alcohol Dependence, 18(3), 225–233.

Appendix A3

RRTCS: An R package for randomized response techniques in complex surveys

Rueda, María del Mar; Cobo, Beatriz; Arcos, Antonio (2015)
RRTCS: An R package for randomized response techniques in complex surveys.
Applied psychological measurement, vol. 40, number 1, pp. 78 - 80.
DOI: 10.1177/0146621615605090



SOCIAL SCIENCES, MATHEMATICAL METHODS			
JCR Year	Impact factor	Rank	Quartile
2015	1	28/49	Q3

Abstract

Randomized response (RR) techniques may be used to compile more reliable data, to protect the responden's confidentiality, and to avoid an unacceptable rate of nonresponse when the information requested is sensitive (e.g., concerning racism, drug use, abortion, delinquency, AIDS, or academic cheating). Standard RR methods are used primarily in surveys that require a binary response to a sensitive question, and seek to estimate the proportion of people presenting a given (sensitive) characteristic. Nevertheless, some studies have addressed situations in which the response to a sensitive question results in a quantitative variable. RR methods are usually developed assuming that the sample is obtained using simple random sampling. However, in practice, most surveys are complex and involve stratification, clustering, and an unequal probability of selection of the sample. Data from complex survey designs require special consideration with regard to the estimation of finite population parameters and to the corresponding variance estimation procedures, due to the reality of significant departures from the simple random sampling assumption. In such a complex survey design, unbiased variance estimation is not easy to calculate, because of clustering and the involvement of (generally complex) second-order inclusion probabilities. In view of these considerations, a new computer program has been developed to provide a method for estimating the parameters of sensitive characteristics under a variety of complex sampling designs.

1 Program Description

RRTCS is a new (R Development Core Team, 2010) package to perform the point and interval estimation of linear parameters using data obtained from RR surveys under complex sampling designs. The package works with a wide range of sampling designs, including stratified sampling, cluster sampling, unequal probabilities sampling, and any combination of these. The package consists of 21 main functions, each of which implements one of the following RR procedures for complex surveys:

- *RR procedures to estimate parameters of a qualitative sensitive characteristic:* Christofides model (Christofides, 2003); Devore model (Devore, 1977); forced response model (Boruch, 1972); Horvitz model (Greenberg, Abul-Ela, Simmons, & Horvitz, 1969); Horvitz model with unknown B (Chaudhuri, 2011, p. 42); Kuk model (Kuk, 1990); Mangat model (Mangat, 1992); Mangat model with unknown B (Chaudhuri, 2011, p. 53); Mangat and Singh model (Mangat & Singh, 1990); Mangat, Singh, and Singh model (Mangat, Singh, & Singh, 1992); Mangat, Singh, and Singh model with unknown B (Chaudhuri, 2011, p. 54); Singh and Joarder model (Singh & Joarder, 1997); Soberanis-Cruz model (Soberanis-Cruz, Ramírez Valverde, Pérez Elizalde, & González Cossio, 2008); and Warner model (Warner, 1965).

- *RR procedures to estimate parameters of a quantitative sensitive characteristic:* Bar-Lev model (Bar-Lev, Bobovitch, & Boukai, 2004), Chaudhuri and Christofides model (Chaudhuri & Christofides, 2013, p. 97), Diana and Perri 1 model (Diana & Perri, 2010, p. 1877), Diana and Perri 2 model (Diana & Perri, 2010, p. 1879), Eichhorn and Hayre model (Eichhorn & Hayre, 1983), Eriksson model (Eriksson, 1973), and Saha model (Saha, 2007).

The package includes an additional function that provides variance estimates of the RR estimators using resampling methods (Wolter, 2007) under stratified, cluster, and unequal probabilities sampling. These include the jackknife method (Quenouille, 1949), the Escobar-Berger method (Escobar & Berger, 2013), and the Campbell-Berger-Skinner method (Berger & Skinner, 2005). Finally, the package includes 20 data sets with observations from different surveys conducted in real and simulated populations using different RR techniques.

2 Availability, Documentation, and Distribution

The *RRTCS* package is available free of charge from the website http://www.r-project.org and works under Windows, Linux, and MacOS. A reference manual (in PDF format) is also available from the same website. Some examples to illustrate its use in real surveys related to sensitive issues are included in this manual on CRAN. The current version of *RRTCS* is 0.0.1. Version 3.1.3 (or later) of the R software should be installed for optimal functioning of RRTCS.

Declaration of Conflicting Interests

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

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The development of the RRTCS package was partly funded by Ministerio de Educación, Cultura y Deporte (Grant MTM2012-35650 and FPU grant program, Spain) and by Consejería de Economía, Innovación, Ciencia y Empleo (Grant SEJ2954, Junta de Andalucía, Spain).

References

[1] Bar-Lev, S. K., Bobovitch, E., & Boukai, B. (2004). A note on randomized response models for quantitative data. Metrika, 60, 255-260.

[2] Berger, Y. G., & Skinner, C. J. (2005). A jackknife variance estimator for unequal probability sampling. Journal of the Royal Statistical Society: Series B, 67, 79-89.

[3] Boruch, R. F. (1972). Relations among statistical methods for assuring confidentiality of social research data. Social Science Research, 1, 403-414.

[4] Chaudhuri, A. (2011). Randomized response and indirect questioning techniques in surveys. London, England: Chapman & Hall.

[5] Chaudhuri, A., & Christofides, T. C. (2013). Indirect questioning in sample surveys. Berlin, Germany: Springer-Verlag.

[6] Christofides, T. C. (2003). A generalized randomized response technique. Metrika, 57, 195-200.

[7] Devore, J. L. (1977). A note on the randomized response technique. Communications in Statistics-Theory and Methods, 6, 1525-1529.

[8] Diana, G., & Perri, P. F. (2010). A new scrambled response models for estimating the mean of a sensitive quantitative character. Journal of Applied Statistics, 37, 1875-1890.

[9] Eichhorn, B., & Hayre, L. S. (1983). Scrambled randomized response methods for obtaining sensitive quantitative data. Journal of Statistical Planning and Inference, 7, 307-316.

[10] Eriksson, S. A. (1973). A new model for randomized response. International Statistical Review, 41, 40-43.

[11] Escobar, E. L., & Berger, Y. G. (2013). A new replicate variance estimator for unequal probability sampling without replacement. Canadian Journal of Statistics, 41, 508-524.

[12] Greenberg, B. G., Abul-Ela, A. L., Simmons, W. R., & Horvitz, D. G. (1969). The unrelated question RR model: Theoretical framework. Journal of the American Statistical Association, 64, 520-539.

[13] Kuk, A. Y. C. (1990). Asking sensitive questions indirectly. Biometrika, 77, 436-438.

[14] Mangat, N.S. (1992) Two stage randomized response sampling procedure using unrelated question. Journal of the Indian Society of Agricultural Statistics, 44, 82-87.

[15] Mangat, N. S., & Singh, R. (1990). An alternative randomized response procedure. Biometrika, 77, 439-442.

[16] Mangat, N. S., Singh, R., & Singh, S. (1992). An improved unrelated question randomized response strategy. Calcutta Statistical Association Bulletin, 42, 277-281.

[17] Quenouille, M. H. (1949). Problems in plane sampling. The Annals of Mathematical Statistics, 20, 355-375.

[18] R Development Core Team. (2010). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Available from http://www.R-project.org

[19] Saha, A. (2007). A simple randomized response technique in complex surveys. Metron, LXV, 59-66.

[20] Singh, S., & Joarder, A. H. (1997). Unknown repeated trials in randomized response sampling. Journal of the Indian Statistical Association, 30, 109-122.

[21] Soberanis-Cruz, V., Ramírez Valverde, G., Pérez Elizalde, S., & González Cossio, F. (2008). Randomized response sampling in finite populations: A unifying approach. Agrociencia, 42, 537-549.

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

[23] Wolter, K. M. (2007). Introduction to variance estimation (2nd ed.). New York, NY: Springer.

BIBLIOGRAPHY

Package RRTCS: Randomized Response Techniques for Complex Surveys

Cobo, Beatriz; Rueda, María del Mar; Arcos, Antonio (2015)
RRTCS: Randomized Response Techniques for Complex Surveys
The package documentation is on the website:
https://cran.r-project.org/web/packages/RRTCS/

Package 'RRTCS'

November 3, 2015

Type Package

Title Randomized Response Techniques for Complex Surveys

Version 0.0.3

Date 2015-04-22

Author Beatriz Cobo Rodríguez, María del Mar Rueda García, Antonio Arcos Cebrián

Maintainer Beatriz Cobo Rodríguez <beacr@ugr.es>

Description Point and interval estimation of linear parameters with data obtained from complex surveys (including stratified and clustered samples) when randomization techniques are used. The randomized response technique was developed to obtain estimates that are more valid when studying sensitive topics. Estimators and variances for 14 randomized response methods for qualitative variables and 7 randomized response methods for quantitative variables are also implemented. In addition, some data sets from surveys with these randomization methods are included in the package.

License GPL (>= 2)

Encoding UTF-8

Imports sampling, samplingVarEst, stats

Suggests knitr

VignetteBuilder knitr

NeedsCompilation no

Repository CRAN

Date/Publication 2015-11-03 13:36:04

R topics documented:

RRTCS-package *Randomized Response Techniques for Complex Surveys*

Description

The aim of this package is to calculate point and interval estimation for linear parameters with data obtained from randomized response surveys. Twenty one RR methods are implemented for complex surveys:

- Randomized response procedures to estimate parameters of a qualitative stigmatizing characteristic: Christofides model, Devore model, Forced-Response model, Horvitz model, Horvitz model with unknown B, Kuk model, Mangat model, Mangat model with unknown B, Mangat-Singh model, Mangat-Singh-Singh model, Mangat-Singh-Singh model with unknown B, Singh-Joarder model, SoberanisCruz model and Warner model.

- Randomized response procedures to estimate parameters of a quantitative stigmatizing characteristic: BarLev model, Chaudhuri-Christofides model, Diana-Perri-1 model, Diana-Perri-2 model, Eichhorn-Hayre model, Eriksson model and Saha model.

Using the usual notation in survey sampling, we consider a finite population $U = \{1, \ldots, i, \ldots, N\}$, consisting of N different elements. Let y_i be the value of the sensitive aspect under study for the ith population element. Our aim is to estimate the finite population total $Y = \sum_{i=1}^{N} y_i$ of the variable of interest y or the population mean $\bar{Y} = \frac{1}{N}\sum_{i=1}^{N} y_i$. If we can estimate the proportion of the population presenting a certain stigmatized behaviour A, the variable y_i takes the value 1 if $i \in G_A$ (the group with the stigmatized behaviour) and the value zero otherwise. Some qualitative models use an innocuous or related attribute B whose population proportion can be known or unknown.

Assume that a sample s is chosen according to a general design p with inclusion probabilities $\pi_i = \sum_{s \ni i} p(s), i \in U$.

In order to include a wide variety of RR procedures, we consider the unified approach given by Arnab (1994). The interviews of individuals in the sample s are conducted in accordance with the RR model. For each $i \in s$ the RR induces a random response z_i (denoted scrambled response) so that the revised randomized response r_i (Chaudhuri and Christofides, 2013) is an unbiased estimation of y_i. Then, an unbiased estimator for the population total of the sensitive characteristic y is given by

$$\widehat{Y}_R = \sum_{i \in s} \frac{r_i}{\pi_i}$$

The variance of this estimator is given by:

$$V(\widehat{Y}_R) = \sum_{i \in U} \frac{V_R(r_i)}{\pi_i} + V_{HT}(r)$$

where $V_R(r_i)$ is the variance of r_i under the randomized device and $V_{HT}(r)$ is the design-variance of the Horvitz Thompson estimator of r_i values.

This variance is estimated by:

$$\widehat{V}(\widehat{Y}_R) = \sum_{i \in s} \frac{\widehat{V}_R(r_i)}{\pi_i} + \widehat{V}(r)$$

where $\widehat{V}_R(r_i)$ varies with the RR device and the estimation of the design-variance, $\widehat{V}(r)$, is obtained using Deville's method (Deville, 1993).

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

$$ci = \left(\widehat{Y}_R - z_{1-\frac{\alpha}{2}}\sqrt{\widehat{V}(\widehat{Y}_R)}, \widehat{Y}_R + z_{1-\frac{\alpha}{2}}\sqrt{\widehat{V}(\widehat{Y}_R)}\right)$$

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

Similarly, an unbiased estimator for the population mean $\bar{Y}$ is given by

$$\widehat{\bar{Y}}_R = \frac{1}{N}\sum_{i \in s}\frac{r_i}{\pi_i}$$

and an unbiased estimator for its variance is calculated as:

$$\widehat{V}(\widehat{\bar{Y}}_R) = \frac{1}{N^2}\left(\sum_{i \in s}\frac{\widehat{V}_R(r_i)}{\pi_i} + \widehat{V}(r)\right)$$

In cases where the population size N is unknown, we consider Hàjek-type estimators for the mean:

$$\widehat{\bar{Y}}_{RH} = \frac{\sum_{i \in s} r_i}{\sum_{i \in s}\frac{1}{\pi_i}}$$

and Taylor-series linearization variance estimation of the ratio (Wolter, 2007) is used.

In qualitative models, the values r_i and $\widehat{V}_R(r_i)$ for $i \in s$ are described in each model.

In some quantitative models, the values r_i and $\widehat{V}_R(r_i)$ for $i \in s$ are calculated in a general form (Arcos et al, 2015) as follows:

The randomized response given by the person i is

$$z_i = \begin{cases} y_i & \text{with probability } p_1 \\ y_i S_1 + S_2 & \text{with probability } p_2 \\ S_3 & \text{with probability } p_3 \end{cases}$$

with $p_1 + p_2 + p_3 = 1$ and where S_1, S_2 and S_3 are scramble variables whose distributions are assumed to be known. We denote by μ_i and σ_i respectively the mean and standard deviation of the variable $S_i, (i = 1, 2, 3)$.

The transformed variable is

$$r_i = \frac{z_i - p_2\mu_2 - p_3\mu_3}{p_1 + p_2\mu_1},$$

its variance is

$$V_R(r_i) = \frac{1}{(p_1 + p_2\mu_1)^2}(y_i^2 A + y_i B + C)$$

where

$$A = p_1(1-p_1) + \sigma_1^2 p_2 + \mu_1^2 p_2 - \mu_1^2 p_2^2 - 2p_1 p_2 \mu_1$$

$$B = 2p_2\mu_1\mu_2 - 2\mu_1\mu_2 p_2^2 - 2p_1 p_2 \mu_2 - 2\mu_3 p_1 p_3 - 2\mu_1\mu_3 p_2 p_3$$

$$C = (\sigma_2^2 + \mu_2^2)p_2 + (\sigma_3^2 + \mu_3^2)p_3 - (\mu_2 p_2 + \mu_3 p_3)^2$$

and the estimated variance is

$$\widehat{V}_R(r_i) = \frac{1}{(p_1 + p_2\mu_1)^2}(r_i^2 A + r_i B + C).$$

Some of the quantitative techniques considered can be viewed as particular cases of the above described procedure. Other models are described in the respective function.

Alternatively, the variance can be estimated using certain resampling methods.

Author(s)

Beatriz Cobo Rodríguez, Department of Statistics and Operations Research. University of Granada `<beacr@ugr.es>`

María del Mar Rueda García, Department of Statistics and Operations Research. University of Granada `<mrueda@ugr.es>`

Antonio Arcos Cebrián, Department of Statistics and Operations Research. University of Granada `<arcos@ugr.es>`

Maintainer: Beatriz Cobo Rodríguez `<beacr@ugr.es>`

References

Arcos, A., Rueda, M., Singh, S. (2015). *A generalized approach to randomised response for quantitative variables.* Quality and Quantity 49, 1239-1256.

Arnab, R. (1994). *Non-negative variance estimator in randomized response surveys.* Comm. Stat. Theo. Math. 23, 1743-1752.

Chaudhuri, A., Christofides, T.C. (2013). *Indirect Questioning in Sample Surveys* Springer-Verlag Berlin Heidelberg.

Deville, J.C. (1993). *Estimation de la variance pour les enquêtes en deux phases.* Manuscript, INSEE, Paris.

Wolter, K.M. (2007). *Introduction to Variance Estimation.* 2nd Edition. Springer.

`BarLev` *BarLev model*

Description

Computes the randomized response estimation, its variance estimation and its confidence interval through the BarLev model. The function can also return the transformed variable. The BarLev model was proposed by Bar-Lev et al. in 2004.

Usage

```
BarLev(z,p,mu,sigma,pi,type=c("total","mean"),cl,N=NULL,pij=NULL)
```

Arguments

z	vector of the observed variable; its length is equal to n (the sample size)
p	probability of direct response
mu	mean of the scramble variable S
sigma	standard deviation of the scramble variable S
pi	vector of the first-order inclusion probabilities
type	the estimator type: total or mean
cl	confidence level
N	size of the population. By default it is NULL
pij	matrix of the second-order inclusion probabilities. By default it is NULL

Details

The randomized response given by the person i is

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

where S is a scramble variable, whose mean μ and standard deviation σ are known.

Value

Point and confidence estimates of the sensitive characteristics using the BarLev model. The transformed variable is also reported, if required.

References

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

See Also

BarLevData

ResamplingVariance

Examples

```
data(BarLevData)
dat=with(BarLevData,data.frame(z,Pi))
p=0.6
mu=1
sigma=1
cl=0.95
BarLev(dat$z,p,mu,sigma,dat$Pi,"total",cl)
```

BarLevData *Randomized Response Survey on industrial company income*

Description

This data set contains observations from a randomized response survey conducted in a population of 2396 industrial companies in a city to investigate their income. The sample is drawn by stratified sampling with probabilities proportional to the size of the company. The randomized response technique used is the BarLev model (Bar-Lev et al, 2004) with parameter $p = 0.6$ and scramble variable $S = exp(1)$.

Usage

```
BarLevData
```

Format

A data frame containing 370 observations of a sample of companies divided into three strata. The variables are:

- ID: Survey ID
- ST: Strata ID
- z: The randomized response to the question: What was the company's income in the previous fiscal year?
- Pi: first-order inclusion probabilities

References

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

See Also

BarLev

Examples

```
data(BarLevData)
```

ChaudhuriChristofides *Chaudhuri-Christofides model*

Description

Computes the randomized response estimation, its variance estimation and its confidence interval through the Chaudhuri-Christofides model. The function can also return the transformed variable. The Chaudhuri-Christofides model can be seen in Chaudhuri and Christofides (2013, page 97).

Usage

```
ChaudhuriChristofides(z,mu,sigma,pi,type=c("total","mean"),cl,N=NULL,pij=NULL)
```

Arguments

z	vector of the observed variable; its length is equal to n (the sample size)
mu	vector with the means of the scramble variables
sigma	vector with the standard deviations of the scramble variables
pi	vector of the first-order inclusion probabilities
type	the estimator type: total or mean
cl	confidence level
N	size of the population. By default it is NULL
pij	matrix of the second-order inclusion probabilities. By default it is NULL

Details

The randomized response given by the person i is $z_i = y_i S_1 + S_2$ where S_1, S_2 are scramble variables, whose mean μ and standard deviation σ are known.

Value

Point and confidence estimates of the sensitive characteristics using the Chaudhuri-Christofides model. The transformed variable is also reported, if required.

References

Chaudhuri, A., and Christofides, T.C. (2013) *Indirect Questioning in Sample Surveys.* Springer-Verlag Berlin Heidelberg.

See Also

ChaudhuriChristofidesData

ChaudhuriChristofidesDatapij

ResamplingVariance

Examples

```
N=417
data(ChaudhuriChristofidesData)
dat=with(ChaudhuriChristofidesData,data.frame(z,Pi))
mu=c(6,6)
sigma=sqrt(c(10,10))
cl=0.95
data(ChaudhuriChristofidesDatapij)
ChaudhuriChristofides(dat$z,mu,sigma,dat$Pi,"mean",cl,pij=ChaudhuriChristofidesDatapij)
```

`ChaudhuriChristofidesData` *Randomized Response Survey on agricultural subsidies*

Description

This data set contains observations from a randomized response survey conducted in a population of 417 individuals in a municipality to investigate the agricultural subsidies. The sample is drawn by sampling with unequal probabilities (probability proportional to agricultural subsidies in the previous year). The randomized response technique used is the Chaudhuri-Christofides model (Chaudhuri and Christofides, 2013) with scramble variables $S_1 = U(1, ..., 11)$ and $S_2 = U(1, ..., 11)$.

Usage

```
ChaudhuriChristofidesData
```

Format

A data frame containing 100 observations. The variables are:

- ID: Survey ID
- z: The randomized response to the question: What are your annual agricultural subsidies?
- Pi: first-order inclusion probabilities

References

Chaudhuri, A., and Christofides, T.C. (2013) *Indirect Questioning in Sample Surveys.* Springer-Verlag Berlin Heidelberg.

See Also

`ChaudhuriChristofides`

`ChaudhuriChristofidesDatapij`

Examples

```
data(ChaudhuriChristofidesData)
```

`ChaudhuriChristofidesDatapij` *Matrix of the second-order inclusion probabilities*

Description

This dataset consists of a square matrix of dimension 100 with the first and second order inclusion probabilities for the units included in sample s, drawn from a population of size $N = 417$ according to a sampling with unequal probabilities (probability proportional to agricultural subsidies in the previous year).

Usage

```
ChaudhuriChristofidesDatapij
```

See Also

ChaudhuriChristofides

ChaudhuriChristofidesData

Examples

```
data(ChaudhuriChristofidesDatapij)
#Now, let select only the first-order inclusion probabilities
diag(ChaudhuriChristofidesDatapij)
```

`Christofides` *Christofides model*

Description

Computes the randomized response estimation, its variance estimation and its confidence interval through the Christofides model. The function can also return the transformed variable. The Christofides model was proposed by Christofides in 2003.

Usage

```
Christofides(z,mm,pm,pi,type=c("total","mean"),cl,N=NULL,pij=NULL)
```

Arguments

`z`	vector of the observed variable; its length is equal to n (the sample size)
`mm`	vector with the marks of the cards
`pm`	vector with the probabilities of previous marks
`pi`	vector of the first-order inclusion probabilities
`type`	the estimator type: total or mean
`cl`	confidence level
`N`	size of the population. By default it is NULL
`pij`	matrix of the second-order inclusion probabilities. By default it is NULL

Details

In the Christofides randomized response technique, a sampled person i is given a box with identical cards, each bearing a separate mark as $1, \ldots, k, \ldots m$ with $m \geq 2$ but in known proportions $p_1, \ldots, p_k, \ldots p_m$ with $0 < p_k < 1$ for $k = 1, \ldots, m$ and $\sum_{k=1}^m p_k = 1$. The person sampled is requested to draw one of the cards and respond

$$z_i = \begin{cases} k & \text{if a card marked } k \text{ is drawn and the person bears } A^c \\ m-k+1 & \text{if a card marked } k \text{ is drawn but the person bears } A \end{cases}$$

The transformed variable is $r_i = \frac{z_i - \mu}{m+1-2\mu}$ where $\mu = \sum_{k=1}^m kp_k$ and the estimated variance is $\widehat{V}_R(r_i) = \frac{V_R(k)}{(m+1-2\mu)^2}$, where $V_R(k) = \sum_{k=1}^m k^2 p_k - \mu^2$.

Value

Point and confidence estimates of the sensitive characteristics using the Christofides model. The transformed variable is also reported, if required.

References

Christofides, T.C. (2003). *A generalized randomized response technique.* Metrika, 57, 195-200.

See Also

`ChristofidesData`

`ResamplingVariance`

Examples

```
N=802
data(ChristofidesData)
dat=with(ChristofidesData,data.frame(z,Pi))
mm=c(1,2,3,4,5)
pm=c(0.1,0.2,0.3,0.2,0.2)
cl=0.95
Christofides(dat$z,mm,pm,dat$Pi,"mean",cl,N)
```

ChristofidesData *Randomized Response Survey on eating disorders*

Description

This data set contains observations from a randomized response survey conducted in a university to investigate eating disorders. The sample is drawn by simple random sampling without replacement. The randomized response technique used is the Christofides model (Christofides, 2003) with parameters, $mm = (1, 2, 3, 4, 5)$ and $pm = (0.1, 0.2, 0.3, 0.2, 0.2)$.

Usage

```
ChristofidesData
```

Format

A data frame containing 150 observations from a population of $N = 802$ students. The variables are:

- ID: Survey ID of student respondent
- z: The randomized response to the question: Do you have problems of anorexia or bulimia?
- Pi: first-order inclusion probabilities

References

Christofides, T.C. (2003). *A generalized randomized response technique.* Metrika, 57, 195-200.

See Also

Christofides

Examples

```
data(ChristofidesData)
```

Devore *Devore model*

Description

Computes the randomized response estimation, its variance estimation and its confidence interval through the Devore model. The function can also return the transformed variable. The Devore model was proposed by Devore in 1977.

Usage

```
Devore(z,p,pi,type=c("total","mean"),cl,N=NULL,pij=NULL)
```

Arguments

`z`	vector of the observed variable; its length is equal to n (the sample size)
`p`	proportion of cards bearing the mark A
`pi`	vector of the first-order inclusion probabilities
`type`	the estimator type: total or mean
`cl`	confidence level
`N`	size of the population. By default it is NULL
`pij`	matrix of the second-order inclusion probabilities. By default it is NULL

Details

In the Devore model, the randomized response device presents to the sampled person labelled i a box containing a large number of identical cards with a proportion p, $(0 < p < 1)$ bearing the mark A and the rest marked B (an innocuous attribute). The response solicited denoted by z_i takes the value y_i if i bears A and the card drawn is marked A. Otherwise z_i takes the value 1.

The transformed variable is $r_i = \frac{z_i-(1-p)}{p}$ and the estimated variance is $\widehat{V}_R(r_i) = r_i(r_i - 1)$.

Value

Point and confidence estimates of the sensitive characteristics using the Devore model. The transformed variable is also reported, if required.

References

Devore, J.L. (1977). *A note on the randomized response technique.* Communications in Statistics Theory and Methods 6: 1525-1529.

See Also

`DevoreData`

`ResamplingVariance`

Examples

```
data(DevoreData)
dat=with(DevoreData,data.frame(z,Pi))
p=0.7
cl=0.95
Devore(dat$z,p,dat$Pi,"total",cl)
```

DevoreData *Randomized Response Survey on instant messaging*

Description

This data set contains observations from a randomized response survey conducted in a university to investigate the use of instant messaging. The sample is drawn by stratified sampling by academic year. The randomized response technique used is the Devore model (Devore, 1977) with parameter $p = 0.7$. The unrelated question is: Are you alive?

Usage

```
DevoreData
```

Format

A data frame containing 240 observations divided into four strata. The sample is selected from a population of $N = 802$ students. The variables are:

- ID: Survey ID of student respondent
- ST: Strata ID
- z: The randomized response to the question: Do you use whatsapp / line or similar instant messaging while you study?
- Pi: first-order inclusion probabilities

References

Devore, J.L. (1977). *A note on the randomized response technique.* Communications in Statistics Theory and Methods 6: 1525-1529.

See Also

Devore

Examples

```
data(DevoreData)
```

`DianaPerri1` *Diana-Perri-1 model*

Description

Computes the randomized response estimation, its variance estimation and its confidence interval through the Diana-Perri-1 model. The function can also return the transformed variable. The Diana-Perri-1 model was proposed by Diana and Perri (2010, page 1877).

Usage

```
DianaPerri1(z,p,mu,pi,type=c("total","mean"),cl,N=NULL,method="srswr")
```

Arguments

`z`	vector of the observed variable; its length is equal to n (the sample size)
`p`	probability of direct response
`mu`	vector with the means of the scramble variables W and U
`pi`	vector of the first-order inclusion probabilities
`type`	the estimator type: total or mean
`cl`	confidence level
`N`	size of the population. By default it is NULL
`method`	method used to draw the sample: srswr or srswor. By default it is srswr

Details

In the Diana-Perri-1 model let $p \in [0,1]$ be a design parameter, controlled by the experimenter, which is used to randomize the response as follows: with probability p the interviewer responds with the true value of the sensitive variable, whereas with probability $1-p$ the respondent gives a coded value, $z_i = W(y_i + U)$ where W, U are scramble variables whose distribution is assumed to be known.

To estimate $\bar{Y}$ a sample of respondents is selected according to simple random sampling with replacement. The transformed variable is

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

where μ_W, μ_U are the means of W, U scramble variables, respectively.

The estimated variance in this model is

$$\widehat{V}(\widehat{\bar{Y}}_R) = \frac{s_z^2}{n(p + (1-p)\mu_W)^2}$$

where $s_z^2 = \sum_{i=1}^n \frac{(z_i - \bar{z})^2}{n-1}$.

If the sample is selected by simple random sampling without replacement, the estimated variance is

$$\widehat{V}(\widehat{\bar{Y}}_R) = \frac{s_z^2}{n(p + (1-p)\mu_W)^2}\left(1 - \frac{n}{N}\right)$$

Value

Point and confidence estimates of the sensitive characteristics using the Diana-Perri-1 model. The transformed variable is also reported, if required.

References

Diana, G., Perri, P.F. (2010). *New scrambled response models for estimating the mean of a sensitive quantitative character.* Journal of Applied Statistics 37 (11), 1875-1890.

See Also

DianaPerri1Data

DianaPerri2

ResamplingVariance

Examples

```
N=417
data(DianaPerri1Data)
dat=with(DianaPerri1Data,data.frame(z,Pi))
p=0.6
mu=c(5/3,5/3)
cl=0.95
DianaPerri1(dat$z,p,mu,dat$Pi,"mean",cl,N,"srswor")
```

DianaPerri1Data *Randomized Response Survey on defrauded taxes*

Description

This data set contains observations from a randomized response survey conducted in a population of 417 individuals in a municipality to investigate defrauded taxes. The sample is drawn by simple random sampling without replacement. The randomized response technique used is the Diana and Perri 1 model (Diana and Perri, 2010) with parameters $p = 0.6$, $W = F(10, 5)$ and $U = F(5, 5)$.

Usage

```
DianaPerri1Data
```

Format

A data frame containing 150 observations from a population of $N = 417$. The variables are:

- ID: Survey ID
- z: The randomized response to the question: What quantity of your agricultural subsidy do you declare in your income tax return?
- Pi: first-order inclusion probabilities

References

Diana, G., Perri, P.F. (2010). *New scrambled response models for estimating the mean of a sensitive quantitative character.* Journal of Applied Statistics 37 (11), 1875–1890.

See Also

`DianaPerri1`

Examples

```
data(DianaPerri1Data)
```

`DianaPerri2` *Diana-Perri-2 model*

Description

Computes the randomized response estimation, its variance estimation and its confidence interval through the Diana-Perri-2 model. The function can also return the transformed variable. The Diana-Perri-2 model was proposed by Diana and Perri (2010, page 1879).

Usage

```
DianaPerri2(z,mu,beta,pi,type=c("total","mean"),cl,N=NULL,method="srswr")
```

Arguments

`z`	vector of the observed variable; its length is equal to n (the sample size)
`mu`	vector with the means of the scramble variables W and U
`beta`	the constant of weighting
`pi`	vector of the first-order inclusion probabilities
`type`	the estimator type: total or mean
`cl`	confidence level
`N`	size of the population. By default it is NULL
`method`	method used to draw the sample: srswr or srswor. By default it is srswr

Details

In the Diana-Perri-2 model, each respondent is asked to report the scrambled response $z_i = W(\beta U + (1-\beta)y_i)$ where $\beta \in [0,1)$ is a suitable constant controlled by the researcher and W, U are scramble variables whose distribution is assumed to be known.

To estimate $\bar{Y}$ a sample of respondents is selected according to simple random sampling with replacement. The transformed variable is

$$r_i = \frac{z_i - \beta\mu_W\mu_U}{(1-\beta)\mu_W}$$

where μ_W, μ_U are the means of W, U scramble variables, respectively.

The estimated variance in this model is

$$\widehat{V}(\widehat{\overline{Y}}_R) = \frac{s_z^2}{n(1-\beta)^2\mu_W^2}$$

where $s_z^2 = \sum_{i=1}^{n} \frac{(z_i-\bar{z})^2}{n-1}$.

If the sample is selected by simple random sampling without replacement, the estimated variance is

$$\widehat{V}(\widehat{\overline{Y}}_R) = \frac{s_z^2}{n(1-\beta)^2\mu_W^2}\left(1-\frac{n}{N}\right)$$

Value

Point and confidence estimates of the sensitive characteristics using the Diana-Perri-2 model. The transformed variable is also reported, if required.

References

Diana, G., Perri, P.F. (2010). *New scrambled response models for estimating the mean of a sensitive quantitative character.* Journal of Applied Statistics 37 (11), 1875-1890.

See Also

DianaPerri2Data

DianaPerri1

ResamplingVariance

Examples

```
N=100000
data(DianaPerri2Data)
dat=with(DianaPerri2Data,data.frame(z,Pi))
beta=0.8
mu=c(50/48,5/3)
cl=0.95
DianaPerri2(dat$z,mu,beta,dat$Pi,"mean",cl,N,"srswor")
```

DianaPerri2Data — *Randomized Response Survey of a simulated population*

Description

This data set contains observations from a simulated randomized response survey. The interest variable is a normal distribution with mean 1500 and standard deviation 4. The sample is drawn by simple random sampling without replacement. The randomized response technique used is the Diana and Perri 2 model (Diana and Perri, 2010) with parameters $W = F(10, 50), U = F(1, 5)$ and $\beta = 0.8$.

Usage

```
DianaPerri2Data
```

Format

A data frame containing 1000 observations from a population of $N = 100000$. The variables are:

- ID: Survey ID
- z: The randomized response
- Pi: first-order inclusion probabilities

References

Diana, G., Perri, P.F. (2010). *New scrambled response models for estimating the mean of a sensitive quantitative character.* Journal of Applied Statistics 37 (11), 1875–1890.

See Also

DianaPerri2

Examples

```
data(DianaPerri2Data)
```

EichhornHayre *Eichhorn-Hayre model*

Description

Computes the randomized response estimation, its variance estimation and its confidence interval through the Eichhorn-Hayre model. The function can also return the transformed variable. The Eichhorn-Hayre model was proposed by Eichhorn and Hayre in 1983.

Usage

```
EichhornHayre(z,mu,sigma,pi,type=c("total","mean"),cl,N=NULL,pij=NULL)
```

Arguments

z	vector of the observed variable; its length is equal to n (the sample size)
mu	mean of the scramble variable S
sigma	standard deviation of the scramble variable S
pi	vector of the first-order inclusion probabilities
type	the estimator type: total or mean
cl	confidence level
N	size of the population. By default it is NULL
pij	matrix of the second-order inclusion probabilities. By default it is NULL

Details

The randomized response given by the person labelled i is $z_i = y_i S$ where S is a scramble variable whose distribution is assumed to be known.

Value

Point and confidence estimates of the sensitive characteristics using the Eichhorn-Hayre model. The transformed variable is also reported, if required.

References

Eichhorn, B.H., Hayre, L.S. (1983). *Scrambled randomized response methods for obtaining sensitive quantitative data.* Journal of Statistical Planning and Inference, 7, 306-316.

See Also

EichhornHayreData

ResamplingVariance

Examples

```
data(EichhornHayreData)
dat=with(EichhornHayreData,data.frame(z,Pi))
mu=1.111111
sigma=0.5414886
cl=0.95
#This line returns a warning showing why the variance estimation is not possible.
#See ResamplingVariance for several alternatives.
EichhornHayre(dat$z,mu,sigma,dat$Pi,"mean",cl)
```

EichhornHayreData *Randomized Response Survey on family income*

Description

This data set contains observations from a randomized response survey conducted in a population of families to investigate their income. The sample is drawn by stratified sampling by house ownership. The randomized response technique used is the Eichhorn and Hayre model (Eichhorn and Hayre, 1983) with scramble variable $S = F(20, 20)$.

Usage

```
EichhornHayreData
```

Format

A data frame containing 150 observations of a sample extracted from a population of families divided into two strata. The variables are:

- ID: Survey ID
- ST: Strata ID
- z: The randomized response to the question: What is the annual household income?
- Pi: first-order inclusion probabilities

References

Eichhorn, B.H., Hayre, L.S. (1983). *Scrambled randomized response methods for obtaining sensitive quantitative data.* Journal of Statistical Planning and Inference, 7, 306-316.

See Also

`EichhornHayre`

Examples

```
data(EichhornHayreData)
```

`Eriksson` *Eriksson model*

Description

Computes the randomized response estimation, its variance estimation and its confidence interval through the Eriksson model. The function can also return the transformed variable. The Eriksson model was proposed by Eriksson in 1973.

Usage

```
Eriksson(z,p,mu,sigma,pi,type=c("total","mean"),cl,N=NULL,pij=NULL)
```

Arguments

`z`	vector of the observed variable; its length is equal to n (the sample size)
`p`	probability of direct response
`mu`	mean of the scramble variable S
`sigma`	standard deviation of the scramble variable S
`pi`	vector of the first-order inclusion probabilities
`type`	the estimator type: total or mean
`cl`	confidence level
`N`	size of the population. By default it is NULL
`pij`	matrix of the second-order inclusion probabilities. By default it is NULL

Details

The randomized response given by the person labelled i is y_i with probability p and a discrete uniform variable S with probabilities $q_1, q_2, ..., q_j$ verifying $q_1 + q_2 + ... + q_j = 1 - p$.

Value

Point and confidence estimates of the sensitive characteristics using the Eriksson model. The transformed variable is also reported, if required.

References

Eriksson, S.A. (1973). *A new model for randomized response.* International Statistical Review 41, 40-43.

See Also

EriksssonData

ResamplingVariance

Examples

```
N=53376
data(ErikssonData)
dat=with(ErikssonData,data.frame(z,Pi))
p=0.5
mu=mean(c(0,1,3,5,8))
sigma=sqrt(4/5*var(c(0,1,3,5,8)))
cl=0.95
Eriksson(dat$z,p,mu,sigma,dat$Pi,"mean",cl,N)
```

ErikssonData *Randomized Response Survey on student cheating*

Description

This data set contains observations from a randomized response survey conducted in a university to investigate cheating behaviour in exams. The sample is drawn by stratified sampling by university faculty with uniform allocation. The randomized response technique used is the Eriksson model (Eriksson, 1973) with parameter $p = 0.5$ and S a discrete uniform variable at the points (0,1,3,5,8).

The data were used by Arcos et al. (2015).

Usage

```
ErikssonData
```

Format

A data frame containing 102 students of a sample extracted from a population of $N = 53376$ divided into four strata. The variables are:

- ID: Survey ID of student respondent
- ST: Strata ID
- z: The randomized response to the question: How many times have you cheated in an exam in the past year?
- Pi: first-order inclusion probabilities

References

Arcos, A., Rueda, M. and Singh, S. (2015). *A generalized approach to randomised response for quantitative variables.* Quality and Quantity 49, 1239-1256.

Eriksson, S.A. (1973). *A new model for randomized response.* International Statistical Review 41, 40-43.

See Also

`Eriksson`

Examples

```
data(ErikssonData)
```

`ForcedResponse` *Forced-Response model*

Description

Computes the randomized response estimation, its variance estimation and its confidence interval through the Forced-Response model. The function can also return the transformed variable. The Forced-Response model was proposed by Boruch in 1972.

Usage

```
ForcedResponse(z,p1,p2,pi,type=c("total","mean"),cl,N=NULL,pij=NULL)
```

Arguments

`z`	vector of the observed variable; its length is equal to n (the sample size)
`p1`	proportion of cards marked "Yes"
`p2`	proportion of cards marked "No"
`pi`	vector of the first-order inclusion probabilities
`type`	the estimator type: total or mean

`cl`	confidence level
`N`	size of the population. By default it is NULL
`pij`	matrix of the second-order inclusion probabilities. By default it is NULL

Details

In the Forced-Response scheme, the sampled person i is offered a box with cards: some are marked "Yes" with a proportion p_1, some are marked "No" with a proportion p_2 and the rest are marked "Genuine", in the remaining proportion $p_3 = 1-p_1-p_2$, where $0 < p_1, p_2 < 1, p_1 \neq p_2, p_1+p_2 < 1$. The person is requested to randomly draw one of them, to observe the mark on the card, and to respond

$$z_i = \begin{cases} 1 & \text{if the card is type "Yes"} \\ 0 & \text{if the card is type "No"} \\ y_i & \text{if the card is type "Genuine"} \end{cases}$$

The transformed variable is $r_i = \frac{z_i - p_1}{1-p_1-p_2}$ and the estimated variance is $\widehat{V}_R(r_i) = r_i(r_i - 1)$.

Value

Point and confidence estimates of the sensitive characteristics using the Forced-Response model. The transformed variable is also reported, if required.

References

Boruch, R.F. (1972). *Relations among statistical methods for assuring confidentiality of social research data.* Social Science Research, 1, 403-414.

See Also

ForcedResponseData

ForcedResponseDataSt

ResamplingVariance

Examples

```
data(ForcedResponseData)
dat=with(ForcedResponseData,data.frame(z,Pi))
p1=0.2
p2=0.2
cl=0.95
ForcedResponse(dat$z,p1,p2,dat$Pi,"total",cl)

#Forced Response with strata
data(ForcedResponseDataSt)
dat=with(ForcedResponseDataSt,data.frame(ST,z,Pi))
p1=0.2
p2=0.2
cl=0.95
ForcedResponse(dat$z,p1,p2,dat$Pi,"total",cl)
```

ForcedResponseData *Randomized Response Survey of a simulated population*

Description

This data set contains observations from a randomized response survey obtained from a simulated population. The main variable is a binomial distribution with a probability 0.5. The sample is drawn by simple random sampling without replacement. The randomized response technique used is the Forced Response model (Boruch, 1972) with parameters $p_1 = 0.2$ and $p_2 = 0.2$.

Usage

```
ForcedResponseData
```

Format

A data frame containing 1000 observations from a population of $N = 10000$. The variables are:

- ID: Survey ID
- z: The randomized response
- Pi: first-order inclusion probabilities

References

Boruch, R.F. (1972). *Relations among statistical methods for assuring confidentiality of social research data.* Social Science Research, 1, 403-414.

See Also

ForcedResponse

Examples

```
data(ForcedResponseData)
```

ForcedResponseDataSt *Randomized Response Survey on infertility*

Description

This data set contains observations from a randomized response survey to determine the prevalence of infertility among women of childbearing age in a population-base study. The sample is drawn by stratified sampling. The randomized response technique used is the Forced Response model (Boruch, 1972) with parameters $p_1 = 0.2$ and $p_2 = 0.2$.

Usage

```
ForcedResponseDataSt
```

Format

A data frame containing 442 observations. The variables are:

- ID: Survey ID
- ST: Strata ID
- z: The randomized response to the question: Did you ever have some medical treatment for the infertility?
- Pi: first-order inclusion probabilities

References

Boruch, R.F. (1972). *Relations among statistical methods for assuring confidentiality of social research data.* Social Science Research, 1, 403-414.

See Also

ForcedResponse

Examples

```
data(ForcedResponseDataSt)
```

Horvitz *Horvitz model*

Description

Computes the randomized response estimation, its variance estimation and its confidence interval through the Horvitz model. The function can also return the transformed variable. The Horvitz model was proposed by Horvitz et al. (1967) and by Greenberg et al. (1969).

Usage

```
Horvitz(z,p,alpha,pi,type=c("total","mean"),cl,N=NULL,pij=NULL)
```

Arguments

z	vector of the observed variable; its length is equal to n (the sample size)
p	proportion of marked cards with the sensitive question
alpha	proportion of people with the innocuous attribute
pi	vector of the first-order inclusion probabilities
type	the estimator type: total or mean

`cl`	confidence level
`N`	size of the population. By default it is NULL
`pij`	matrix of the second-order inclusion probabilities. By default it is NULL

Details

In the Horvitz model, the randomized response device presents to the sampled person labelled i a box containing a large number of identical cards, with a proportion p, $(0 < p < 1)$ bearing the mark A and the rest marked B (an innocuous attribute whose population proportion α is known). The response solicited denoted by z_i takes the value y_i if i bears A and the card drawn is marked A or if i bears B and the card drawn is marked B. Otherwise z_i takes the value 0.

The transformed variable is $r_i = \frac{z_i-(1-p)\alpha}{p}$ and the estimated variance is $\widehat{V}_R(r_i) = r_i(r_i-1)$.

Value

Point and confidence estimates of the sensitive characteristics using the Horvitz model. The transformed variable is also reported, if required.

References

Greenberg, B.G., Abul-Ela, A.L., Simmons, W.R., Horvitz, D.G. (1969). *The unrelated question RR model: Theoretical framework.* Journal of the American Statistical Association, 64, 520-539.

Horvitz, D.G., Shah, B.V., Simmons, W.R. (1967). *The unrelated question RR model.* Proceedings of the Social Statistics Section of the American Statistical Association. 65-72. Alexandria, VA: ASA.

See Also

HorvitzData

HorvitzDataStCl

HorvitzDataRealSurvey

HorvitzUB

SoberanisCruz

ResamplingVariance

Examples

```
N=10777
data(HorvitzData)
dat=with(HorvitzData,data.frame(z,Pi))
p=0.5
alpha=0.6666667
cl=0.95
Horvitz(dat$z,p,alpha,dat$Pi,"mean",cl,N)

#Horvitz real survey
N=10777
n=710
```

```
data(HorvitzDataRealSurvey)
p=0.5
alpha=1/12
pi=rep(n/N,n)
cl=0.95
Horvitz(HorvitzDataRealSurvey$sex,p,alpha,pi,"mean",cl,N)
```

HorvitzData *Randomized Response Survey on student bullying*

Description

This data set contains observations from a randomized response survey conducted in a university to investigate bullying. The sample is drawn by simple random sampling without replacement. The randomized response technique used is the Horvitz model (Horvitz et al., 1967 and Greenberg et al., 1969) with parameter $p = 0.5$. The unrelated question is: Were you born between the 1st and 20th of the month? with $\alpha = 0.6666667$.

Usage

```
HorvitzData
```

Format

A data frame containing a sample of 411 observations from a population of $N = 10777$ students. The variables are:

- ID: Survey ID of student respondent
- z: The randomized response to the question: Have you been bullied?
- Pi: first-order inclusion probabilities

References

Greenberg, B.G., Abul-Ela, A.L., Simmons, W.R., Horvitz, D.G. (1969). *The unrelated question RR model: Theoretical framework.* Journal of the American Statistical Association, 64, 520-539.

Horvitz, D.G., Shah, B.V., Simmons, W.R. (1967). *The unrelated question RR model.* Proceedings of the Social Statistics Section of the American Statistical Association. 65-72. Alexandria, VA: ASA.

See Also

Horvitz

Examples

```
data(HorvitzData)
```

HorvitzDataRealSurvey *Randomized Response Survey on a sensitive questions*

Description

This data set contains observations from a randomized response survey conducted in a university to sensitive questions described below. The sample is drawn by simple random sampling without replacement. The randomized response technique used is the Horvitz model (Horvitz et al., 1967 and Greenberg et al., 1969) with parameter $p = 0.5$. Each sensitive question is associated with a unrelated question.

1. Were you born in July? with $\alpha = 1/12$
2. Does your ID number end in 2? with $\alpha = 1/10$
3. Were you born of 1 to 20 of the month? with $\alpha = 20/30$
4. Does your ID number end in 5? with $\alpha = 1/10$
5. Were you born of 15 to 25 of the month? with $\alpha = 10/30$
6. Were you born in April? with $\alpha = 1/12$

Usage

```
HorvitzData
```

Format

A data frame containing a sample of 710 observations from a population of $N = 10777$ students. The variables are:

- copied: The randomized response to the question: Have you ever copied in an exam?
- fought: The randomized response to the question: Have you ever fought with a teacher?
- bullied: The randomized response to the question: Have you been bullied?
- bullying: The randomized response to the question: Have you ever bullied someone?
- drug: The randomized response to the question: Have you ever taken drugs on the campus?
- sex: The randomized response to the question: Have you had sex on the premises of the university?

References

Greenberg, B.G., Abul-Ela, A.L., Simmons, W.R., Horvitz, D.G. (1969). *The unrelated question RR model: Theoretical framework.* Journal of the American Statistical Association, 64, 520-539.

Horvitz, D.G., Shah, B.V., Simmons, W.R. (1967). *The unrelated question RR model.* Proceedings of the Social Statistics Section of the American Statistical Association. 65-72. Alexandria, VA: ASA.

See Also

Horvitz

Examples

```
data(HorvitzDataRealSurvey)
```

`HorvitzDataStCl` *Randomized Response Survey on infidelity*

Description

This data set contains observations from a randomized response survey conducted in a university to investigate the infidelity. The sample is drawn by stratified (by faculty) cluster (by group) sampling. The randomized response technique used is the Horvitz model (Horvitz et al., 1967 and Greenberg et al., 1969) with parameter $p = 0.6$. The unrelated question is: Does your identity card end in an odd number? with a probability $\alpha = 0.5$.

Usage

```
HorvitzDataStCl
```

Format

A data frame containing 365 observations from a population of $N = 1500$ students divided into two strata. The first strata has 14 cluster and the second has 11 cluster. The variables are:

- ID: Survey ID of student respondent
- ST: Strata ID
- CL: Cluster ID
- z: The randomized response to the question: Have you ever been unfaithful?
- Pi: first-order inclusion probabilities

References

Greenberg, B.G., Abul-Ela, A.L., Simmons, W.R., Horvitz, D.G. (1969). *The unrelated question RR model: Theoretical framework.* Journal of the American Statistical Association, 64, 520-539.

Horvitz, D.G., Shah, B.V., Simmons, W.R. (1967). *The unrelated question RR model.* Proceedings of the Social Statistics Section of the American Statistical Association. 65-72. Alexandria, VA: ASA.

See Also

Horvitz

Examples

```
data(HorvitzDataStCl)
```

`HorvitzUB` *Horvitz-UB model*

Description

Computes the randomized response estimation, its variance estimation and its confidence interval through the Horvitz model (Horvitz et al., 1967, and Greenberg et al., 1969) when the proportion of people bearing the innocuous attribute is unknown. The function can also return the transformed variable. The Horvitz-UB model can be seen in Chaudhuri (2011, page 42).

Usage

```
HorvitzUB(I,J,p1,p2,pi,type=c("total","mean"),cl,N=NULL,pij=NULL)
```

Arguments

`I`	first vector of the observed variable; its length is equal to n (the sample size)
`J`	second vector of the observed variable; its length is equal to n (the sample size)
`p1`	proportion of marked cards with the sensitive attribute in the first box
`p2`	proportion of marked cards with the sensitive attribute in the second box
`pi`	vector of the first-order inclusion probabilities
`type`	the estimator type: total or mean
`cl`	confidence level
`N`	size of the population. By default it is NULL
`pij`	matrix of the second-order inclusion probabilities. By default it is NULL

Details

In the Horvitz model, when the population proportion α is not known, two independent samples are taken. Two boxes are filled with a large number of similar cards except that in the first box a proportion $p_1(0 < p_1 < 1)$ of them is marked A and the complementary proportion $(1 - p_1)$ each bearing the mark B, while in the second box these proportions are p_2 and $1 - p_2$, maintaining p_2 different from p_1. A sample is chosen and every person sampled is requested to draw one card randomly from the first box and to repeat this independently with the second box. In the first case, a randomized response should be given, as

$$I_i = \begin{cases} 1 & \text{if card type draws "matches" the sensitive trait } A \text{ or the innocuous trait } B \\ 0 & \text{if there is "no match" with the first box} \end{cases}$$

and the second case given a randomized response as

$$J_i = \begin{cases} 1 & \text{if there is "match" for the second box} \\ 0 & \text{if there is "no match" for the second box} \end{cases}$$

The transformed variable is $r_i = \frac{(1-p_2)I_i-(1-p_1)J_i}{p_1-p_2}$ and the estimated variance is $\widehat{V}_R(r_i) = r_i(r_i - 1)$.

Value

Point and confidence estimates of the sensitive characteristics using the Horvitz-UB model. The transformed variable is also reported, if required.

References

Chaudhuri, A. (2011). *Randomized response and indirect questioning techniques in surveys.* Boca Raton: Chapman and Hall, CRC Press.

Greenberg, B.G., Abul-Ela, A.L., Simmons, W.R., Horvitz, D.G. (1969). *The unrelated question RR model: Theoretical framework.* Journal of the American Statistical Association, 64, 520-539.

Horvitz, D.G., Shah, B.V., Simmons, W.R. (1967). *The unrelated question RR model.* Proceedings of the Social Statistics Section of the American Statistical Association. 65-72. Alexandria, VA: ASA.

See Also

HorvitzUBData

Horvitz

ResamplingVariance

Examples

```
N=802
data(HorvitzUBData)
dat=with(HorvitzUBData,data.frame(I,J,Pi))
p1=0.6
p2=0.7
cl=0.95
HorvitzUB(dat$I,dat$J,p1,p2,dat$Pi,"mean",cl,N)
```

HorvitzUBData *Randomized Response Survey on drugs use*

Description

This data set contains observations from a randomized response survey conducted in a university to investigate drugs use. The sample is drawn by cluster sampling with the probabilities proportional to the size. The randomized response technique used is the Horvitz-UB model (Chaudhuri, 2011) with parameters $p_1 = 0.6$ and $p_2 = 0.7$.

Usage

```
HorvitzUBData
```

Format

A data frame containing a sample of 188 observations from a population of $N = 802$ students divided into four cluster. The variables are:

- ID: Survey ID of student respondent
- CL: Cluster ID
- I: The first randomized response to the question: Have you ever used drugs?
- J: The second randomized response to the question: Have you ever used drugs?
- Pi: first-order inclusion probabilities

References

Chaudhuri, A. (2011). *Randomized response and indirect questioning techniques in surveys.* Boca Raton: Chapman and Hall, CRC Press.

Greenberg, B.G., Abul-Ela, A.L., Simmons, W.R., Horvitz, D.G. (1969). *The unrelated question RR model: Theoretical framework.* Journal of the American Statistical Association, 64, 520-539.

Horvitz, D.G., Shah, B.V., Simmons, W.R. (1967). *The unrelated question RR model.* Proceedings of the Social Statistics Section of the American Statistical Association. 65-72. Alexandria, VA: ASA.

See Also

HorvitzUB

Examples

```
data(HorvitzUBData)
```

Kuk *Kuk model*

Description

Computes the randomized response estimation, its variance estimation and its confidence through the Kuk model. The function can also return the transformed variable. The Kuk model was proposed by Kuk in 1990.

Usage

```
Kuk(z,p1,p2,k,pi,type=c("total","mean"),cl,N=NULL,pij=NULL)
```

Arguments

`z`	vector of the observed variable; its length is equal to n (the sample size)
`p1`	proportion of red cards in the first box
`p2`	proportion of red cards in the second box
`k`	total number of cards drawn
`pi`	vector of the first-order inclusion probabilities
`type`	the estimator type: total or mean
`cl`	confidence level
`N`	size of the population. By default it is NULL
`pij`	matrix of the second-order inclusion probabilities. By default it is NULL

Details

In the Kuk randomized response technique, the sampled person i is offered two boxes. Each box contains cards that are identical exception colour, either red or white, in sufficiently large numbers with proportions p_1 and $1 - p_1$ in the first and p_2 and $1 - p_2$, in the second ($p_1 \neq p_2$). The person sampled is requested to use the first box, if his/her trait is A and the second box if his/her trait is A^c and to make k independent draws of cards, with replacement each time. The person is asked to reports $z_i = f_i$, the number of times a red card is drawn.

The transformed variable is $r_i = \frac{f_i/k-p_2}{p_1-p_2}$ and the estimated variance is $\widehat{V}_R(r_i) = br_i + c$, where $b = \frac{1-p_1-p_2}{k(p_1-p_2)}$ and $c = \frac{p_2(1-p_2)}{k(p_1-p_2)^2}$.

Value

Point and confidence estimates of the sensitive characteristics using the Kuk model. The transformed variable is also reported, if required.

References

Kuk, A.Y.C. (1990). *Asking sensitive questions indirectly.* Biometrika, 77, 436-438.

See Also

`KukData`

`ResamplingVariance`

Examples

```
N=802
data(KukData)
dat=with(KukData,data.frame(z,Pi))
p1=0.6
p2=0.2
k=25
cl=0.95
Kuk(dat$z,p1,p2,k,dat$Pi,"mean",cl,N)
```

`KukData` *Randomized Response Survey on excessive sexual activity*

Description

This data set contains the data from a randomized response survey conducted in a university to investigate excessive sexual activity. The sample is drawn by simple random sampling without replacement. The randomized response technique used is the Kuk model (Kuk, 1990) with parameters $p_1 = 0.6$, $p_2 = 0.2$ and $k = 25$.

Usage

```
KukData
```

Format

A data frame containing 200 observations from a population of $N = 802$ students. The variables are:

- ID: Survey ID of student respondent
- z: The randomized response to the question: Do you practice excessive sexual activity?
- Pi: first-order inclusion probabilities

References

Kuk, A.Y.C. (1990). *Asking sensitive questions indirectly.* Biometrika, 77, 436-438.

See Also

`Kuk`

Examples

```
data(KukData)
```

`Mangat` *Mangat model*

Description

Computes the randomized response estimation, its variance estimation and its confidence interval through the Mangat model. The function can also return the transformed variable. The Mangat model was proposed by Mangat in 1992.

Usage

```
Mangat(z,p,alpha,t,pi,type=c("total","mean"),cl,N=NULL,pij=NULL)
```

Arguments

`z`	vector of the observed variable; its length is equal to n (the sample size)
`p`	proportion of marked cards with the sensitive attribute in the second box
`alpha`	proportion of people with the innocuous attribute
`t`	proportion of marked cards with "True" in the first box
`pi`	vector of the first-order inclusion probabilities
`type`	the estimator type: total or mean
`cl`	confidence level
`N`	size of the population. By default it is NULL
`pij`	matrix of the second-order inclusion probabilities. By default it is NULL

Details

In Mangat's method, there are two boxes, the first containing cards marked "True" and "RR" in proportions t and $(1-t), (0 < t < 1)$. A person drawing a "True" marked card is asked to tell the truth about bearing A or A^c. A person drawing and "RR" marked card is then asked to apply Horvitz's device by drawing a card from a second box with cards marked A and B in proportions p and $(1-p)$. If an A marked card is now drawn the truthful response will be about bearing the sensitive attribute A and otherwise about B. The true proportion of people bearing A is to be estimated but α, the proportion of people bearing the innocuous trait B unrelated to A, is assumed to be known. The observed variable is

$$z_i = \begin{cases} y_i & \text{if a card marked "True" is drawn from the first box} \\ I_i & \text{if a card marked "RR" is drawn} \end{cases}$$

where

$$I_i = \begin{cases} 1 & \text{if the type of card drawn from the second box matches trait } A \text{ or } B \\ 0 & \text{if the type of card drawn from the second box does not match trait } A \text{ or } B. \end{cases}$$

The transformed variable is $r_i = \frac{z_i-(1-t)(1-p)\alpha}{t+(1-t)p}$ and the estimated variance is $\widehat{V}_R(r_i) = r_i(r_i - 1)$.

Value

Point and confidence estimates of the sensitive characteristics using the Mangat model. The transformed variable is also reported, if required.

References

Mangat, N.S. (1992). *Two stage randomized response sampling procedure using unrelated question.* Journal of the Indian Society of Agricultural Statistics, 44, 82-87.

See Also

MangatUB

ResamplingVariance

MangatSingh *Mangat-Singh model*

Description

Computes the randomized response estimation, its variance estimation and its confidence interval through the Mangat-Singh model. The function can also return the transformed variable. The Mangat-Singh model was proposed by Mangat and Singh in 1990.

Usage

```
MangatSingh(z,p,t,pi,type=c("total","mean"),cl,N=NULL,pij=NULL)
```

Arguments

z	vector of the observed variable; its length is equal to n (the sample size)
p	proportion of marked cards with the sensitive attribute in the second box
t	proportion of marked cards with "True" in the first box
pi	vector of the first-order inclusion probabilities
type	the estimator type: total or mean
cl	confidence level
N	size of the population. By default it is NULL
pij	matrix of the second-order inclusion probabilities. By default it is NULL

Details

In the Mangat-Singh model, the sampled person is offered two boxes of cards. In the first box a known proportion $t, (0 < t < 1)$ of cards is marked "True" and the remaining ones are marked "RR". One card is to be drawn, observed and returned to the box. If the card drawn is marked "True", then the respondent should respond "Yes" if he/she belongs to the sensitive category, otherwise "No". If the card drawn is marked "RR", then the respondent must use the second box and draw a card from it. This second box contains a proportion $p, (0 < p < 1, p \neq 0.5)$ of cards marked A and the remaining ones are marked A^c. If the card drawn from the second box matches his/her status as related to the stigmatizing characteristic, he/she must respond "Yes", otherwise "No". The randomized response from a person labelled i is assumed to be:

$$z_i = \begin{cases} y_i & \text{if a card marked "True" is drawn from the first box} \\ I_i & \text{if a card marked "RR" is drawn} \end{cases}$$

$$I_i = \begin{cases} 1 & \text{if the "card type" } A \text{ or } A^c \text{ "matches" the genuine trait } A \text{ or } A^c \\ 0 & \text{if a "mismatch" is observed} \end{cases}$$

The transformed variable is $r_i = \frac{z_i-(1-t)(1-p)}{t+(1-t)(2p-1)}$ and the estimated variance is $\widehat{V}_R(r_i) = r_i(r_i - 1)$.

Value

Point and confidence estimates of the sensitive characteristics using the Mangat-Singh model. The transformed variable is also reported, if required.

References

Mangat, N.S., Singh, R. (1990). *An alternative randomized response procedure.* Biometrika, 77, 439-442.

See Also

MangatSinghData

ResamplingVariance

Examples

```
N=802
data(MangatSinghData)
dat=with(MangatSinghData,data.frame(z,Pi))
p=0.7
t=0.55
cl=0.95
MangatSingh(dat$z,p,t,dat$Pi,"mean",cl,N)
```

MangatSinghData *Randomized Response Survey on cannabis use*

Description

This data set contains observations from a randomized response survey conducted in a university to investigate cannabis use. The sample is drawn by stratified sampling by academic year. The randomized response technique used is the Mangat-Singh model (Mangat and Singh, 1990) with parameters $p = 0.7$ and $t = 0.55$.

Usage

```
MangatSinghData
```

Format

A data frame containing 240 observations from a population of $N = 802$ students divided into four strata. The variables are:

- ID: Survey ID of student respondent
- ST: Strata ID
- z: The randomized response to the question: Have you ever used cannabis?
- Pi: first-order inclusion probabilities

References

Mangat, N.S., Singh, R. (1990). *An alternative randomized response procedure.* Biometrika, 77, 439-442.

See Also

MangatSingh

Examples

```
data(MangatSinghData)
```

MangatSinghSingh *Mangat-Singh-Singh model*

Description

Computes the randomized response estimation, its variance estimation and its confidence interval through the Mangat-Singh-Singh model. The function can also return the transformed variable. The Mangat-Singh-Singh model was proposed by Mangat, Singh and Singh in 1992.

Usage

```
MangatSinghSingh(z,p,alpha,pi,type=c("total","mean"),cl,N=NULL,pij=NULL)
```

Arguments

z	vector of the observed variable; its length is equal to n (the sample size)
p	proportion of marked cards with the sensitive attribute in the box
alpha	proportion of people with the innocuous attribute
pi	vector of the first-order inclusion probabilities
type	the estimator type: total or mean
cl	confidence level
N	size of the population. By default it is NULL
pij	matrix of the second-order inclusion probabilities. By default it is NULL

Details

In the Mangat-Singh-Singh scheme, a person labelled i, if sampled, is offered a box and told to answer "yes" if the person bears A. But if the person bears A^c then the person is to draw a card from the box with a proportion $p(0 < p < 1)$ of cards marked A and the rest marked B; if the person draws a card marked B he/she is told to say "yes" again if he/she actually bears B; in any other case, "no" is to be answered.

The transformed variable is $r_i = \frac{z_i-(1-p)\alpha}{1-(1-p)\alpha}$ and the estimated variance is $\widehat{V}_R(r_i) = r_i(r_i - 1)$.

Value

Point and confidence estimates of the sensitive characteristics using the Mangat-Singh-Singh model. The transformed variable is also reported, if required.

References

Mangat, N.S., Singh, R., Singh, S. (1992). *An improved unrelated question randomized response strategy.* Calcutta Statistical Association Bulletin, 42, 277-281.

See Also

MangatSinghSinghData

MangatSinghSinghUB

ResamplingVariance

Examples

```
data(MangatSinghSinghData)
dat=with(MangatSinghSinghData,data.frame(z,Pi))
p=0.6
alpha=0.5
cl=0.95
MangatSinghSingh(dat$z,p,alpha,dat$Pi,"total",cl)
```

MangatSinghSinghData *Randomized Response Survey on internet betting*

Description

This data set contains observations from a randomized response survey conducted in a university to investigate internet betting. The sample is drawn by stratified (by faculty) cluster (by group) sampling. The randomized response technique used is the Mangat-Singh-Singh model (Mangat, Singh and Singh, 1992) with parameter $p = 0.6$. The unrelated question is: Does your identity card end in an even number? with a probability $\alpha = 0.5$.

Usage

```
MangatSinghSinghData
```

Format

A data frame containing 802 observations from a population of students divided into eight strata. Each strata has a certain number of clusters, totalling 23. The variables are:

- ID: Survey ID of student respondent
- ST: Strata ID
- CL: Cluster ID

- z: The randomized response to the question: In the last year, did you bet on internet?
- Pi: first-order inclusion probabilities

References

Mangat, N.S., Singh, R., Singh, S. (1992). *An improved unrelated question randomized response strategy.* Calcutta Statistical Association Bulletin, 42, 277-281.

See Also

MangatSinghSingh

Examples

```
data(MangatSinghSinghData)
```

MangatSinghSinghUB *Mangat-Singh-Singh-UB model*

Description

Computes the randomized response estimation, its variance estimation and its confidence interval through the Mangat-Singh-Singh model (Mangat el al., 1992) when the proportion of people bearing the innocuous attribute is unknown. The function can also return the transformed variable. The Mangat-Singh-Singh-UB model can be seen in Chauduri (2011, page 54).

Usage

```
MangatSinghSinghUB(I,J,p1,p2,pi,type=c("total","mean"),cl,N=NULL,pij=NULL)
```

Arguments

I	first vector of the observed variable; its length is equal to n (the sample size)
J	second vector of the observed variable; its length is equal to n (the sample size)
p1	proportion of marked cards with the sensitive attribute in the first box
p2	proportion of marked cards with the sensitive attribute in the second box
pi	vector of the first-order inclusion probabilities
type	the estimator type: total or mean
cl	confidence level
N	size of the population. By default it is NULL
pij	matrix of the second-order inclusion probabilities. By default it is NULL

Details

A person labelled i who is chosen, is instructed to say "yes" if he/she bears A, and if not, to randomly take a card from a box containing cards marked A, B in proportions p_1 and $(1-p_1), (0 < p_1 < 1)$; they are then told to report the value x_i if a B-type card is chosen and he/she bears B; otherwise he/she is told to report "No". This entire exercise is to be repeated independently with the second box with A and B-marked cards in proportions p_2 and $(1-p_2), (0 < p_2 < 1, p_2 \neq p_1)$. Let I_i the first response and J_i the second response for the respondent i.

The transformed variable is $r_i = \frac{(1-p_2)I_i-(1-p_1)J_i}{p_1-p_2}$ and the estimated variance is $\widehat{V}_R(r_i) = r_i(r_i - 1)$.

Value

Point and confidence estimates of the sensitive characteristics using the Mangat-Singh-Singh-UB model. The transformed variable is also reported, if required.

References

Chaudhuri, A. (2011). *Randomized response and indirect questioning techniques in surveys.* Boca Raton: Chapman and Hall, CRC Press.

Mangat, N.S., Singh, R., Singh, S. (1992). *An improved unrelated question randomized response strategy.* Calcutta Statistical Association Bulletin, 42, 277-281.

See Also

MangatSinghSinghUBData

MangatSinghSingh

ResamplingVariance

Examples

```
N=802
data(MangatSinghSinghUBData)
dat=with(MangatSinghSinghUBData,data.frame(I,J,Pi))
p1=0.6
p2=0.8
cl=0.95
MangatSinghSinghUB(dat$I,dat$J,p1,p2,dat$Pi,"mean",cl,N)
```

MangatSinghSinghUBData *Randomized Response Survey on overuse of the internet*

Description

This data set contains observations from a randomized response survey conducted in a university to investigate overuse of the internet. The sample is drawn by simple random sampling without replacement. The randomized response technique used is the Mangat-Singh-Singh-UB model (Chaudhuri, 2011) with parameters $p_1 = 0.6$ and $p_2 = 0.8$.

Usage

```
MangatSinghSinghUBData
```

Format

A data frame containing 500 observations. The variables are:

- ID: Survey ID of student respondent
- z: The randomized response to the question: Do you spend a lot of time surfing the internet?
- Pi: first-order inclusion probabilities

References

Chaudhuri, A. (2011). *Randomized response and indirect questioning techniques in surveys.* Boca Raton: Chapman and Hall, CRC Press.

Mangat, N.S., Singh, R., Singh, S. (1992). *An improved unrelated question randomized response strategy.* Calcutta Statistical Association Bulletin, 42, 277-281.

See Also

MangatSinghSinghUB

Examples

```
data(MangatSinghSinghUBData)
```

MangatUB *Mangat-UB model*

Description

Computes the randomized response estimation, its variance estimation and its confidence interval through the Mangat model (Mangat, 1992) when the proportion of people bearing the innocuous attribute is unknown. The function can also return the transformed variable. The Mangat-UB model can be seen in Chaudhuri (2011, page 53).

Usage

```
MangatUB(I,J,p1,p2,t,pi,type=c("total","mean"),cl,N=NULL,pij=NULL)
```

Arguments

I	first vector of the observed variable; its length is equal to n (the sample size)
J	second vector of the observed variable; its length is equal to n (the sample size)
p1	proportion of marked cards with the sensitive attribute in the second box
p2	proportion of marked cards with the sensitive attribute in the third box
t	probability of response to the sensitive questions without using random response in the first box
pi	vector of the first-order inclusion probabilities
type	the estimator type: total or mean
cl	confidence level
N	size of the population. By default it is NULL
pij	matrix of the second-order inclusion probabilities. By default it is NULL

Details

In Mangat's extended scheme, three boxes containing cards are presented to the sampled person, labelled i. The first box contains cards marked "True" and "RR" in proportions t and $1-t$, the second one contains A and B-marked cards in proportions p_1 and $(1-p_1),(0<p_1<1)$ and the third box contains A and B-marked cards in proportions p_2 and $1-p_2,(0<p_2<1), p_1 \neq p_2$. The subject is requested to draw a card from the first box. The sample respondent i is then instructed to tell the truth, using "the first box and if necessary also the second box" and next, independently, to give a second truthful response also using "the first box and if necessary, the third box." Let I_i represent the first response and J_i the second response for respondent i.

The transformed variable is $r_i = \frac{(1-p_2)I_i-(1-p_1)J_i}{p_1-p_2}$ and the estimated variance is $\widehat{V}_R(r_i) = r_i(r_i - 1)$.

Value

Point and confidence estimates of the sensitive characteristics using the Mangat-UB model. The transformed variable is also reported, if required.

References

Chaudhuri, A. (2011). *Randomized response and indirect questioning techniques in surveys.* Boca Raton: Chapman and Hall, CRC Press.

Mangat, N.S. (1992). *Two stage randomized response sampling procedure using unrelated question.* Journal of the Indian Society of Agricultural Statistics, 44, 82-87.

See Also

Mangat

ResamplingVariance

ResamplingVariance *Resampling variance of randomized response models*

Description

To estimate the variance of the randomized response estimators using resampling methods.

Usage

```
ResamplingVariance(output,pi,type=c("total","mean"),option=1,N=NULL,pij=NULL,str=NULL,
clu=NULL,srswr=FALSE)
```

Arguments

`output`	output of the qualitative or quantitative method depending on the variable of interest
`pi`	vector of the first-order inclusion probabilities. By default it is NULL
`type`	the estimator type: total or mean
`option`	method used to calculate the variance (1: Jackknife, 2: Escobar-Berger, 3: Campbell-Berger-Skinner). By default it is 1
`N`	size of the population
`pij`	matrix of the second-order inclusion probabilities. This matrix is necessary for the Escobar-Berger and Campbell-Berger-Skinner options. By default it is NULL
`str`	strata ID. This vector is necessary for the Jackknife option. By default it is NULL
`clu`	cluster ID. This vector is necessary for the Jackknife option. By default it is NULL
`srswr`	variable indicating whether sampling is with replacement. By default it is NULL

Details

Functions to estimate the variance under stratified, cluster and unequal probability sampling by resampling methods (Wolter, 2007). The function ResamplingVariance allows us to choose from three models:

- The Jackknife method (Quenouille, 1949)

- The Escobar-Berger method (Escobar and Berger, 2013)

- The Campbell-Berger-Skinner method (Campbell, 1980; Berger and Skinner, 2005).

The Escobar-Berger and Campbell-Berger-Skinner methods are implemented using the functions defined in samplingVarEst package:

VE.EB.SYG.Total.Hajek, VE.EB.SYG.Mean.Hajek;

VE.Jk.CBS.SYG.Total.Hajek, VE.Jk.CBS.SYG.Mean.Hajek (see López, E., Barrios, E., 2014, for a detailed description of its use).

Note: Both methods require the matrix of the second-order inclusion probabilities. When this matrix is not an input, the program will give a warning and, by default, a jackknife method is used.

Value

The resampling variance of the randomized response technique

References

Berger, Y.G., Skinner, C.J. (2005). *A jackknife variance estimator for unequal probability sampling.* Journal of the Royal Statistical Society B, 67, 79-89.

Campbell, C. (1980). *A different view of finite population estimation.* Proceedings of the Survey Research Methods Section of the American Statistical Association, 319-324.

Escobar, E.L., Berger, Y.G. (2013). *A new replicate variance estimator for unequal probability sampling without replacement.* Canadian Journal of Statistics 41, 3, 508-524.

López, E., Barrios, E. (2014). *samplingVarEst: Sampling Variance Estimation.* R package version 0.9-9. Online http://cran.r-project.org/web/packages/survey/index.html

Quenouille, M.H. (1949). *Problems in Plane Sampling.* The Annals of Mathematical Statistics 20, 355-375.

Wolter, K.M. (2007). *Introduction to Variance Estimation.* 2nd Edition. Springer.

See Also

Warner

ChaudhuriChristofides

EichhornHayre

SoberanisCruz

Horvitz

Examples

```
N=417
data(ChaudhuriChristofidesData)
dat=with(ChaudhuriChristofidesData,data.frame(z,Pi))
mu=c(6,6)
sigma=sqrt(c(10,10))
cl=0.95
data(ChaudhuriChristofidesDatapij)
out=ChaudhuriChristofides(dat$z,mu,sigma,dat$Pi,"mean",cl,pij=ChaudhuriChristofidesDatapij)
out
ResamplingVariance(out,dat$Pi,"mean",2,N,ChaudhuriChristofidesDatapij)

#Resampling with strata
data(EichhornHayreData)
dat=with(EichhornHayreData,data.frame(ST,z,Pi))
mu=1.111111
```

```
sigma=0.5414886
cl=0.95
out=EichhornHayre(dat$z,mu,sigma,dat$Pi,"mean",cl)
out
ResamplingVariance(out,dat$Pi,"mean",1,str=dat$ST)

#Resampling with cluster
N=1500
data(SoberanisCruzData)
dat=with(SoberanisCruzData, data.frame(CL,z,Pi))
p=0.7
alpha=0.5
cl=0.90
out=SoberanisCruz(dat$z,p,alpha,dat$Pi,"total",cl)
out
ResamplingVariance(out,dat$Pi,"total",2,N,samplingVarEst::Pkl.Hajek.s(dat$Pi))

#Resampling with strata and cluster
N=1500
data(HorvitzDataStCl)
dat=with(HorvitzDataStCl, data.frame(ST,CL,z,Pi))
p=0.6
alpha=0.5
cl=0.95
out=Horvitz(dat$z,p,alpha,dat$Pi,"mean",cl,N)
out
ResamplingVariance(out,dat$Pi,"mean",3,N,samplingVarEst::Pkl.Hajek.s(dat$Pi))
```

Saha *Saha model*

Description

Computes the randomized response estimation, its variance estimation and its confidence interval through the Saha model. The function can also return the transformed variable. The Saha model was proposed by Saha in 2007.

Usage

```
Saha(z,mu,sigma,pi,type=c("total","mean"),cl,N=NULL,method="srswr")
```

Arguments

`z`	vector of the observed variable; its length is equal to n (the sample size)
`mu`	vector with the means of the scramble variables W and U
`sigma`	vector with the standard deviations of the scramble variables W and U
`pi`	vector of the first-order inclusion probabilities
`type`	the estimator type: total or mean

cl	confidence level
N	size of the population. By default it is NULL
method	method used to draw the sample: srswr or srswor. By default it is srswr

Details

In the Saha model, each respondent selected is asked to report the randomized response $z_i = W(y_i + U)$ where W, U are scramble variables whose distribution is assumed to be known.

To estimate $\bar{Y}$ a sample of respondents is selected according to simple random sampling with replacement. The transformed variable is

$$r_i = \frac{z_i - \mu_W \mu_U}{\mu_W}$$

where μ_W, μ_U are the means of W, U scramble variables respectively

The estimated variance in this model is

$$\widehat{V}(\widehat{\bar{Y}}_R) = \frac{s_z^2}{n\mu_W^2}$$

where $s_z^2 = \sum_{i=1}^n \frac{(z_i - \bar{z})^2}{n-1}$.

If the sample is selected by simple random sampling without replacement, the estimated variance is

$$\widehat{V}(\widehat{\bar{Y}}_R) = \frac{s_z^2}{n\mu_W^2}\left(1 - \frac{n}{N}\right)$$

Value

Point and confidence estimates of the sensitive characteristics using the Saha model. The transformed variable is also reported, if required.

References

Saha, A. (2007). *A simple randomized response technique in complex surveys.* Metron LXV, 59-66.

See Also

SahaData

ResamplingVariance

Examples

```
N=228
data(SahaData)
dat=with(SahaData,data.frame(z,Pi))
mu=c(1.5,5.5)
sigma=sqrt(c(1/12,81/12))
cl=0.95
Saha(dat$z,mu,sigma,dat$Pi,"mean",cl,N)
```

SahaData *Randomized Response Survey on spending on alcohol*

Description

This data set contains observations from a randomized response survey conducted in a population of students to investigate spending on alcohol. The sample is drawn by simple random sampling with replacement. The randomized response technique used is the Saha model (Saha, 2007) with scramble variables $W = U(1, 2)$ and $U = U(1, 10)$.

Usage

```
SahaData
```

Format

A data frame containing 100 observations. The variables are:

- ID: Survey ID
- z: The randomized response to the queston: How much money did you spend on alcohol, last weekend?
- Pi: first-order inclusion probabilities

References

Saha, A. (2007). *A simple randomized response technique in complex surveys.* Metron LXV, 59-66.

See Also

Saha

Examples

```
data(SahaData)
```

SinghJoarder *Singh-Joarder model*

Description

Computes the randomized response estimation, its variance estimation and its confidence interval through the Singh-Joarder model. The function can also return the transformed variable. The Singh-Joarder model was proposed by Singh and Joarder in 1997.

Usage

```
SinghJoarder(z,p,pi,type=c("total","mean"),cl,N=NULL,pij=NULL)
```

Arguments

z	vector of the observed variable; its length is equal to n (the sample size)
p	proportion of marked cards with the sensitive question
pi	vector of the first-order inclusion probabilities
type	the estimator type: total or mean
cl	confidence level
N	size of the population. By default it is NULL
pij	matrix of the second-order inclusion probabilities. By default it is NULL

Details

The basics of the Singh-Joarder scheme are similar to Warner's randomized response device, with the following difference. If a person labelled i bears A^c he/she is told to say so if so guided by a card drawn from a box of A and A^c marked cards in proportions p and $(1-p), (0 < p < 1)$. However, if he/she bears A and is directed by the card to admit it, he/she is told to postpone the reporting based on the first draw of the card from the box but to report on the basis of a second draw. Therefore,

$$z_i = \begin{cases} 1 & \text{if person } i \text{ responds "Yes"} \\ 0 & \text{if person } i \text{ responds "No"} \end{cases}$$

The transformed variable is $r_i = \frac{z_i-(1-p)}{(2p-1)+p(1-p)}$ and the estimated variance is $\widehat{V}_R(r_i) = r_i(r_i-1)$.

Value

Point and confidence estimates of the sensitive characteristics using the Singh-Joarder model. The transformed variable is also reported, if required.

References

Singh, S., Joarder, A.H. (1997). *Unknown repeated trials in randomized response sampling.* Journal of the Indian Statistical Association, 30, 109-122.

See Also

SinghJoarderData

ResamplingVariance

Examples

```
N=802
data(SinghJoarderData)
dat=with(SinghJoarderData,data.frame(z,Pi))
p=0.6
cl=0.95
SinghJoarder(dat$z,p,dat$Pi,"mean",cl,N)
```

`SinghJoarderData` *Randomized Response Survey on compulsive spending*

Description

This data set contains observations from a randomized response survey conducted in a university to investigate compulsive spending. The sample is drawn by simple random sampling without replacement. The randomized response technique used is the Singh-Joarder model (Singh and Joarder, 1997) with parameter $p = 0.6$.

Usage

```
SinghJoarderData
```

Format

A data frame containing 170 observations from a population of $N = 802$ students. The variables are:

- ID: Survey ID of student respondent
- z: The randomized response to the question: Do you have spend compulsively?
- Pi: first-order inclusion probabilities

References

Singh, S., Joarder, A.H. (1997). *Unknown repeated trials in randomized response sampling.* Journal of the Indian Statistical Association, 30, 109-122.

See Also

`SinghJoarder`

Examples

```
data(SinghJoarderData)
```

SoberanisCruz *SoberanisCruz model*

Description

Computes the randomized response estimation, its variance estimation and its confidence interval through the SoberanisCruz model. The function can also return the transformed variable. The SoberanisCruz model was proposed by Soberanis Cruz et al. in 2008.

Usage

```
SoberanisCruz(z,p,alpha,pi,type=c("total","mean"),cl,N=NULL,pij=NULL)
```

Arguments

z	vector of the observed variable; its length is equal to n (the sample size)
p	proportion of marked cards with the sensitive question
alpha	proportion of people with the innocuous attribute
pi	vector of the first-order inclusion probabilites
type	the estimator type: total or mean
cl	confidence leve
N	size of the population. By default it is NULL
pij	matrix of the second-order inclusion probabilities. By default it is NULL

Details

The SoberanisCruz model considers the introduction of an innocuous variable correlated with the sensitive variable. This variable does not affect individual sensitivity, and maintains reliability. The sampling procedure is the same as in the Horvitz model.

Value

Point and confidence estimates of the sensitive characteristics using the SoberanisCruz model. The transformed variable is also reported, if required.

References

Soberanis Cruz, V., Ramírez Valverde, G., Pérez Elizalde, S., González Cossio, F. (2008). *Muestreo de respuestas aleatorizadas en poblaciones finitas: Un enfoque unificador.* Agrociencia Vol. 42 Núm. 5 537-549.

See Also

SoberanisCruzData

Horvitz

ResamplingVariance

Examples

```
data(SoberanisCruzData)
dat=with(SoberanisCruzData,data.frame(z,Pi))
p=0.7
alpha=0.5
cl=0.90
SoberanisCruz(dat$z,p,alpha,dat$Pi,"total",cl)
```

SoberanisCruzData   *Randomized Response Survey on speeding*

Description

This data set contains observations from a randomized response survey conducted in a population of 1500 families in a Spanish town to investigate speeding. The sample is drawn by cluster sampling by district. The randomized response technique used is the SoberanisCruz model (Soberanis Cruz et al., 2008) with parameter $p = 0.7$. The innocuous question is: Is your car medium/high quality? with $\alpha = 0.5$.

Usage

```
SoberanisCruzData
```

Format

A data frame containing 290 observations from a population of $N = 1500$ families divided into twenty cluster. The variables are:

- ID: Survey ID
- CL: Cluster ID
- z: The randomized response to the question: Do you often break the speed limit?
- Pi: first-order inclusion probabilities

References

Soberanis Cruz, V., Ramírez Valverde, G., Pérez Elizalde, S., González Cossio, F. (2008). *Muestreo de respuestas aleatorizadas en poblaciones finitas: Un enfoque unificador.* Agrociencia Vol. 42 Núm. 5 537-549.

See Also

SoberanisCruz

Examples

```
data(SoberanisCruzData)
```

`Warner` *Warner model*

Description

Computes the randomized response estimation, its variance estimation and its confidence interval through the Warner model. The function can also return the transformed variable. The Warner model was proposed by Warner in 1965.

Usage

```
Warner(z,p,pi,type=c("total","mean"),cl,N=NULL,pij=NULL)
```

Arguments

`z`	vector of the observed variable; its length is equal to n (the sample size)
`p`	proportion of marked cards with the sensitive attribute
`pi`	vector of the first-order inclusion probabilities
`type`	the estimator type: total or mean
`cl`	confidence level
`N`	size of the population. By default it is NULL
`pij`	matrix of the second-order inclusion probabilities. By default it is NULL

Details

Warner's randomized response device works as follows. A sampled person labelled i is offered a box of a considerable number of identical cards with a proportion $p, (0 < p < 1, p \neq 0.5)$ of them marked A and the rest marked A^c. The person is requested, randomly, to draw one of them, to observe the mark on the card, and to give the response

$$z_i = \begin{cases} 1 & \text{if card type "matches" the trait } A \text{ or } A^c \\ 0 & \text{if a "no match" results} \end{cases}$$

The randomized response is given by $r_i = \frac{z_i-(1-p)}{2p-1}$ and the estimated variance is $\widehat{V}_R(r_i) = r_i(r_i - 1)$.

Value

Point and confidence estimates of the sensitive characteristics using the Warner model. The transformed variable is also reported, if required.

References

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

See Also

WarnerData

ResamplingVariance

Examples

```
N=802
data(WarnerData)
dat=with(WarnerData,data.frame(z,Pi))
p=0.7
cl=0.95
Warner(dat$z,p,dat$Pi,"total",cl)
```

WarnerData *Randomized Response Survey on alcohol abuse*

Description

This data set contains observations from a randomized response survey related to alcohol abuse. The sample is drawn by simple random sampling without replacement. The randomized response technique used is the Warner model (Warner, 1965) with parameter $p = 0.7$.

Usage

```
WarnerData
```

Format

A data frame containing 125 observations from a population of $N = 802$ students. The variables are:

- ID: Survey ID of student respondent
- z: The randomized response to the question: During the last month, did you ever have more than five drinks (beer/wine) in succession?
- Pi: first-order inclusion probabilities

References

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

See Also

Warner

Examples

```
data(WarnerData)
```

Index

Appendix A4

Advances in estimation by the item sum technique using auxiliary information in complex surveys.

Rueda, María del Mar; Perri, Pier Francesco; Cobo, Beatriz. (2017)
Advances in estimation by the item sum technique using auxiliary information in complex surveys.
Advances in Statistical Analysis
DOI: 10.1007/s10182-017-0315-2

STATISTICS & PROBABILITY			
JCR Year	Impact factor	Rank	Quartile
2016	0.868	71/124	Q3

Abstract

To collect sensitive data, survey statisticians have designed many strategies to reduce nonresponse rates and social desirability response bias. In recent years, the item count technique (ICT) has gained considerable popularity and credibility as an alternative mode of indirect questioning survey, and several variants of this technique have been proposed as new needs and challenges arise. The item sum technique (IST), which was introduced by Chaudhuri and Christofides (Indirect questioning in sample surveys, Springer-Verlag, Berlin, 2013) and Trappmann et al. (J Surv Stat Methodol 2:58-7, 2014), is one such variant, used to estimate the mean of a sensitive quantitative variable. In this approach, sampled units are asked to respond to a two-list of items containing a sensitive question related to the study variable and various innocuous, nonsensitive, questions. To the best of our knowledge, very few theoretical and applied papers have addressed the IST. In this article, therefore, we present certain methodological advances as a contribution to appraising the use of the IST in real-world surveys. In particular, we employ a generic sampling design to examine the problem of how to improve the estimates of the sensitive mean when auxiliary information on the population under study is available and is used at the design and estimation stages. A Horvitz-Thompson-type estimator and a calibration-type estimator are proposed and their efficiency is evaluated by means of an extensive simulation study. Using simulation experiments, we show that estimates obtained by the IST are nearly equivalent to those obtained using "true data" and that in general they outperform the estimates provided by a competitive randomized response method. Moreover, the variance estimation may be considered satisfactory. These results open up new perspectives for academics, researchers and survey practitioners, and could justify the use of the IST as a valid alternative to traditional direct questioning survey modes.

1 Introduction

In many fields of applied research, and particularly in sociological, economic, demographic, ecological and medical studies, the investigator very often has to gather information concerning highly personal, sensitive, stigmatizing and perhaps incriminating issues such as abortion, drug addiction, HIV/AIDS infection status, duration of suffering from a disease, sexual behavior, domestic violence, racial prejudice or noncompliance with laws and regulations. In these situations, collecting data by means of survey modes based on direct questioning (DQ) methods of interview is likely to encounter two serious problems: (i) participants in the survey may deliberately release untruthful or misleading answers, or (ii) participants may refuse to respond ("unit nonresponse" or "item nonresponse") due to the social stigma or because they feel threatened by such inquiries and fear that their personal information may be released to third parties for purposes other than those of the survey. Misleading information and refusal to answer are nonsampling errors that are difficult to deal with and can seriously flaw the validity of final analyses. To reduce this problem, the level of cooperation obtained from the respondent must be increased. Since the decision to cooperate, in terms of providing complete and honest answers, depends on how interviewees perceive their privacy will be protected, survey modes which ensure full anonymity go some way to increasing the probability of cooperation and, consequently, that of obtaining more reliable information on sensitive topics. In this respect, survey statisticians and practitioners have developed many different strategies to ensure interviewees' anonymity and to reduce the incidence of evasive answers and underreporting of social taboos when direct questions are posed on sensitive issues. One possibility is to limit the influence of the interviewer, by providing self-administered questionnaires, enabling computer-assisted self-interviews or by conducting online surveys. Alternatively, the randomized response (RR) theory (RRT), conceived by Warner (1965), may be employed. In its original version, this nonstandard survey approach adopts a randomization device such as a deck of cards, dice, coins, coloured numbered balls, spinners or even a computer to conceal the true answer, in the sense that respondents reply to one of two or more selected questions depending on the result of the device. Specifically, the randomization device 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 stigmatizing behavior. The randomization device generates a probabilistic relation between the sensitive question and a given answer which is used to draw inference about unknown parameters of interest, for instance the prevalence of a sensitive attribute in the target population. The rationale of the RRT is that interviewees 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 both to the interviewer and to the researcher and, consequently, respondents' privacy is preserved.

Since Warner's pioneering work, a large number of RR mechanisms have been considered, with continual innovations of existing devices as well as novel proposals. Such procedures have been amply discussed, for example, by Fox and Tracy (1986), Chaudhuri and Mukerjee (1988), Chaudhuri (2011) and Chaudhuri and Christofides (2013). Contextually, many studies have assessed the validity of RR methods, showing that they can produce more reliable answers than conventional data collection methods (e.g., DQ in face-to-face interviews, self-administered questionnaires with paper and pencil and computer-assisted self interviews). In this respect, see van der Heijden et al. (2000), Lara et al. (2004) and Lensvelt-Mulders et al. (2005), to name just a few. Finally, let us note that considerable use is made of the RRT and its variants in real-life studies of a great variety of topics including, for instance, the use of drug, athletic and cognitive performance-enhancing substances (Goodstadt and Gruson, 1975; Kerkvliet, 1994; Simon et al., 2006; Striegel et al., 2010; James et al., 2013; Stubbe et al., 2013; Dietz et al., 2013; Shamsipour et al., 2014), the estimation of the prevalence of fraud in the area of disability benefits (van der Heijden et al., 2000; Lensvelt-Mulders et al., 2006), racial prejudice in Germany (Ostapczuk et al., 2009; Krumpal, 2012), 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 et al., 2006; Oliveras and Letamo, 2010), voting turnout (Holbrook and Krosnick, 2010a), tax evasion (Houston and Tran, 2001; Korndörfer et al., 2014), plagiarism in Swiss and German student papers (Jann et al., 2012), induced abortion and irregular immigrant status among foreign women in Calabria (Perri et al., 2016) and the illegal use of natural resources (Chaloupka, 1985; Schill and Kline, 1995; Solomon et al., 2007; Blank and Gavin, 2009; Arias and Sutton, 2013; Conteh et al. 2015).

Despite the good reputation that the RRT has acquired over time as a tool to obtain reliable data while protecting respondents' confidentiality, avoiding unacceptable rates of nonresponse and reducing social desirability response bias, the approach, at least in its basic idea, suffers from some inadequacies that have limited its complete acceptance among survey statisticians and practitioners. The main limitations may be summarized in the following points: (i) RRT surveys are, in general, more time-demanding and costly than other types of survey modes; (ii) RRT estimates are subject to greater sampling variance (i.e., lower efficiency) than DQ estimates. This loss of efficiency represents the cost of obtaining more reliable information by reducing response bias. Consequently, achieving estimates which are comparably efficient with those obtained under DQ may require a considerably larger sample with the consequent increase in cost, an aspect which is rarely acceptable; (iii) RRT surveys lack reproducibility, in the sense that the same respondent may give different information if asked to repeat the survey. This is because his/her answer depends on the outcome of the randomization device. Hence, conditioned to a selected sample of respondents, the estimation process may yield different estimates according to the outcome of the device; (iv) lack of understanding and trust among respondents. Chaudhuri and Christofides (2007) observed that the RRT basically asks respondents to provide information that may seem useless or even deceitful. When the respondent does not understand the mathematical logic underlying the technique, then the entire procedure may be suspect, leading the respondent to believe there might be a way for the interviewer to determine his/her exact status regarding the sensitive characteristic by processing the response provided. Moreover, respondents may not understand the instructions for using the RR device and/or not trust the privacy protection offered. Hence, they might intentionally refuse to participate in the survey or break the rules of the RR design; (v) RR procedures require a randomization device to drive the answer. In Warner's original model, the suggested device was a spinner but any other physical device, like dice, a deck of cards or coloured numbered balls, could be used. Using physical devices limits the application of the RRT exclusively to face-to-face personal interviews and may also be more time consuming (the procedure must be explained to each survey participant) and costly (the devices must be obtained) than DQ. Other means of survey administration, such as telephone interview, self-administered mail questionnaire and internet-delivered interviews, seem to be precluded. In addition, respondents could find it difficult to use a physical device, for instance due to reduced motor capacity, or be suspicious of using something provided by the interviewer.

Mindful of these drawbacks, alternative indirect questioning techniques have been proposed which overcome some of the limitations affecting the RRT and enable sensitive information to be acquired while preserving respondents' confidentiality. Such alternative methods are encompassed in different approaches which include the nominative technique (Miller, 1985), the three card method (Droitcour et al., 2002), the non-randomized response technique (Tian and Tang, 2014) and the item count technique (hereafter ICT; Raghavarao and Federer, 1979; Miller, 1984; Droitcour et al., 1991). All of these alternatives were originally conceived for surveys requiring a "yes" or "no" response to a sensitive question, or a choice of responses from a set of nominal categories, and do not address quantitative sensitive characteristics. Recently, Chaudhuri and Christofides (2013) and Trappmann et al. (2014) have proposed a generalization of the ICT that can be used to survey a quantitative sensitive characteristic and to estimate its mean. This variant of the ICT is called the item sum technique (hereafter IST) and is the focus of the present article, which has a twofold aim: (i) to provide a general framework for the IST by extending the results of Chaudhuri and Christofides (2013) and Trappmann et al. (2014) from simple random sampling to a generic complex sampling design; (ii) to investigate the effectiveness of employing auxiliary information to improve, without incurring additional costs or increasing the sample size, the efficiency of estimates when the IST is used to obtain data from a complex survey. The first of these study aims is motivated by the fact that real surveys are customarily conducted by using complex sampling designs such as stratified and/or cluster sampling, with units selected according to a specific varying probability scheme. The second concerns the fact that, in sampling practice, DQ techniques for collecting information about nonsensitive characteristics make use of auxiliary variables to improve sampling designs and to achieve higher precision in the estimates of unknown population parameters. Nevertheless, and although a number of proposals have been made to improve the estimation of the population proportion and the population mean of sensitive variables in the RRT (see, among others, Diana and Perri 2009, 2010, 2011, 2012; Perri and Diana 2013), very few such procedures have been suggested to improve the performance of the IST. To the best of our knowledge, there is only the paper by Trappmann et al. (2014) who outlined a procedure to estimate regression models for the IST, and that of Hussain et al. (2017) who discussed ratio, product and regression methods. Hence, we seek to fill this gap, giving prominence to the use of auxiliary information.

The rest of this article is organized as follows. Sections 2 and 3 describe the ICT and the IST, respectively. In Section 4, we discuss methodological advances for IST estimation under a generic sampling design. Specifically, a Horvitz-Thompson-type estimator is examined in Section 4.1, a calibration-type estimator is proposed in Section 4.2, and in Section 4.3 the calibration approach is employed for domain estimation. The results of various simulation experiments are presented and commented on in Section 5. In particular, Section 5.1 includes: (i) a numerical comparison of DQ, IST and RR estimates under three sampling designs; (ii) an analysis of the effect on the Horvitz-Thompson and calibration-type estimators caused by the presence of a different correlation coefficient between the target variable and the innocuous variable; (iii) an analysis of the performance of the Horvitz-Thompson and calibration-type estimators for the domain of interest. Section 5.2 is then devoted to an analysis of just the IST calibration estimators, investigating their performance when the number of nonsensitive variables used in the IST design is increased. The accuracy of the variance estimator is also investigated. Section 6 concludes the paper with some final remarks.

2 The item count technique

Assume that the researcher wishes to use the ICT to determine the prevalence of a sensitive attribute A in a population, for instance the amount of work performed and not declared to the tax authorities. The ICT (also known as "the unmatched count technique", "block total response" or "list experiment") was originally conceived by Raghavarao and Federer (1979) and Miller (1984), and consists of drawing two independent samples from the target population, say s_1 and s_2. Without loss of generality, units belonging to sample s_1 are provided with a long list (LL) of items containing $(G + 1)$ dichotomous questions, of which G are nonsensitive, while the remaining one refers to the sensitive attribute A. The sampled units are instructed to consider the LL, and to count and report the number of items that apply to them (i.e., the number of "yes" responses) without answering each question individually. Consequently, respondents' privacy is protected since their true sensitive status remains undisclosed unless they report that none or all of the items in the list apply to them. By contrast, units belonging to sample s_2 are asked to make a similar response to a short list (SL) of items, containing only the G innocuous questions which are identical to those present in the LL. The innocuous items should be chosen and worded in sufficient quantity as to ensure the necessary variability in their application to the units in the population.

The answers given by samples s_1 and s_2 are then pooled to obtain an estimate of the prevalence π_A of units bearing the sensitive attribute A. An unbiased estimator of π_A is termed the difference-in-means estimator, and is obtained as the difference between the means of the answers in sample s_1 and in sample s_2:

$$\hat{\pi}_A = \hat{\mu}_1 - \hat{\mu}_2. \tag{1}$$

Following Miller (1984), the body of research literature on the subject expanded rapidly, discussing alternative techniques and item count schemes to increase the efficiency of the estimator of π_A and to overcome some shortcomings of the original version. For instance, Chaudhuri and Christofides (2007) proposed a modification of the method aimed at protecting against a possible "negative" value for the estimate, which might arise from (1), and at increasing privacy protection should all or none of the $(G+1)$ items be applicable to a respondent in sample s_1. The revised ICT requires that an innocuous characteristic B, unrelated to the sensitive one and possessed by a known proportion π_B of the population, be considered. Then, units in sample s_1 are presented with a list of $(G + 1)$ items of which the first G are innocuous and the $(G + 1)$st item stands for "I have the characteristic A, or B or both". Similarly, units in the second sample s_2 are given a list of $(G + 1)$ items, of which the first G items are exactly the same as those in sample s_1 while the $(G + 1)$st item stands for "I do not have either characteristic A or B". Using the same notation as in the original ICT, an unbiased estimator of π_A is obtained as

$\hat{\pi}_A = \hat{\mu}_1 - \hat{\mu}_2 + 1 - \pi_B$. Under this variant, privacy protection is guaranteed, provided that at least one of the innocuous items applies. In order to overcome this minimum requirement, Christofides (2015) presented a new version of the ICT in which respondents' privacy is fully protected since no answer reveals whether the sensitive attribute is possessed. Subsequently, Shaw (2016) revised Chaudhuri and Christofides' method (2007) and proposed a procedure based on a single sample. Other attempts to improve the ICT and thus contribute to its growing use among survey practitioners have been made, among others, by the following: Droitcour et al. (1991) proposed a design in which π_A is estimated by using two-list experiment applied to the same units in such a way as to reduce sampling variability; Glynn (2013) suggested an adjustment to the estimator given in (1) which yields greater efficiency, although at the cost of greater bias; Blair and Imai (2010) introduced the *list* R package to conduct statistical analysis for the ICT, implementing the methods described by Imai (2011), Blair and Imai (2012), Blair et al. (2014), Imai et al. (2015) Aronow et al. (2015) and Hussain et al. (2012) provided the variance expression of the estimator $\hat{\pi}_A$ under simple random sampling and suggested an improved ICT that does not require two samples; Aronow et al. (2015) proposed a method to combine ICT and DQ estimates; Holbrook and Krosnick (2010b), in order to compare direct and list experiment estimates within the same target population in a real-world study, randomly split the selected sample into three groups: the first received the SL, the second received the LL and the third was surveyed only by DQ, with no list at all; Chaudhuri and Christofides (2013) discussed a three-sample procedure, extending the variant suggested by Chaudhuri and Christofides (2007).

3 The item sum technique

Standard item count methods are primarily used in surveys which require a binary response to a sensitive question, and seek to estimate the proportion of people bearing a given sensitive characteristic. Nevertheless, in practice many situations may be encountered in which the response to a sensitive question results in a quantitative variable. For instance, sensitive questions may refer to the number of extramarital relationships, the amount of personal income or wealth, the number of times income taxes are evaded, and so on. For situations like these, Chaudhuri and Christofides (2013) presented a variant of the ICT, suitable for quantitative sensitive characteristics, that Trappmann et al. (2014) termed the item sum technique (IST) and used in a CATI survey on undeclared work in Germany. The IST works in a similar way to the ICT. Two independent simple random samples are drawn from the population. Units belonging to one of the two samples are presented with the LL of items containing the sensitive question and a number of nonsensitive questions; units in the other sample receive only the SL of items consisting of the nonsensitive questions. All of the items refer to quantitative variables, possibly measured on the same scale as that of the sensitive variable. Respondents are then asked to report the total score of their answers to all of the questions in their list, without revealing the individual score for each question. Like the ICT, the mean difference of the answers between the LL-sample and the SL-sample is then used as an unbiased estimator of the population mean of the sensitive variable.

Hussain et al. (2017) proposed a one-sample variant of the IST, in which each of the units in the simple random sample is provided with a list of items and just one of these items contains queries about stigmatizing and non-stigmatizing variables. These authors also considered ratio, product and regression estimators to incorporate auxiliary information into the IST estimation procedure. The one-sample approach to the IST has also been considered by Shaw (2015).

To the best of our knowledge, to date there have been no other contributions regarding the IST. Motivated by this perceived research gap, and seeking to contribute to the development of the IST in real-world studies, we suggest some methodological advances based on the use of auxiliary information at both the design and the estimation stages. Specifically, we introduce a general framework for estimating the population mean of a sensitive quantitative variable by assuming that the samples are randomly obtained under a generic sampling design. Hence, we discuss the use of the calibration technique to improve the efficiency of the estimates and then extend this calibration approach to the estimation of domains. In addition, we discuss variance estimation and the impact on the estimates of including an increased number of innocuous questions in the list of items. Part of the discussion is based on an extensive simulation study.

4 Advances in IST estimation

4.1 Estimation under a generic sampling design: the Horvitz-Thompson-type estimator

Consider a finite population $U = \{1, \ldots, N\}$ consisting of N different and identifiable units. Let y_i be the value of the sensitive character under study, say y, for the ith population unit. Our aim is to estimate the population mean $\bar{Y} = N^{-1} \sum_{i \in U} y_i$.

Let us assume a generic sampling design $p(\cdot)$ with positive first- and second-order inclusion probabilities $\pi_i = \sum_{s \ni i} p(s)$ and $\pi_{ij} = \sum_{s \ni i,j} p(s), i, j \in U$. Let $d_i = \pi_i^{-1}$ denote the known sampling design-basic weight for unit $i \in U$, and $\mathbb{E}_p$ and $\mathbb{V}_p$ the operators expectation and variance under the sampling design $p(\cdot)$. Two independent samples, s_1 and s_2, are selected from U according to the design $p(\cdot)$. One of the samples, say s_1, is confronted with a LL of items containing $(G+1)$ questions of which G refer to nonsensitive characteristics and one is related to the sensitive characteristic under study. The other sample s_2 receives a SL of items that only contains the G innocuous questions. All sensitive and nonsensitive items are quantitative in nature. Respondents in both samples are requested to report the total score of all the items applicable to them, without revealing the individual score on each of the items. Without loss of generality, let t be the variable denoting the total score applicable to the G nonsensitive questions, and $z = y + t$ the total score applicable to the nonsensitive questions and the sensitive question. Hence, the answer of the ith respondent will be $z_i = y_i + t_i$ if $i \in s_1$ or t_i if $i \in s_2$. We observe that for $G = 1$, the variable t simply denotes the innocuous variable and t_i its value on the ith unit.

Under the design $p(\cdot)$, let

$$\hat{\bar{Z}}_{HT} = \frac{1}{N}\sum_{i \in s_1} d_i z_i \quad \text{and} \quad \hat{\bar{T}}_{HT} = \frac{1}{N}\sum_{i \in s_2} d_i t_i$$

be the unbiased Horvitz-Thompson (hereafter HT; Horvitz and Thompson, 1952) 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 $\bar{Y}$ can be immediately obtained as:

$$\hat{\bar{Y}}_{HT} = \hat{\bar{Z}}_{HT} - \hat{\bar{T}}_{HT}. \tag{2}$$

From the unbiasedness of $\hat{\bar{Z}}_{HT}$ and $\hat{\bar{T}}_{HT}$, it readily follows that the estimator $\hat{\bar{Y}}_{HT}$ is unbiased for $\bar{Y}$. In fact

$$\begin{aligned}
\mathbb{E}_p(\hat{\bar{Y}}_{HT}) &= \mathbb{E}_p(\hat{\bar{Z}}_{HT}) - \mathbb{E}_p(\hat{\bar{T}}_{HT}) = \frac{1}{N}\sum_{i \in U} z_i - \frac{1}{N}\sum_{i \in U} t_i \\
&= \frac{1}{N}\sum_{i \in U}(z_i - t_i) = \frac{1}{N}\sum_{i \in U} y_i.
\end{aligned}$$

The variance of $\hat{\bar{Y}}_{HT}$, as long as the two samples s_1 and s_2 are independent, can be expressed as:

$$\begin{aligned}
\mathbb{V}_p(\hat{\bar{Y}}_{HT}) &= \mathbb{V}_p(\hat{\bar{Z}}_{HT}) + \mathbb{V}_p(\hat{\bar{T}}_{HT}) \\
&= \frac{1}{N^2}\left(\sum_{i \in U}\sum_{j \in U} \Delta_{ij}(d_i z_i)(d_j z_j) + \sum_{i \in U}\sum_{j \in U} \Delta_{ij}(d_i t_i)(d_j t_j)\right),
\end{aligned}$$

where $\Delta_{ij} = \pi_{ij} - \pi_i \pi_j$. Finally, an unbiased estimator of $\mathbb{V}(\hat{\bar{Y}}_{HT})$ is achieved by means of

$$\hat{\mathbb{V}}_p(\hat{\bar{Y}}_{HT}) = \frac{1}{N^2}\left(\sum_{i \in s_1}\sum_{j \in s_1} \check{\Delta}_{ij}(d_i z_i)(d_j z_j) + \sum_{i \in s_2}\sum_{j \in s_2} \check{\Delta}_{ij}(d_i t_i)(d_j t_j)\right),$$

where $\check{\Delta}_{ij} = \Delta_{ij}/\pi_{ij}$.

4.2 Estimation in the presence of auxiliary information: the calibration-type estimator

The growing availability of population information derived from census data, administrative registers and previous surveys provides a wide range of variables that can be used to increase the efficiency of the estimation procedure. In this respect, a useful approach is that calibration by which new sampling weights are constructed to match benchmark constraints on auxiliary variables while remaining "close" to the design-basic weights (Deville and Särndal, 1992). Särndal (2007) provides an overview of several developments in calibration estimation, showing that this tool can be used to combine and/or align estimates from different surveys. Calibration is also widely used as a tool to reduce nonresponse and coverage error. This aspect has been discussed at length by Särndal and Lundström (2005), and further explored by Kott and Chang (2010) and, more recently, by Kott (2014).

Let us now discuss how calibration estimation may be extended to address IST surveys. In so doing, we assume that a vector $\mathbf{x}$ of Q auxiliary variables is available from different sources such that the vector of values $\mathbf{x}_i = (x_{i1}, \ldots, x_{iQ})^t$ is known $\forall i \in U$. Additionally, let $\bar{\mathbf{X}} = N^{-1} \sum_{i \in U} \mathbf{x}_i$ denote the vector for the known population means of the Q auxiliary variables. Our goal is to estimate the population mean $\bar{Y}$ by using the observations of the variables z, t and $\mathbf{x}$ in the samples s_1 and s_2, and the known vector values $\bar{\mathbf{X}}$ in the population. In order to obtain a calibration estimator of $\bar{Y}$ in the IST setting, we follow Deville and Särndal (1992) to obtain a new system of weights ω_{ij} based on sample s_j, $j = 1, 2$, by minimizing the chi-squared distance function

$$\Phi_{s_j}(d_i, \omega_{ij}) = \sum_{i \in s_j} \frac{(\omega_{ij} - d_i)^2}{d_i q_i}, \quad j = 1, 2 \tag{3}$$

subject to the calibration equations

$$\frac{1}{N} \sum_{i \in s_j} \omega_{ij} \mathbf{x}_i = \bar{\mathbf{X}}, \tag{4}$$

where the q_i's are known positive constants unrelated to the d_i's. Minimization of (3) under (4) then yields the weights ω_{ij} given by:

$$\omega_{ij} = d_i + \frac{d_i q_i \boldsymbol{\lambda}^t \mathbf{x}_i}{N}, \quad j = 1, 2 \tag{5}$$

where $\boldsymbol{\lambda} = (\lambda_1 \ldots, \lambda_Q)^t$ is the vector of the Lagrange multipliers given by:

$$\boldsymbol{\lambda} = N^2 \mathbf{F}_{s_j}^{-1} (\bar{\mathbf{X}} - \hat{\bar{\mathbf{X}}}_{HT}),$$

with $\mathbf{F}_{s_j} = \sum_{i \in s_j} d_i q_i \mathbf{x}_i \mathbf{x}_i^t$ and where $\hat{\bar{\mathbf{X}}}_{HT}$ denotes the vector of the HT estimators of the population means $\bar{\mathbf{X}}$ based on the sample s_j.

According to the calibration weights obtained from (5), we define a calibration-type estimator of $\bar{Y}$ as:

$$\hat{\bar{Y}}_C = \hat{\bar{Z}}_C - \hat{\bar{T}}_C, \tag{6}$$

where

$$\hat{\bar{Z}}_C = \frac{1}{N}\sum_{i\in s_1}\omega_{i1}z_i = \hat{\bar{Z}}_{HT} + (\bar{\mathbf{X}} - \hat{\bar{\mathbf{X}}}_{HT})^t\hat{\mathbf{B}}_{s_1}$$

is the calibration estimator of $\bar{Z}$ obtained on the basis of the LL-sample s_1, with $\hat{\mathbf{B}}_{s_1} = \mathbf{F}_{s_1}^{-1}\sum_{i\in s_1} d_i q_i \mathbf{x}_i z_i$, and

$$\hat{\bar{T}}_{\mathrm{C}} = \frac{1}{N}\sum_{i\in s_2}\omega_{i2}t_i = \hat{\bar{T}}_{HT} + (\bar{\mathbf{X}} - \hat{\bar{\mathbf{X}}}_{HT})^t\hat{\mathbf{B}}_{s_2}$$

is the calibration estimator of $\bar{T}$ obtained from the SL-sample s_2, with $\hat{\mathbf{B}}_{s_2} = \mathbf{F}_{s_2}^{-1}\sum_{i\in s_2} d_i q_i \mathbf{x}_i t_i$.

Following Deville and Särndal (1992), it can be shown that the estimator $\hat{\bar{Y}}_C$ is asymptotically unbiased for $\bar{Y}$ and its asymptotic variance is given by:

$$\begin{aligned}\mathbb{V}_p(\hat{\bar{Y}}_C) &= \mathbb{V}_p(\hat{\bar{Z}}_C) + V_p(\hat{\bar{T}}_C)\\ &= \frac{1}{N^2}\left(\sum_{i\in U}\sum_{j\in U}\Delta_{ij}(d_iE_i)(d_jE_j) + \sum_{i\in U}\sum_{j\in U}\Delta_{ij}(d_iG_i)(d_jG_j)\right),\end{aligned}$$

where

$$E_i = z_i - \mathbf{x}_i^t\mathbf{B}_1 \quad \text{with} \quad \mathbf{B}_1 = \left(\sum_{i\in U} q_i\mathbf{x}_i\mathbf{x}_i^t\right)^{-1}\sum_{i\in U} q_i\mathbf{x}_i z_i$$

and

$$G_i = t_i - \mathbf{x}_i^t\mathbf{B}_2 \quad \text{with} \quad \mathbf{B}_2 = \left(\sum_{i\in U} q_i\mathbf{x}_i\mathbf{x}_i^t\right)^{-1}\sum_{i\in U} q_i\mathbf{x}_i t_i.$$

An estimator for this variance is:

$$\hat{\mathbb{V}}_p(\hat{\bar{Y}}_C) = \frac{1}{N^2}\left(\sum_{i\in s_1}\sum_{j\in s_1}\check{\Delta}_{ij}(d_ie_i)(d_je_j) + \sum_{i\in s_2}\sum_{j\in s_2}\check{\Delta}_{ij}(d_ig_i)(d_jg_j)\right), \tag{7}$$

where

$$e_i = z_i - \mathbf{x}_i^t\hat{\mathbf{B}}_{s_1} \quad \text{and} \quad g_i = t_i - \mathbf{x}_i^t\hat{\mathbf{B}}_{s_2}.$$

4.3 Estimation for domains

As in Section 4.1, let U denote the target population from which two samples, s_1 and s_2, are drawn according to the sampling design $p(\cdot)$. Let $U_d \subset U$ denote a domain of interest of N_d units, δ_{di} the domain identifier taking the value 1 if $i \in U_d$, and s_{jd} the subset of s_j containing units from U_d, $s_{jd} = s_j \cap U_d$, with $j = 1, 2$. It is straightforwardly determined that the sizes of s_{1d} and s_{2d} are random variables.

In order to obtain an estimate of the domain mean $\bar{Y}_d = N_d^{-1} \sum_{i \in U_d} y_i$, let us first consider, following (2), the HT-type estimator defined as:

$$\hat{\hat{Y}}_{\text{HT},d} = \frac{1}{N_d} \sum_{i \in s_{1d}} d_i z_i - \frac{1}{N_d} \sum_{i \in s_{2d}} d_i t_i.$$

The estimator $\hat{\hat{Y}}_{HT,d}$ is design-unbiased. In fact,

$$\begin{aligned}
\mathbb{E}_p(\hat{\hat{Y}}_{HT,d}) &= \frac{1}{N_d}\mathbb{E}_p\left(\sum_{i \in s_{1d}} d_i z_i\right) - \frac{1}{N_d}\mathbb{E}_p\left(\sum_{i \in s_{2d}} d_i t_i\right) \\
&= \frac{1}{N_d}\mathbb{E}_p\left(\sum_{i \in s_1} d_i z_i \delta_{di}\right) - \frac{1}{N_d}\mathbb{E}_p\left(\sum_{i \in s_2} d_i t_i \delta_{di}\right) \\
&= \frac{1}{N_d}\sum_U z_i \delta_{di} - \frac{1}{N_d}\sum_U t_i \delta_{di} \\
&= \frac{1}{N_d}\sum_{i \in U_d} z_i - \frac{1}{N_d}\sum_{i \in U_d} t_i \\
&= \frac{1}{N_d}\sum_{i \in U_d} (z_i - t_i) = \frac{1}{N_d}\sum_{i \in U_d} y_i.
\end{aligned}$$

The variance of $\hat{\hat{Y}}_{HT,d}$ is given by:

$$\mathbb{V}_p(\hat{\hat{Y}}_{HT,d}) = \frac{1}{N_d^2}\left(\sum_{i \in U_d}\sum_{j \in U_d} \Delta_{ij}(d_i z_i)(d_j z_j) + \sum_{i \in U_d}\sum_{j \in U_d} \Delta_{ij}(d_i t_i)(d_j t_j)\right),$$

which can be unbiasedly estimated with

$$\hat{\mathbb{V}}_p(\hat{\hat{Y}}_{HT,d}) = \frac{1}{N_d^2}\left(\sum_{i \in s_{1d}}\sum_{j \in s_{1d}} \check{\Delta}_{ij}(d_i z_i)(d_j z_j) + \sum_{i \in s_{2d}}\sum_{j \in s_{2d}} \check{\Delta}_{ij}(d_i t_i)(d_j t_j)\right).$$

This variance may be unacceptably large for certain domains. Notwithstanding, it may be improved by using calibration when (multi-)auxiliary information on the domains is available. In this paper, however, we only discuss design-based estimation for sufficiently large domains. If the (random) size of the domain sample s_d is insufficient to meet demands concerning the precision of the estimates, small-area (model-based) estimation may be needed.

Using the same notation as in Section 4.2, if the vector of the population means $\bar{\mathbf{X}}$ is known in the domain U_d, the domain calibration-type estimator can be defined as:

$$\hat{\bar{Y}}_{C,d} = \frac{1}{N_d}\sum_{i\in s_1}\omega_{i1}z_i\delta_{di} - \frac{1}{N_d}\sum_{i\in s_2}\omega_{i2}t_i\delta_{di},$$

where weights ω_{ij}, $j = 1, 2$, are determined by minimizing the χ^2 distance function

$$\Phi_{s_{jd}}(d_i, \omega_{ij}) = \sum_{i\in s_{jd}} \frac{(\omega_{ij} - d_i)^2}{d_i q_i}, \quad j = 1, 2$$

subject to the conditions

$$\bar{\mathbf{X}}_{U_d} = \frac{1}{N_d}\sum_{i\in U_d}\mathbf{x}_i = \frac{1}{N_d}\sum_{i\in s_j}\omega_{ij}\mathbf{x}_i\delta_{di}$$

and

$$N_d = \sum_{i\in s_j}\omega_{ij}\delta_{di}, \quad j = 1, 2.$$

The expressions of $\hat{\bar{Y}}_{C,d}$, of its variance, and of the variance estimator can easily be obtained by adapting the results given in Section 4.2.

5 Simulation study

This section presents two simulation studies to numerically investigate the performance of the HT and calibration-type estimators when sensitive quantitative data are to be obtained by the IST. The first study is designed to: (i) compare the proposed IST estimators and a competitor RRT estimator which uses two different scrambling variables; (ii) evaluate, within the IST framework, the effects of using innocuous items with different correlations with the target sensitive variable; (iii) evaluate the performance of the IST for domain estimation. The second simulation study highlights the accuracy of the variance estimators and enables us to evaluate the effects of using more than one nonsensitive item in the calibration setting.

5.1 Simulation 1: comparisons and correlations

The study is based on real data obtained by World Bank Enterprise Surveys compiled in China between December 2011 and February 2013 (http://www.enterprisesurveys.org). During this period, 2700 privately-owned firms and 148 state-owned firms were interviewed. The total sales value for 2011 was taken as the study variable (y). In order to perform the IST procedure, the total annual cost of electricity was taken as the innocuous variable (t). The estimation for the entire population and for the study domains are discussed below. To estimate the population mean $\bar{Y}$ in the IST setting, we first calculated the HT-type estimator (2) and compared it with the calibration-type estimator (6). Calibration was performed with respect to the following auxiliary variables: total annual sales three years ago (2009), permanent/full-time workers three fiscal years ago (2009), and firm's yearly average inventory of finished goods in 2011. To determine the cost in terms of loss of efficiency of using the IST to increase respondents' privacy protection, we also considered the corresponding estimators of $\bar{Y}$, say $\hat{\bar{Y}}_{HT_y}$ and $\hat{\bar{Y}}_{C_y}$, which were computed on the basis of the true value of the target variable. Additionally, the HT and calibration-type estimates were compared with the estimates derived from another indirect questioning method referable to the RRT. Thus, the responses were assumed to be randomized by the scrambled response model (SRM) proposed by Bar-Lev et al. (2004). According to this model, the ith survey unit provides the randomized response z_i defined as:

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

where w_i is a random number generated from the scrambling variable w whose distribution is completely known to the researcher. Hence, an unbiased HT estimator for $\bar{Y}$ is obtained as:

$$\hat{\bar{Y}}_{SRM} = \frac{1}{N}\sum_{i\in s} d_i r_i,$$

with

$$r_i = \frac{z_i}{\theta + (1-\theta)\bar{W}}$$

and where $\bar{W}$ denotes the known mean of w. We assumed $\theta = 0.5$ and then investigated the performance of the estimates under two different distribution laws for the scrambling variable w:

- $w \sim F_{10,10}$ as in Eichhorn and Hayre (1983) and Arcos et al. (2015). We refer to this choice as SRM_1;
- $w \sim exp(1)$ as in Rueda et al. (2017). We refer to this choice as SRM_2.

In our study, available data at firm-level were taken as the target population from which a sample of size n was selected according to: (i) simple random sampling without replacement (SRSWOR); (ii) stratified SRSWOR; (iii) Midzuno sampling design (see, e.g., Sukhatme et al., 1984). The sample size ranges from 25 to 200 firms. The population was then stratified into three industrial sectors, termed "manufacturing", "retail" and "other services", after recoding the available variables. From each stratum, a number of samples were selected according to SRSWOR with proportional allocation from 5% to 15% of the population size. The Midzuno sampling design was implemented with first-order inclusion probabilities proportional to the number of establishments owned by the firm.

In order to evaluate the performance of the HT and calibration estimators under the DQ, IST and RRT survey modes, the absolute relative bias (RB) and relative mean squared error (RMSE) were computed for the estimator $\hat{\hat{Y}}^* = \hat{\hat{Y}}_{HT_y}, \hat{\hat{Y}}_{C_y}, \hat{\hat{Y}}_{HT}, \hat{\hat{Y}}_C, \hat{\hat{Y}}_{SRM_1}, \hat{\hat{Y}}_{SRM_2}, \hat{\hat{Y}}_{C_1}, \hat{\hat{Y}}_{C_2}$:

$$|\text{RB}(\hat{\hat{Y}}^*)| = \left| \frac{\mathbb{E}_M(\hat{\hat{Y}}^*) - \bar{Y}}{\bar{Y}} \right| \quad \text{and} \quad \text{RMSE}(\hat{\hat{Y}}^*) = \frac{\mathbb{E}_M(\hat{\hat{Y}}^* - \bar{Y})^2}{\bar{Y}^2},$$

where $\hat{\hat{Y}}_{C_i}$ denotes the calibration estimator of $\bar{Y}$ under the SRM$_i$, $i = 1, 2$, while $\mathbb{E}_M$ denotes the mean operator evaluated on the basis of 10,000 Monte Carlo replications for different sample sizes.

The results of the simulation study for the three different sampling designs are illustrated in Figure 1. Although the behavior of the SRM estimates appears irregular, there is no evidence of any significant bias for all the estimators considered, at least as the sample size increases. In fact, for all the estimators, the absolute RB falls within a reasonable range. In terms of RMSE, the IST estimators perform well. Overall, these findings are very interesting and highlight the successful use of auxiliary information at the IST estimation stage. While the HT estimator based on the true values y_i slightly outperforms, as expected, the HT-type estimator based on the IST values z_i, the calibration estimators are unexpectedly nearly equivalent, both in terms of (absolute) bias and of mean squared error. On the other hand, the behavior of the SRM estimators is less stable and less satisfactory than that of the IST estimator $\hat{\hat{Y}}_{HT}$. This is particularly true for the estimates obtained using SRM$_2$, which are generally less efficient than those provided by $\hat{\hat{Y}}_{HT}$. As regards the estimates under SRM$_1$, in some cases across the three sampling designs, they appear to be slightly more efficient than $\hat{\hat{Y}}_{HT}$ but, in general, the IST seems to outperform the RRT approach, at least for the scrambling models considered in the present study. This results also holds when SRM and IST estimates are compared under the calibration setting.

For all the estimators considered, it is also evident that using auxiliary information at the design stage through stratification and sampling with varying probability can improve the efficiency of the estimates with respect to SRSWOR. In this study, the improvement obtained by stratification is notable.

Finally, the mean squared error of the estimators tendentially decreases as the sample size increases, which is an evident indication of the consistency of all the estimates produced.

[Figure 1]

We then focused on the IST approach and investigated the influence on the estimates produced by innocuous variables which exhibit different degrees of correlation with the target variable. Therefore, the above simulation was repeated, but considering, as well as the nonsensitive variable “total annual cost of electricity” ($t = t_1$ with $\rho_{yt_1} = 0.753$), the variable “total annual rental cost of machinery, vehicles and equipment” ($t = t_2$ with $\rho_{yt_2} = 0.526$) and the variable “total annual cost of raw materials” ($t = t_3$ with

$\rho_{yt_3} = 0.811$).

The results of the simulation concerning only the performance of the estimators in the IST framework are illustrated in Figure 2.

[Figure 2]

We observe that two HT-type estimators $\hat{\hat{Y}}_{HT_1}$ and $\hat{\hat{Y}}_{HT_2}$, which employ t_1 and t_2 as auxiliary variables, show a similar performance while, when using the auxiliary variable t_3, the efficiency of the estimates decreases. Hence, the choice of which innocuous variable to use is a matter of some importance for the researcher. On the contrary, no striking differences are apparent when the IST calibration estimators are considered, and the results appear to be robust to the choice of the innocuous variable. For the IST calibrated estimators, the correlation between the target variable and the calibration variable is more important than that between the target and the innocuous variable.

Finally, we investigated the behavior of the estimators when we wish to obtain estimates for population domains. For this purpose, the above study was repeated, but splitting the firms into domains according to the numbers of employers. In this case, three domains were considered: small, medium and large firms. Again, we focused only on the IST approach. For brevity, Figure 3 shows only the outcomes of stratified sampling. The results obtained are very similar to those of the first simulation study and confirm that the IST can also be profitably used in more complex survey situations.

[Figure 3]

5.2 Simulation 2: focusing on the IST calibration estimator

In the previous simulation study, we ascertained the very good performance of the IST calibration estimators. Accordingly, we then focused on the calibration approach and ran a new simulation in order to explore some additional features concerning: (i) the influence on the estimates of the length of the list; (ii) the accuracy of the variance estimation.

For this purpose, we considered the population included in Shaw (2015). This population is composed of $N = 117$ units and includes, beside the target variable y, five innocuous variables. To perform the calibration we generated a new variable (x) correlated with y ($\rho_{yx} = 0.754$). The population was stratified into three strata using the cut-off values 4 and 7 of y. Hence, 10,000 samples of several sample sizes were selected from the population according to SRSWOR and stratified SRSWOR. In this process, for each sample, the calibration estimates are obtained by increasing the number of innocuous items. Let $\hat{\hat{Y}}_{C,G}$ denote the IST calibration estimator for the list of items which includes G innocuous variables, $G = 1, \ldots, 5$. Hence, for each $\hat{\hat{Y}}_{C,G}$, we computed the absolute RB and RMSE as in Section 5.1.

The results obtained are shown in Figure 4. Clearly, the performance of the estimators strongly depends on the length of the list. As the number of innocuous items increases, both the absolute RB and the RMSE increase, although the RB always remains within an acceptable range of values. The fact that the efficiency of the estimates deteriorates as the length of the list increases is not surprising, since the more innocuous items are included, the higher the variance of the total score t reported by the respondents. The best performance of the estimators is achieved when one or two innocuous variables are used to perturb the true sensitive response. With respect to this point, Trappmann et al. (2014) suggested using a single nonsensitive item in order to improve the efficiency of the procedure.

[Figure 4]

Finally, another simulation was run to investigate the behavior of the variance estimator of $\hat{\hat{Y}}_{C,G}$. This experiment is summarized in the following steps:

1. For all the IST situations considered, calibration-type estimates are computed on the basis of 50,000 samples selected from Shaw's population (sample sizes ranging from 10 to 50 units) according to SRSWOR and stratified SRSWOR. Hence, an approximation of the true theoretical variance of $\hat{\hat{Y}}_{C,G}$ is achieved by the simulated variance:

$$\mathbb{V}_{\text{sim}}(\hat{\hat{Y}}_{C,G}) = \frac{1}{50000} \sum_{k=1}^{50000} (\hat{\hat{Y}}_{C,G}^{(k)} - \bar{Y})^2$$

   where $\hat{\hat{Y}}_{C,G}^{(k)}$ is the calibration-type estimate computed on the kth sample and $G = 1, \ldots, 5$;

2. 10,000 Monte Carlo samples are drawn from Shaw's population according to SRSWOR and stratified SRSWOR, and variance estimates $\hat{\mathbb{V}}(\hat{\hat{Y}}_{C,G})$ are computed as reported in (7);

3. The absolute relative bias and relative mean squared error for the variance estimates are computed as:

$$|\text{RB}(\hat{\mathbb{V}}(\hat{\hat{Y}}_{C,G}))| = \left| \frac{\mathbb{E}_M(\hat{\mathbb{V}}(\hat{\hat{Y}}_{C,G})) - \mathbb{V}_{\text{sim}}(\hat{\hat{Y}}_{C,G})}{\mathbb{V}_{\text{sim}}(\hat{\hat{Y}}_{C,G})} \right|$$

and

$$\text{RMSE}(\mathbb{V}(\hat{\hat{Y}}_{C,G})) = \frac{\mathbb{E}_M(\hat{\mathbb{V}}(\hat{\hat{Y}}_{C,G}) - \mathbb{V}_{\text{sim}}(\hat{\hat{Y}}_{C,G}))^2}{(\mathbb{V}_{\text{sim}}(\hat{\hat{Y}}_{C,G}))^2}.$$

Figure 5 shows the behavior of the absolute RB and the RMSE for different sample sizes and under the two sampling designs.

[Figure 5]

Overall, both the absolute RB and the RMSE of the variance estimator for the suggested IST calibration estimator produce very small values. Moreover, we observe that: (i) the RMSE decreases as the sample size increases; (ii) the satisfactory behavior of the variance estimator does not seem to be affected by the increased number of innocuous variables used to perform the IST.

6 Conclusions

This article describes advances that may be achieved in the use of the IST when auxiliary information is available for the entire population, at no additional cost. This situation is very common in sampling practice and has given rise to many papers discussing the situation when nonsensitive parameters must be estimated. However, to the best of our knowledge, very few studies have addressed the question of estimating a quantitative sensitive characteristic when using the IST and auxiliary information. This is probably due to the fact that the IST has only recently been introduced, as a variant of the much better known ICT, which is suitable for collecting data on sensitive attributes.

In our work, auxiliary information is employed at both the design and the estimation stages. In particular, under a generic sampling design, we introduce, for a two-list experiment, a Horvitz-Thompson-type estimator and a calibration-type estimator in order to efficiently estimate the mean of a sensitive quantitative variable. The performance of the proposed estimators is analyzed extensively by means of simulation experiments based on two data sets. Specifically, the efficiency of the two estimators based on "perturbed data" is compared with that of analogous estimators based on "true data". This comparison is then extended to include the RRT, an indirect questioning mode that represents an alternative to the IST. This comparison is effected under different sampling designs. The results arising from the simulation study are very interesting and promising. For the data considered, at least, our findings reveal that IST surveys can provide estimates which are nearly as efficient as those obtained from a DQ survey while, in general, outperforming RRT estimates. This is particularly true for the calibration-type estimators. Accordingly, we further investigated the behavior of these estimators by running additional simulations in order to assess variance estimation and the impact made on the estimates when the number of innocuous variables is increased.

The idea of using calibration in the IST is certainly original and merits future research attention. We hope that the promising results obtained from this study will encourage academics and researchers to incorporate our proposal into applied studies, to gain a better understanding of the potential of the IST in real-world analyses and to contribute to extending its use as an alternative indirect questioning technique in surveys.

Acknowledgments

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

References

[1] Arcos, A., Rueda, M.M., Singh, S., 2015. Generalized approach to randomized response for quantitative variables. *Quality & Quantity* 49: 1239-1256

[2] Arias, A., Sutton, S.G. 2013. “Understanding recreational fishers compliance with no-take zones in the Great Barrier Reef Marine Park.” *Ecology and Society* 18: Available at http://dx.doi.org/10.5751/ES-05872-180418.

[3] 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.

[4] Aronow, P.M., Coppock, A., Crawford, F.W., Green, D.P. 2015. “Combining list experiment and direct question estimates of sensitive behavior prevalence.” *Journal of Survey Statistics and Methodology* 3:43-66.

[5] Bar-Lev, S.K., Bobovitch, E., Boukai, B. 2004. “A note on randomized response models for quantitative data.” *Metrika* 60:255-260.

[6] Blair, G., Imai, K. 2010. *list: Statistical Methods for the Item Count Technique and List Experiment.* URL=http://CRAN.Rproject.org/package=list.

[7] Blair, G., Imai, K. 2012. “Statistical analysis of list experiments.” *Political Analysis* 20:47-77.

[8] Blair, G., Imai, K., Lyall, J. 2014. “Comparing and combining list and endorsement experiments: Evidence from Afghanistan.” *American Journal of Political Science* 58:1043-1063.

[9] Blank, S.G., Gavin, M.C. 2009. "The randomized response technique as a tool for estimating non-compliance rates in fisheries: a case study of illegal red abalone (Haliotis rufescens) fishing in Northern California." *Environmental Conservation* 36:112-119.

[10] Chaloupka, M. Y. 1985. "Application of the randomized response technique to marine park management: An assessment of permit compliance." *Environmental Management* 9:393-398.

[11] Chaudhuri, A. 2011. *Randomized Response and Indirect Questioning Techniques in Surveys.* Chapman & Hall/CRC, Boca Raton, FL.

[12] Chaudhuri, A., Christofides, T.C. 2007. "Item count technique in estimating the proportion of people with a sensitive feature." *Journal of Statistical Planning and Inference* 137:589-593.

[13] Chaudhuri, A., Christofides, T.C. 2013. *Indirect Questioning in Sample Surveys.* Springer-Verlag Berlin Heidelberg.

[14] Chaudhuri, A., Mukerjee, R. 1988. *Randomized Response: Theory and Techniques.* Marcel Dekker, Inc., New York.

[15] Conteh, A., Gavin M.C., Solomon, J. 2015. "Quantifying illegal hunting: A novel application of the randomised response technique." *Biological Conservation* 189:16-23.

[16] Christofides, T.C. 2015. "A new version of the item count technique." *Model Assisted Statistics and Applications* 10:289-297.

[17] Deville, J.C., and Särndal, C.E. 1992. "Calibration estimators in survey sampling." *Journal of the American Statistical Association* 87:376-382.

[18] Diana, G., Perri, P.F. 2009. "Estimating a sensitive proportion through randomized response procedures based on auxiliary information." *Statistical Papers* 50:661-672.

[19] 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.

[20] Diana, G., Perri, P.F. 2011. "A class of estimators for quantitative sensitive data." *Statistical Papers* 52:633-650.

[21] Diana, G., Perri, P.F. 2012. "A calibration-based approach to sensitive data: A simulation study." *Journal of Applied Statistics* 39:53-65.

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

[23] Droitcour, J.A., Caspar, R.A., Hubbard, M.L., Parsley, T.L., Visseher, W., and Ezzati, T.M. 1991. "The item count technique as a method of indirect questioning: A review of its development and a case study application." Pp. 187-209. In: *Measurement Error in Surveys.* Biemer, P.P., Groves, R.M., Lyburg, L.E., Mathiowetz, N., Sudmar. S. (Eds), Wiley, New York.

[24] Droitcour, J.A., Larson, E.M. 2002. "An innovative technique for asking sensitive questions: The three-card method." *Bulletin of Sociological Methodology* 75:5-23.

[25] Eichhorn, B.H., Hayre, L.S., 1983. "Scrambled randomized response methods for obtaining sensitive quantitative data". *Journal of Statistical Planning and Inference* 7:306-316.

[26] Fox, J.A., Tracy, P.E. 1986. *Randomized Response: A Method for Sensitive Survey*. Sage Publication, Inc., Newbury Park.

[27] Glynn, A.N. 2013. "What can we learn with statistical truth serum? Design and analysis of the list experiment." *Public Opinion Quarterly* 77:159-172.

[28] Goodstadt, M.S., Gruson, V. 1975. "The randomized response technique: A test on drug use." *Journal of the American Statistical Association* 70:814-818.

[29] Holbrook, A.L., and Krosnick J.A. 2010a. "Measuring voter turnout by using the randomized response technique: Evidence calling into question the method's validity." *Public Opinion Quarterly* 74:328-343.

[30] Holbrook, A.L., and Krosnick J.A. 2010b. "Social desirability bias in voter turnout reports: Tests using the item count technique." *Public Opinion Quarterly* 74:37-67.

[31] Horvitz, D.G., Thompson, D.J. 1952. "A generalization of sampling without replacement from a finite universe." *Journal of the American Statistical Association*, 47:663-685.

[32] Houston, J., Tran, A. 2001. "A survey of tax evasion using the randomized response technique." *Advances in Taxation* 13:69-94.

[33] Hussain, Z., Shah, E.A., Shabbir, J. 2012. "An alternative item count technique in sensitive surveys." *Revista Colombiana de Estadistica* 35:39-54.

[34] Hussain, Z. Shabbir, N., Shabbir J. 2017. "An alternative item sum technique for improved estimators of population mean in sensitive surveys." *Hacettepe Journal of Mathematics and Statistics*, 46:907-934

[35] Imai, K. 2011. "Multivariate regression analysis for the item count technique." *Journal of the American Statistical Association* 106:407-416.

[36] Imai, K., Park, B., Greene, K.F. 2015. "Using the predicted responses from list experiments as explanatory variables in regression models." *Political Analysis* 23:180-196.

[37] James R.A., Nepusz T., Naughton D.P., Petróczi A. 2013. "A potential inflating effect in estimation models: Cautionary evidence from comparing performance enhancing drug and herbal hormonal supplement use estimates." *Psychology of Sports and Exercise* 14:84-96.

[38] Jann, B., Jerke, J., Krumpal, I. 2012. "Asking sensitive questions using the crosswise model: An experimental survey measuring plagiarism." *Public Opinion Quarterly* 76:32-49.

[39] Kerkvliet, J. 1994 "Estimating a logit model with randomized data: The case of cocaine use." *Australian Journal of Statistics* 36:9-20.

[40] Korndörfer, M., Krumpal, I., Schmukle, S.C. 2014. "Measuring and explain tax evasion: improving self-reports using the crosswise model." *Journal of Economic Psychology* 45:18-32.

[41] Kott P.S. 2014. "Developing calibration weights and standard-error estimates for a survey of drug-related emergency-department visits." *Journal of Official Statistics* 30:521-532.

[42] Kott P.S., Chang, T. 2010. "Using calibration weighting to adjust for nonignorable unit nonresponse." *Journal of the American Statistical Association* 105:1265-1275.

[43] Krumpal, I. 2012. "Estimating the prevalence of xenophobia and anti-semitism in Germany: A comparison of the randomized response technique and direct questioning." *Social Science Research* 41:1387-1403.

[44] Lara, D., Strickler, J., Olavarrieta, C.D., García, S.G., Ellertson, C. 2004. "Measuring induced abortion in Mexico: A comparison of four methodologies." *Sociological Methods & Research* 32:529-558.

[45] Lara, D., García, S.G., Ellertson, C., Camlin, C., Suaréz, J. 2006. "The measure of induced abortion in Mexico using random response technique." *Sociological Methods & Research* 35:279-301.

[46] Lensvelt-Mulders, G.J.L.M., Hox J.J., van der Heijden, P.G.M., Mass, C.J.M. 2005. "Meta-analysis of randomized response research: Thirty-five years of validation." *Sociological Methods & Research* 33:319-348.

[47] Lensvelt-Mulders, G.J.L.M., van der Heijden, P.G.M., Laudy, O., van Gils, G. 2006. "A validation of a computer-assisted randomized response survey to estimate the prevalence of fraud in social security." *Journal of the Royal Statistical Society-A* 169:305-318.

[48] Miller, J.D. 1984. *A New Survey Technique for Studying Deviant Behavior.* Ph.D. Thesis, The George Washington University.

[49] Miller, J.D. 1985. *The Nominative Technique: A New Method of Estimating Heroin Prevalence.* NIDA Research Monograph 57:104-124.

[50] Oliveras, E., Letamo, G. 2010. "Examples of methods to address underreporting of induced abortion: preceding birth technique and randomized response technique." Pp. 63-69 in *Methodologies for Estimating Abortion Incidence and Abortion-Related Morbidity: A Review.* Singh S., Remez L., Tartaglione A. (Eds.) Guttmacher Institute, New York - International Union for the Scientific Study of Population, Paris.

[51] Ostapczuk, M., Musch, J., Mashagen, M. 2009. "A randomized-response investigation of the education effect in attitudes towards foreigners." *European Journal of Social Psychology* 39:920-931.

[52] Perri, P.F., Diana, G. 2013. "Scrambled response models based on auxiliary variables." Pp. 281-291. In: *Advances in Theoretical and Applied Statistics.* Torelli, N., Pesarin, F. and Bar-Hen, A. (Eds.), Spriger-Verlag, Berlin.

[53] Perri, P.F., Pelle, E., Stranges, M. 2016. "Estimating induced abortion and foreign irregular presence using the randomized response crossed model." *Social Indicators Research* 129:601-618.

[54] Raghavarao, D., Federer, W.F. 1979. "Block total response as an alternative to the randomized response method in survey." *Journal of the Royal Statistical Society-B* 41:40-45.

[55] Rueda, M., Cobo, B, Arcos, A. 2017. "An improved class of estimators in RR surveys." *Mathematical Methods in the Applied Sciences.* In press. DOI: 10.1002/mma.4256

[56] Särndal, C.E. 2007. "The calibration approach in survey theory and practice." *Survey Methodology* 33:99-119.

[57] Särndal, C.E., and Lundström, S. 2005. *Estimation in Survey with Nonresponse.* Wiley, New York.

[58] Schill, D.J., Kline, P.A. 1995. "Use of random response to estimate angler noncompliance with fishing regulations." *North American Journal of Fisheries Management* 15:721-731.

[59] 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.

[60] Shaw, P. 2015. “Estimating a finite population mean of a sensitive quantitative variable from a single probability sample by the Item Count Technique.” *Model Assisted Statistics and Applications* 10:411-419.

[61] Shaw, P. 2016. “Estimating a finite population proportion bearing a sensitive attribute from a single probability sample by Item Count Technique.” Pp 387-404. In: *Handbook of Statistics vol. 34: Data Gathering, Analysis and Protection of Privacy Through Randomized Response Techniques: Qualitative and Quantitative Human Traits.* Chaudhuri, A., Christofides, T.C. and Rao, C.R. (Eds), Elsevier, Amsterdam, Netherlands

[62] 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.

[63] Solomon, J., Jacobson, S.K., Wald, K.D., Gavin, M. 2007. “Estimating illegal resources use at the Ugandan park with the randomized response technique.” *Human Dimensions of Wildlife* 12:75-88.

[64] Striegel, H., Ulrich, R., Simon, P. 2010. “Randomized response estimates for doping and illicit drug use in elite athletes.” *Drug and Alcohol Dependence* 106:230-232.

[65] Stubbe, J.H., Chorus, A.M.J., Frank, L.E., de Hon, O., van der Heijden P.G.M. 2013. “Prevalence of use of performance enhancing drugs by fitness center members.” *Drug Testing and Analysis* 6:434-438.

[66] Sukhatme, P.V., Sukhatme, B.V., Sukhatme, S., Asok, C. 1984. *Sampling Theory of Surveys with Applications* Iowa State University Press, Ames, Iowa.

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

[68] Trappmann, M., Krumpal, I., Kirchner, A. and Jann, B. 2014. “Item sum: A new technique for asking quantitative sensitive questions.” *Journal of Survey Statistics and Methodology* 2:58-77.

[69] van der Heijden, P.G.M., van Gils, G., Bouts, J., Hox, J.J. 2000. “A comparison of randomized response, computer-assisted self-interview, and face-to-face direct questioning.” *Sociological Methods & Research* 28:505-537.

[70] Warner S.L. 1965. “Randomized response: A survey technique for eliminating evasive answer bias.” *Journal of the American Statistical Association* 60:63-69.

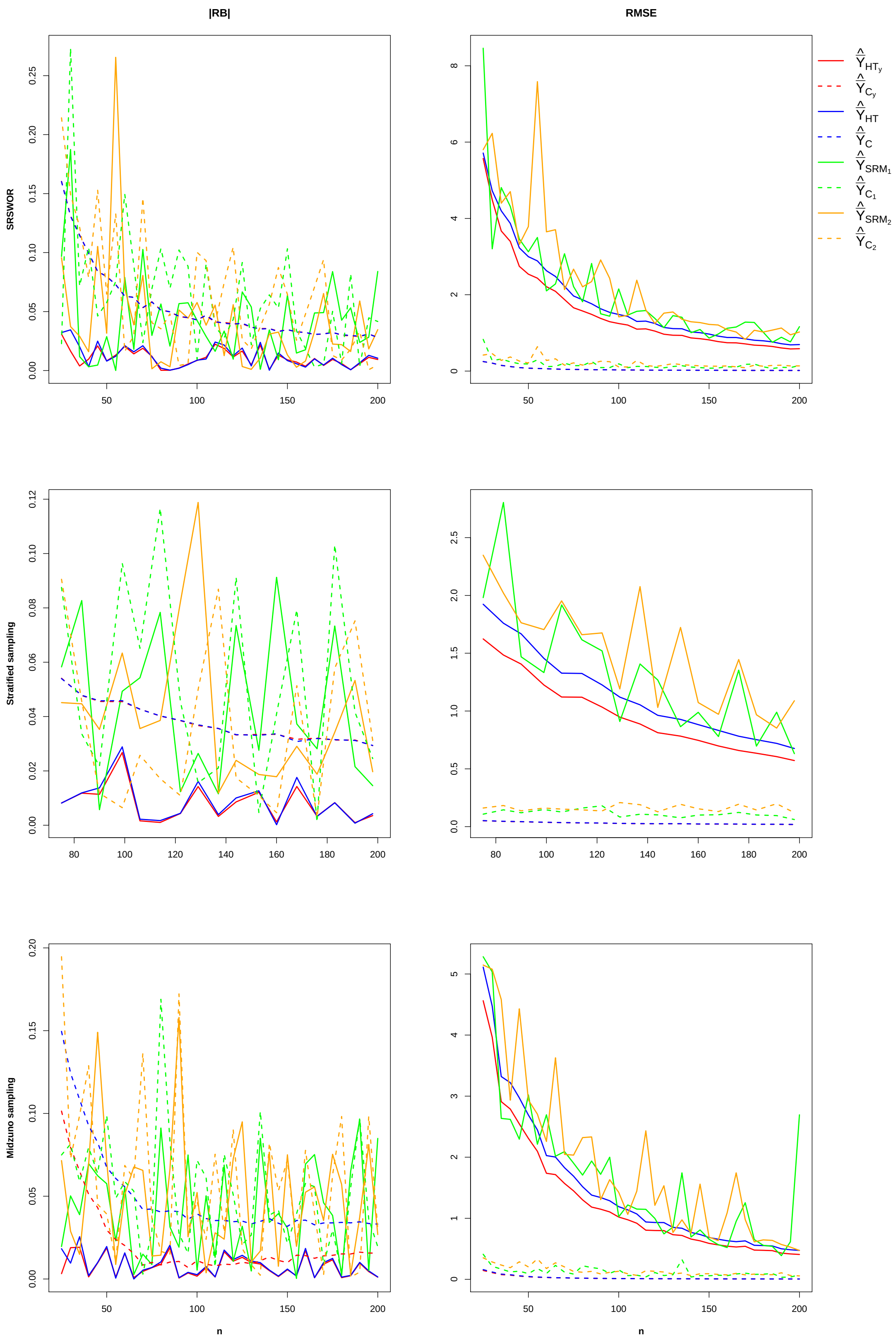


Figure 1: Performance of the HT and calibration estimators under DQ, IST and RRT survey modes

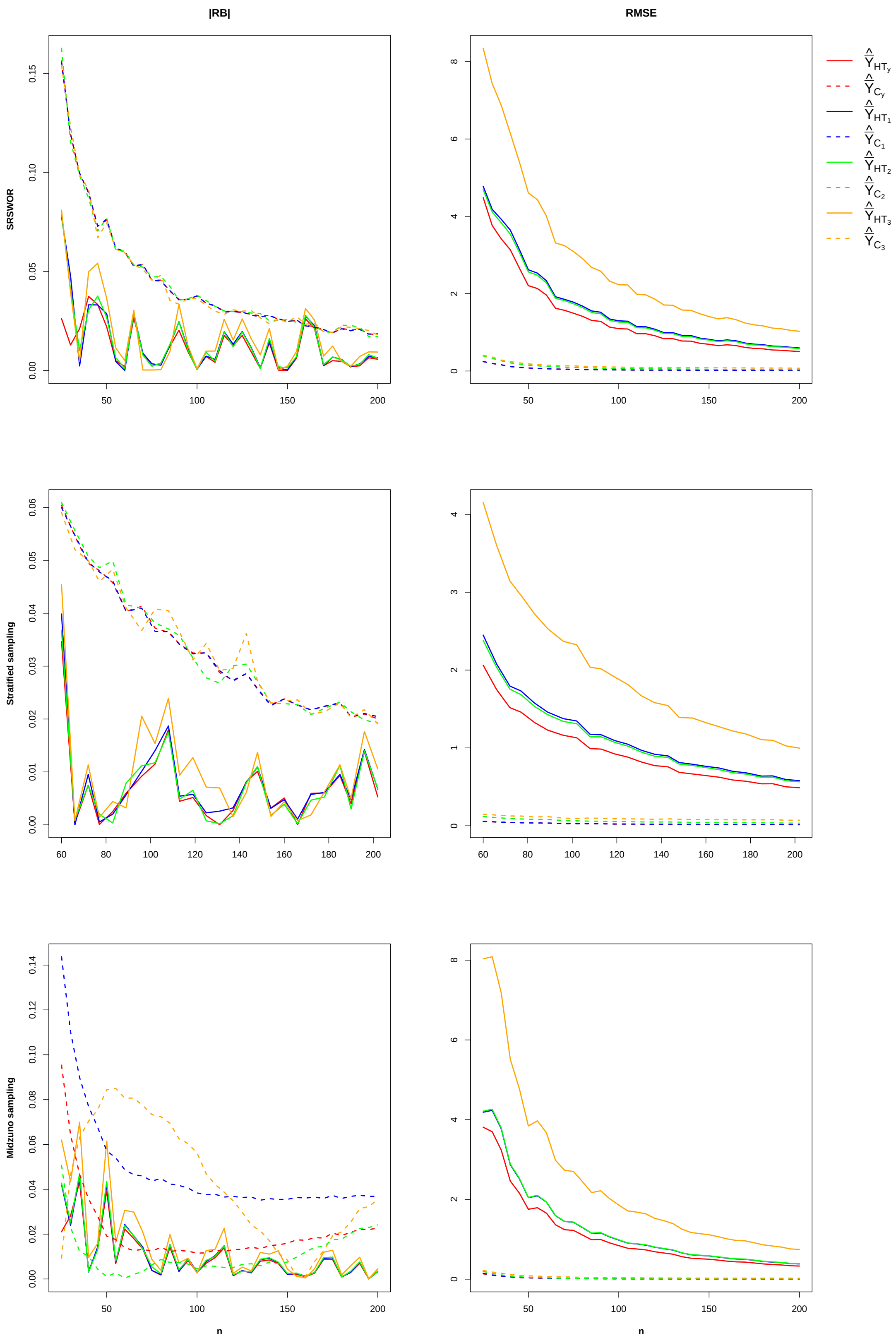


Figure 2: Performance of the HT and calibration estimators under DQ and IST survey modes and for different correlations

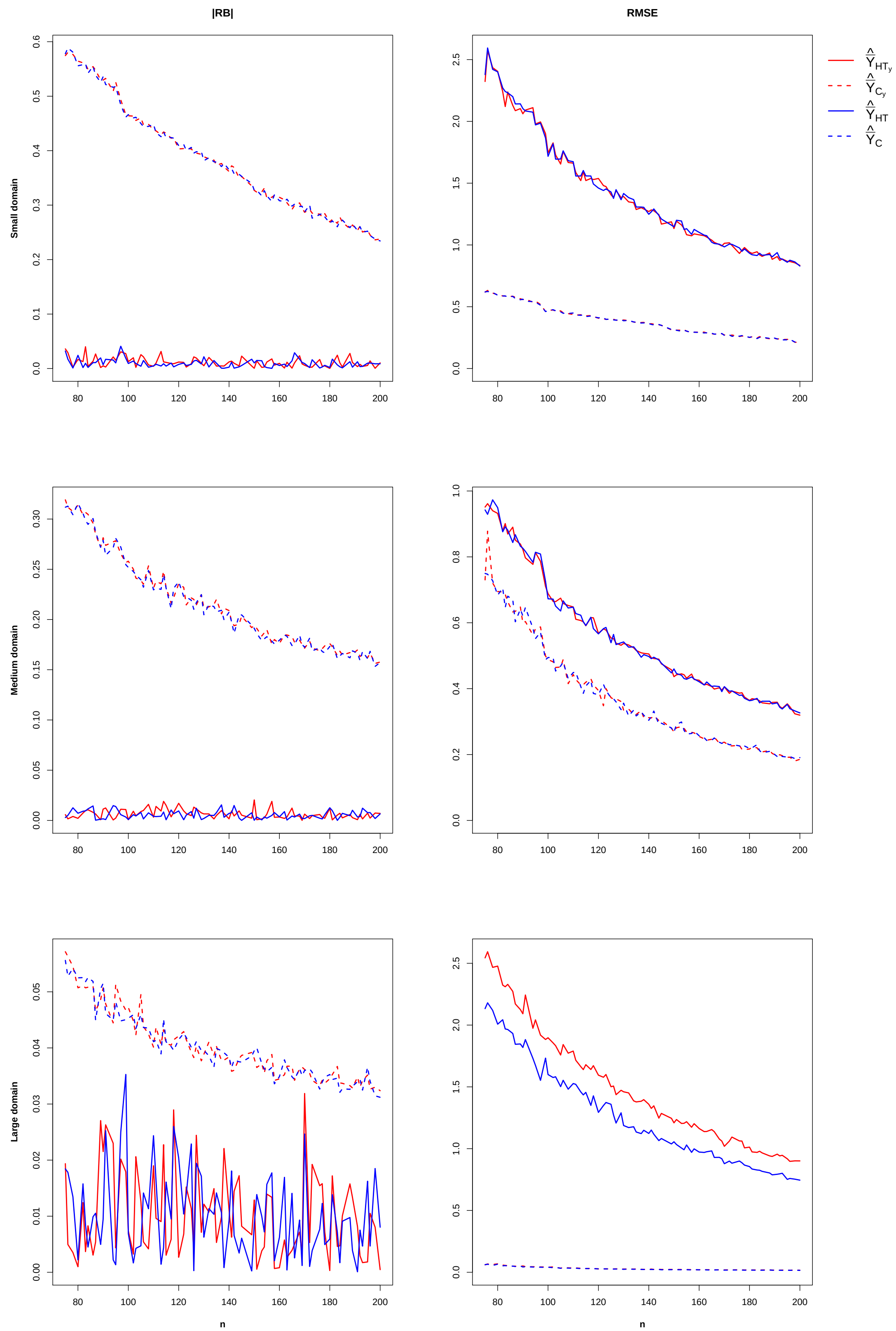


Figure 3: Stratified domain estimates by the HT and calibration estimators under DQ and IST survey modes

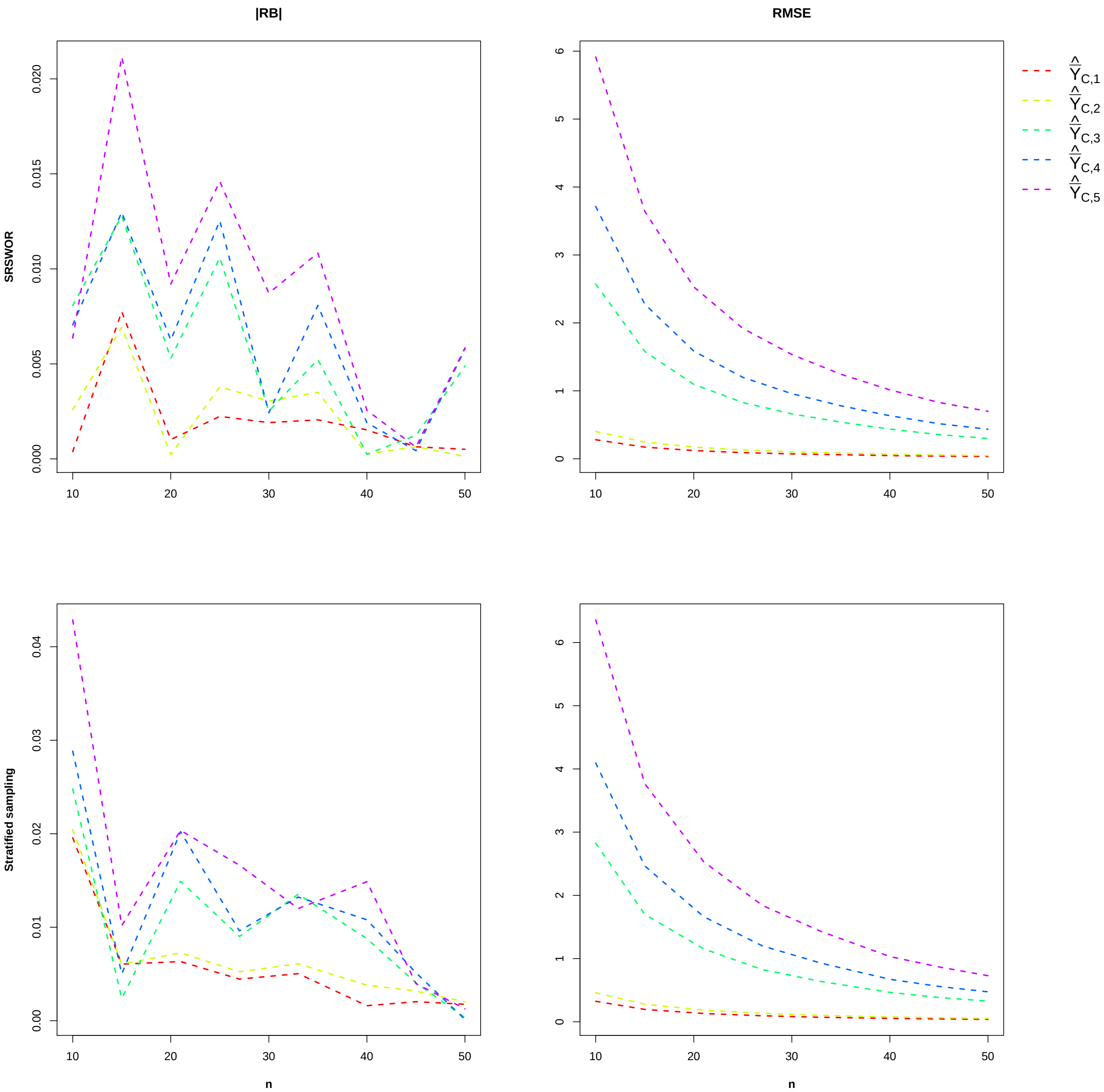


Figure 4: Performance of the IST calibration estimators with an increasing number of innocuous items

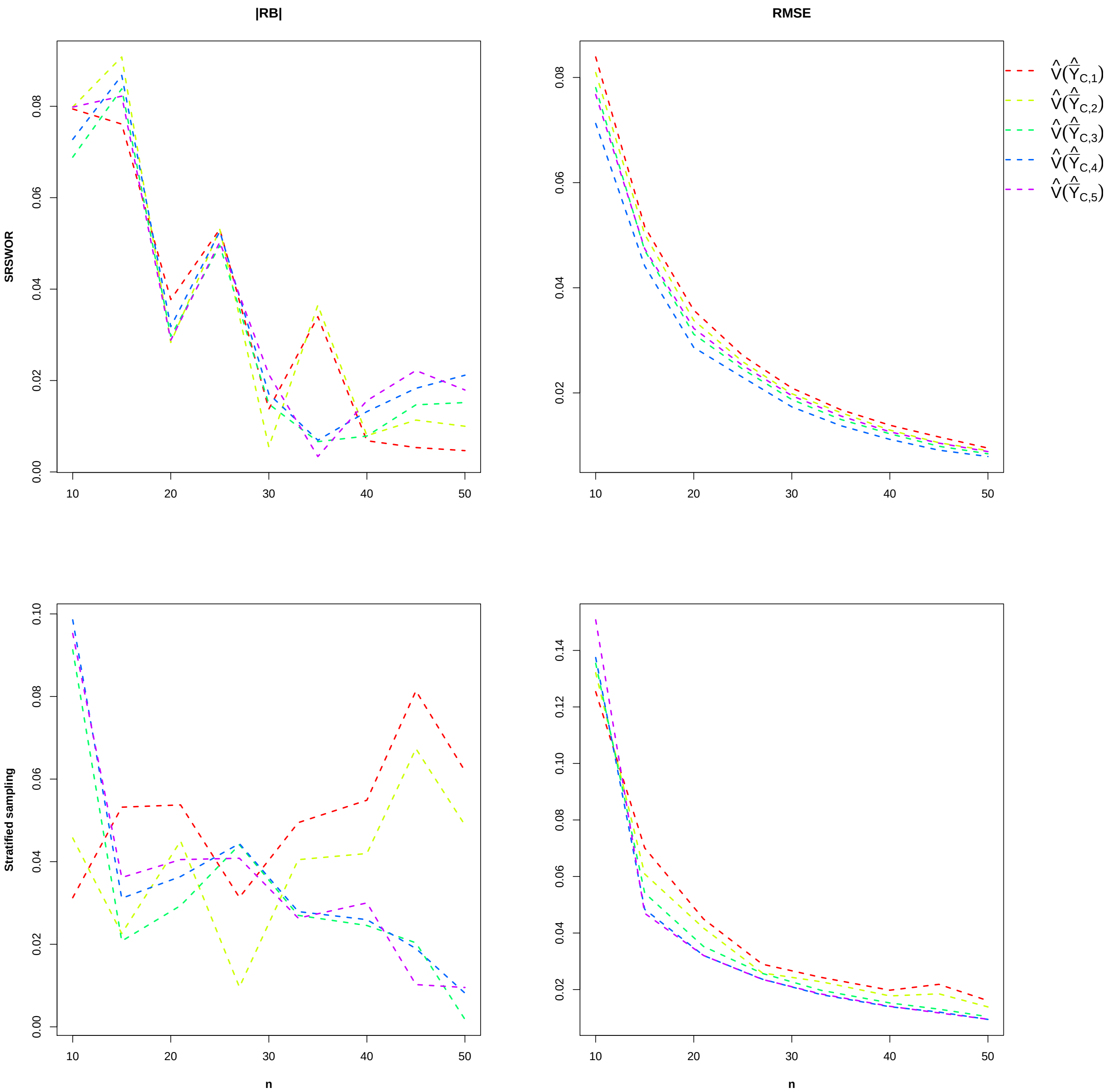
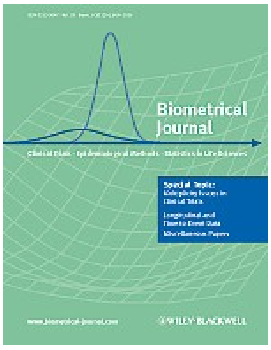


Figure 5: Performance of the variance estimator for the IST calibration estimators with an increasing number of innocuous items

Appendix A5

Multiple sensitive estimation and optimal sample size allocation in the item sum technique

Perri, Pier Francesco; Rueda, María del Mar; Cobo, Beatriz. (2017)
Multiple sensitive estimation and optimal sample size allocation in the item sum technique.
Biometrical Journal, vol. 60, number 1, pp. 155 - 173.
DOI: 10.1002/bimj.201700021



STATISTICS & PROBABILITY			
JCR Year	Impact factor	Rank	Quartile
2016	1.075	55/124	Q2

Abstract

For surveys of sensitive issues in life sciences, statistical procedures can be used to reduce nonresponse and social desirability response bias. Both of these phenomena provoke nonsampling errors that are difficult to deal with and can seriously flaw the validity of the analyses. The item sum technique (IST) is a very recent indirect questioning method derived from the item count technique which seeks to procure more reliable responses on quantitative items than direct questioning while preserving respondents' anonymity. This article addresses two important questions concerning the IST: (i) its implementation when two or more sensitive variables are investigated and efficient estimates of their unknown population means are required; (ii) the determination of the optimal sample size to achieve minimum variance estimates. These aspects are of great relevance for survey practitioners engaged in sensitive research and, to the best of our knowledge, were not studied so far. In this article, theoretical results for multiple estimation and optimal allocation are obtained under a generic sampling design and then particularized to simple random sampling and stratified sampling designs. Theoretical considerations are integrated with a number of simulation studies based on data from two real surveys and conducted to ascertain the efficiency gain derived from optimal allocation in different situations. One of the surveys concerns cannabis consumption among university students. Our findings highlight some methodological advances that can be obtained in life sciences IST surveys when optimal allocation is achieved.

1 Introduction

Studies in life and social sciences addressing highly personal, embarrassing, stigmatizing, threatening or even incriminating issues often yield unreliable estimates of unknown characteristics of the population under study, due to nonresponse (unit-nonresponse or item-nonresponse) and socially desirable responding. In particular, social desirability bias, i.e. the desire to make a favorable impression on others, poses a significant threat to the validity of self-reports in "sensitive research" as well described in Dickson-Swift et al. (2008).

Refusal to answer and false answers constitute nonsampling errors that are difficult to deal with and can seriously flaw the quality of the collected data, thus jeopardizing the usefulness of subsequent analyses including statistical inference of unknown characteristics of the population under study. Although these errors cannot be totally avoided, they may be mitigated by enhancing respondents' cooperation. Since the decision to cooperate fully and honestly greatly depends on how survey participants perceive their privacy being disclosed, survey modes which ensure respondents' anonymity or, at least, a high degree of confidentiality, may go some way to improving cooperation and, consequently, ensure more reliable information on sensitive topics than that derived from direct questioning.

In recent years, indirect questioning survey modes have gained popularity in many research fields, mostly falling in the life and social sciences, as effective methods for eliciting truthful responses to sensitive questions while guaranteeing respondents' privacy. In general, this nonstandard survey approach encourages greater cooperation from respondents and reduces the motivation to falsely report their attitudes. The approach obeys the principle that no direct question is posed to survey participants and, then, there is no need for respondents to openly reveal if they are actually engaged in sensitive behaviors. In this way, privacy is protected since answers 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 cannot be used to know their true sensitive status, the information gathered for all the survey participants can be profitable used to make inference on certain parameters of interest of the population under study, usually the prevalence of a sensitive behavior, its frequency or the mean/total of a sensitive quantitative variable.

The indirect questioning strategies may be classified in three different groups: the randomized response technique, the item count technique (ICT), and the nonrandomized response technique. All the approaches have produced a considerable literature and attracted the interest of health, cognitive and behavioral psychologists, epidemiologists, health-care operators, researchers engaged in organizing, managing and conduction sensitive studies, as well as policy-makers committed in formulating effective diseases and mental disorders control measures and promoting public intervention programs to gauge progress toward improving the behavioral health of a state.

For a comprehensive review of the topic, interested readers are referred to Fox and Tracy (1986), Chaudhuri and Mukerjee (1988), Chaudhuri (2011), Chaudhuri and Christofides (2013), Tian and Tang (2014). Useful and detailed studies on recent methodological advances, more complex estimation problems and new challenges may be found, among others, in Arcos et al. (2015), Barabesi et al. (2013, 2015), Diana and Perri (2011), Fox et al. (2014), Glynn (2013), Groenitz (2014), Hoffmann and Musch (2016), Hoffmann et al. (2016), Hussain et al. (2015), Ibrahim (2016), Imai (2011), Imai et al. (2015), Liu and Tian (2013), Moshagen et al. (2014), Nepusz et al. (2014), Perri and van der Heijden (2012), Petróczi et al. (2011), Rueda et al. (2016), Tsuchiya (2005), Ulrich et al. (2012), Wu and Tang (2016).

Various indirect questioning techniques have been experienced in different branches of life sciences. In particular, these methods have been mainly applied to estimate prevalence of discriminating or embarrassing behaviors in epidemiological and medical studies. Some recent contributions, although not exhaustive, cover a great variety of topics. For instance: the measure of the impact of HIV/AIDS infection in Botswana (Arnab and Singh, 2010); the assessment of sensitive health-risk behaviors in HIV/AIDS positive individuals (Arentoft et al., 2016); the assessment of permissive sexual attitudes and high-risk sexual behaviors to reduce the transmission and acquisition of sexually transmitted infections and HIV/AIDS (De Jong et al., 2012; Starosta and Earleywine, 2014; Geng et al., 2016; Kazemzadeh et al., 2016); patterns of condom use among university students for HIV/AIDS control programs (Safiri, 2016; Vakilian et al., 2016); the prevalence of sexual behaviors such as extradyadic sex (Tu and Hsieh, 2017), commercial sex among homosexual men (Chen et al., 2014) and sexual assault (Krebs et al., 2011); the use of drug, and athletic, cognitive and mood performance-enhancing substances (Striegel et al., 2010; Petróczi et al., 2011; Dietz et al., 2013; Franke et al., 2013; James et al., 2013; Nakhaee et al., 2013; Stubbe et al., 2013; Shamsipour et al., 2014; Khosravi et al., 2015; Cobo et al., 2016); smoking behavior validation studies (Fox et al., 2013); dental hygiene habits of Chinese college students (Moshagen et al., 2010); farmers' transgressionary behaviors and prevalence of animal diseases such us sheep scab in Wales (Cross et al., 2010), African swine fever in Madagascar (Randrianantoandro et al., 2015), or foot and mouth disease-infected animals in Sri Lanka (Gunarathne et al., 2016); estimation of the prevalence of induced abortion (Oliveras and Letamo, 2010; Moseson et al., 2015; Perri et al., 2016); ecological and biological conservation issues including estimation of illegal bushmeat hunting (Nuno et al., 2013; Conteh et al., 2015), illegal fishing (Blank and Gavin, 2009; Arias and Sutton, 2013) and unauthorized natural resources use (Harrison et al., 2015).

This article focuses on a recent variant of the ICT conceived to deal with quantitative sensitive variables. We propose some methodological advances that can be useful in life sciences when multiple sensitive issues are to be investigated, and reliable and accurate estimates of usually underreported characteristics are to be produced.

The ICT has recently attracted much attention among applied researchers. This method, also known as the list experiment or the unmatched count technique, was originally proposed by Miller (1984) for binary variables to estimate the prevalence of a stigmatizing behaviour within the population. Without loss of generality, respondents are asked directly about their own sensitive behaviour and, at the same time, about a number of innocuous behaviours. In the standard setting, the method requires the selection of two samples: a reference sample which receives a short list (SL) of items on questions only about innocuous behaviours, and a treatment sample which receives a long list (LL) containing the innocuous items in the SL-sample and a sensitive question. Units selected in the two samples are asked to report the total number of items that apply to them without revealing which item applies individually.

The ICT is used in surveys which require the study of a qualitative variable. Nonetheless, many practical situations may deal with sensitive variables which are quantitative in nature. To address this situation, Chaudhuri and Christofides (2013) proposed a generalization of the ICT that can be used to estimate the mean (or the total) of a quantitative variable. Trappmann et al. (2014) called this variant the item sum technique (IST) and used it in a survey to estimate the amount of undeclared work in Germany. The IST works in a similar way to the ICT and offers a promising tool for dealing with sensitive issues. Nonetheless, some methodological challenges, conceptually inherited from the ICT, remain to be overcome in order to successful use the technique in applied research. The purpose of the present article is to address these challenges. In particular, two open and unresolved issues are discussed. The first pertains the reduction of the statistical burden when multiple sensitive items are to be investigated and estimates of certain characteristics are required. This situation occurs frequently in real studies where researchers must incorporate $Q \geq 2$ sensitive questions in their surveys. Three different approaches are considered in the article, and pros and cons highlighted. The first two techniques require that sampled units participate in Q distinct IST surveys, one for each sensitive item. The first method is time-consuming and costly since requires the selection of $2Q$ samples, the second instead requires Q samples but burdens the surveyed participants. A third viable alternative, which requires the selection of $Q + 1$ samples and acts as a trade-off between the first two approaches, is therefore proposed and its performance investigated on a number of simulation experiments based on real data.

The second, but not less important, problem we consider is how to split the total sample size into the LL-sample and the SL-sample. A simple solution would be to allocate the same number of units to each sample, irrespective of the variability of the items in the two lists. Although intuitive and easy to implement, this basic solution is inefficient because estimates may be affected by high variability. A possible alternative, discussed in the article, would be to achieve optimal sample size allocation by minimizing the variance of the IST estimates under a budget constraint. This possibility is first formalized and discussed under a generic sample design and, then, results are particularized to the simple random sampling and the stratified sampling designs. Optimal allocation results are finally extended to the multiple sensitive estimation setting.

Methodological developments are integrated with an extensive simulation study aimed at investigating the performance of the proposed techniques and the related estimators under two different sampling designs and for different sample sizes. Most of the simulation study is based on the results of a real sensitive research conducted among university students in Granada (Spain) to investigate the consumption of cannabis for recreational purposes.

The rest of this paper is organized as follows: Section 2 introduces the IST under a very general sampling design. Section 3 discusses some estimation methods for multiple sensitive questions under different approaches. The problem of the optimal sample size allocation is then formulated in Section 4. Allocation is first derived for a general setting and then applied to simple random sampling without replacement and stratified sampling design. In Section 5, a number of simulation experiments are generated from two real surveys to investigate the performance of the optimal allocation for single and multiple sensitive estimation under different scenarios. One of the surveys concerns the number of cannabis cigarettes smoked in last year by university students. Section 6 concludes the article with some final considerations.

2 The item sum technique

Consider a finite population $U = \{1, \dots, N\}$ consisting of N different and identifiable units. Let y_i be the value of the sensitive character under study, say $\mathcal{Y}$, for the i-th population unit. Let us suppose that the population mean $\bar{Y} = N^{-1} \sum_{i \in U}^{N} y_i$ is unknown and has to be estimated in an IST setting. In so doing, two independent samples, say s_{ll} and s_{sl}, are selected from U according to the generic sampling designs $p_{ll}(\cdot)$ and $p_{sl}(\cdot)$ with positive first- and second-order inclusion probabilities $\pi_{i(ll)} = \sum_{s_{ll} \ni i} p_{ll}(s_{ll})$,

$\pi_{ij(ll)} = \sum_{s_{ll}\ni i,j} p_{ll}(s_{ll})$, $\pi_{i(sl)} = \sum_{s_{sl}\ni i} p_{sl}(s_{sl})$ and $\pi_{ij(sl)} = \sum_{s_{sl}\ni i,j} p_{sl}(s_{sl})$ with $i,j \in U$. Let $d_{i(ll)} = \pi_{i(ll)}^{-1}$ and $d_{i(sl)} = \pi_{i(sl)}^{-1}$ denote the known sampling design-basic weight for unit $i \in U$ in each sampling design.

Chaudhuri and Christofides (2013) introduced the IST in the following way: one of the samples, say s_{ll}, is confronted with a LL of items containing $(G+1)$ questions of which G refer to nonsensitive characteristics and one is related to the sensitive characteristic under study. The other sample, s_{sl}, receives a SL of items that only contains the G innocuous questions present in the LL-sample. All sensitive and nonsensitive items are quantitative in nature. Respondents in each sample are requested to report the total score of all the items applicable to them, without revealing the individual scores for the items.

Without loss of generality, let $\mathcal{T}$ be the variable denoting the total score applicable to the G nonsensitive questions, and $\mathcal{Z} = \mathcal{Y} + \mathcal{T}$ the total score applicable to the nonsensitive questions and the sensitive question. When $G = 1$, $\mathcal{T}$ denotes the innocuous variable and t_i its value on unit $i \in U$. Hence, the answer given by the i-th respondent will be $z_i = y_i + t_i$ if $i \in s_{ll}$ or t_i if $i \in s_{sl}$.

Under the sampling designs $p_{ll}(\cdot)$, $p_{sl}(\cdot)$, let:

$$\hat{\bar{Z}} = \frac{1}{N}\sum_{i\in s_{ll}} d_{i(ll)} z_i \quad , \quad \hat{\bar{T}} = \frac{1}{N}\sum_{i\in s_{sl}} d_{i(sl)} t_i$$

be the unbiased Horvitz-Thompson (hereafter 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 readily obtained as:

$$\hat{\bar{Y}} = \hat{\bar{Z}} - \hat{\bar{T}}. \tag{1}$$

From the unbiasedness of $\hat{\bar{Z}}$ and $\hat{\bar{T}}$, it readily follows that the estimator $\hat{\bar{Y}}$ is unbiased for $\bar{Y}$. Furthermore, as long as the two samples are independent, the variance of $\hat{\bar{Y}}$ can be expressed as:

$$\begin{aligned} \mathbb{V}(\hat{\bar{Y}}) &= \mathbb{V}(\hat{\bar{Z}}) + \mathbb{V}(\hat{\bar{T}}) \qquad (2) \\ &= \frac{1}{N^2}\left(\sum\sum_{i,j\in U}\Delta_{ij(ll)} d_{i(ll)} d_{j(ll)} z_i z_j + \sum\sum_{i,j\in U}\Delta_{ij(sl)} d_{i(sl)} d_{j(sl)} t_i t_j\right), \end{aligned}$$

where $\Delta_{ij(a)} = \pi_{ij(a)} - \pi_{i(a)}\pi_{j(a)}$ with $a = ll, sl$. An unbiased estimator of $\mathbb{V}(\hat{\bar{Y}})$ is given by:

$$\hat{\mathbb{V}}(\hat{\bar{Y}}) = \frac{1}{N^2}\left(\sum\sum_{i,j\in s_{ll}}\check{\Delta}_{ij(ll)} d_{i(ll)} d_{j(ll)} z_i z_j + \sum\sum_{i,j\in s_{sl}}\check{\Delta}_{ij(sl)} d_{i(sl)} d_{j(sl)} t_i t_j\right)$$

where $\check{\Delta}_{ij(a)} = \Delta_{ij(a)}/\pi_{ij(a)}$.

3 Multiple sensitive estimation under IST

Traditionally, indirect questioning techniques deal with one sensitive variable. However, in real surveys, the researcher may be interested in investigating more than one sensitive variable. Typical areas of inquiry include: (i) the amount of self-employment income and income from financial assets; (ii) the frequency and amount of tax evasion; (iii) the frequency, quantity and cost of cannabis use. In general, in situations like these concerning multiple estimation of the means of $Q > 1$ quantitative sensitive variables, the implementation of the IST may be not unique and cumbersome, for various reasons. To obtain a reliable estimation, a number of solutions might be adopted. One consists in performing Q separate IST surveys, one for each sensitive item. This approach (hereafter, separate approach) requires for each item the selection of one LL-sample and one SL-sample, for a total of $2Q$ samples. In practice, however, this solution does not appear to be feasible, because it is both time-consuming and costly, and also because possible associations between variables would be lost since each IST survey is independently executed on different subjects. To overcome these problems, a single IST survey could be performed. In this case, just one LL-sample and one SL-sample are selected and respondents are asked to participate in Q separate IST experiments, one for each sensitive item. As can be readily imagined, this procedure (hereafter, all-in-one approach) imposes a heavy statistical burden on the respondents, since they must provide the required information on the single sensitive items by separately implementing the IST Q times. More specifically, each respondent belonging to the SL-sample has to answer on Q different short lists and each respondent in the LL-sample has to answer on Q different long lists. If there are many items to be investigated, the accuracy of the responses may deteriorate during the runs. Respondents may be more willing to participate and concentrate more effectively at the beginning of the process, but lose attention during the course of the survey, and possibly break the rules or drop out. If the all-in-one approach is adopted, the order of the items to be investigated, the question of reducing the statistical burden and the problem of respondent drop out must all be carefully considered in the survey design. In view of the manifest weaknesses of the separate and all-in-one approaches, we now consider a possible solution, one providing a trade-off of costs and benefits. Without loss of generality, let us focus, initially, on two quantitative sensitive variables, $\mathcal{Y}_1$ and $\mathcal{Y}_2$, and on one innocuous variable $\mathcal{T}$. We want to estimate the mean of the two variables, say $\bar{Y}_1$ and $\bar{Y}_2$. Under this approach (hereafter, mixed approach), three independent samples are selected. For ease of notation, let us suppose that the same sampling design $p(\cdot)$ is used. Hence, let:

(i) s_0 be a sample of size n_0. The respondents are given a SL containing only the innocuous variable. The i-th respondent provides the score t_{i_0} with $i = 1, \ldots, n_0$;

(ii) s_1 be a sample of size n_1. The respondents are given a list containing one sensitive variable, for instance $\mathcal{Y}_1$, and the innocuous one. The i-th respondent provides the total score $y_{1i_1} + t_{i_1}$ with

$i = 1, \ldots, n_1$;

(iii) s_2 be a sample of size n_2. The respondents are given a list containing the two sensitive variables and the innocuous one. The i-th respondent provides the total score $y_{1i_2} + y_{2i_2} + t_{i_2}$ with $i = 1, \ldots, n_2$.

Let

$$\hat{\bar{Z}}_0 = \frac{1}{N} \sum_{i_0 \in s_0} \frac{t_{i_0}}{\pi_{i_0}}, \quad \hat{\bar{Z}}_1 = \frac{1}{N} \sum_{i_1 \in s_1} \frac{y_{1i_1} + t_{i_1}}{\pi_{i_1}}, \quad \hat{\bar{Z}}_2 = \frac{1}{N} \sum_{i_2 \in s_2} \frac{y_{1i_2} + y_{2i_2} + t_{i_2}}{\pi_{i_2}}.$$

Hence

$$\hat{\bar{Y}}_1^* = \hat{\bar{Z}}_1 - \hat{\bar{Z}}_0$$

is the HT-unbiased estimator of $\bar{Y}_1$ with

$$\mathbb{V}(\hat{\bar{Y}}_1^*) = \mathbb{V}(\hat{\bar{Z}}_1) + \mathbb{V}(\hat{\bar{Z}}_0).$$

Similarly,

$$\hat{\bar{Y}}_2^* = \hat{\bar{Z}}_2 - \hat{\bar{Z}}_1$$

is the HT-unbiased estimator of $\bar{Y}_2$ with

$$\mathbb{V}(\hat{\bar{Y}}_2^*) = \mathbb{V}(\hat{\bar{Z}}_2) + \mathbb{V}(\hat{\bar{Z}}_1).$$

This framework can be readily extended to the case of $Q \geq 2$ sensitive variables, $\mathcal{Y}_1, \ldots, \mathcal{Y}_Q$, by selecting $Q + 1$ samples. With the same notation as in the case $Q = 2$, let:

$$\hat{\bar{Z}}_k = \frac{1}{N} \sum_{i_k \in s_k} \frac{z_{i_k}}{\pi_{i_k}} = \frac{1}{N} \sum_{i_k \in s_k} \frac{\sum_{j=1}^{Q} y_{ji_k} + t_{i_k}}{\pi_{i_k}},$$

with $k = 1, \ldots, Q$. Hence, the estimator

$$\hat{\bar{Y}}_k^* = \hat{\bar{Z}}_k - \hat{\bar{Z}}_{k-1}$$

is the HT-unbiased estimator of $\bar{Y}_k$, $k = 1, \ldots, Q$. The variance of this estimator is given by:

$$\begin{aligned} \mathbb{V}(\hat{\bar{Y}}_k^*) &= \mathbb{V}(\hat{\bar{Z}}_k) + \mathbb{V}(\hat{\bar{Z}}_{k-1}) \\ &= \frac{1}{N^2} \left(\sum\sum_{i,j \in U} \Delta_{ij(k)} d_{i(k)} d_{j(k)} z_i z_j + \sum\sum_{i,j \in U} \Delta_{ij(k-1)} d_{i(k-1)} d_{j(k-1)} z_i z_j \right), \end{aligned}$$

where, slightly changing the notation, $d_{b(a)} = \pi_{b(a)}^{-1}$ and $\Delta_{ij(a)} = \pi_{ij(a)} - \pi_{i(a)}\pi_{j(a)}$, with $b = i, j$ and

$a = 1, \ldots, k$. Accordingly, an unbiased estimator for $\mathbb{V}(\hat{\hat{Y}}_k^*)$ follows as:

$$\hat{\mathbb{V}}(\hat{\hat{Y}}_k^*) = \frac{1}{N^2} \left(\sum\sum_{i_k, j_k \in s_k} \check{\Delta}_{ij(k)} d_{i(k)} d_{j(k)} z_{i_k} z_{j_k} + \sum\sum_{i_{k-1}, j_{k-1} \in s_{k-1}} \check{\Delta}_{ij(k-1)} d_{i(k-1)} d_{j(k-1)} z_{i_{k-1}} z_{j_{k-1}} \right).$$

Similarly, $G > 1$ innocuous variables, say $\mathcal{T}_1, \ldots, \mathcal{T}_G$, can be considered. In this case, $\mathcal{T}$ denotes the total score of the values of the G innocuous variables and $t_{i_k} = \sum_{g=1}^{G} t_{gi_k}$ is the total score of the G innocuous variables for the i_k-th respondent in the k-th sample s_k.

4 Total sample size allocation in the IST estimation

A key design decision in an IST survey is how to split the total sample into the LL-sample and SL-sample. A simple solution is to allocate the same number of units to each sample irrespective of the variability of the items in the two lists. Clearly, this intuitive and basic solution is not efficient because responses in the LL-sample are tendentially affected by high variability due to the presence of innocuous items: the larger the number of items, the higher the variability of the response and, hence, of the estimates. To the best of our knowledge, the problem of optimal allocation in the IST framework has not been considered so far. Therefore, we propose a possible solution to this problem. First, we consider the standard IST with just one sensitive variable, and assume that the total sample size n is fixed beforehand. Hence, the problem of optimal sample allocation is formulated as one of determining the LL-sample and SL-sample sizes, n_{ll} and n_{sl}, in such a way as to minimize the variance of $\hat{\hat{Y}}$ subject to a fixed cost C.

4.1 Allocation under a generic sampling design

Suppose that an IST design has been decided upon. Let n be the sample size of the IST design, or the expected sample size if the sampling design is not of a fixed size. To estimate the population mean $\bar{Y}$, the HT-estimator defined in (1) is considered. Before selecting the sample, the sample sizes n_{ll} and n_{sl} must be determined. We provide a solution to this allocation problem for the case in which the sampling designs $p_{ll}(\cdot)$ and $p_{sl}(\cdot)$ provide a variance of the estimator which can be formulated as:

$$\mathbb{V}(\hat{\hat{Y}}) = \frac{A_z}{n_{ll}} + \frac{A_t}{n_{sl}} + B, \tag{3}$$

where the terms A_z, A_t and B do not depend on n_{ll} and n_{sl}. The simple random sampling and the stratified random sampling designs meet this requirement.

Let c_0 represent the fixed overhead cost of the survey, and $c_{ll} > 0$ and $c_{sl} > 0$ be the costs of surveying one element in s_{ll} and s_{sl}, respectively. These costs depend on the survey designs adopted.

We assume a linear cost function. Hence, the total data-collection cost for the survey is given by:

$$C = c_0 + c_{ll}n_{ll} + c_{sl}n_{sl}. \tag{4}$$

Under this setup, the following result holds.

Theorem 1. *For an IST design which admits $\mathbb{V}(\hat{\hat{Y}})$ in the form given by (3), the optimal sample size allocation under the linear cost function $C = c_0 + c_{ll}n_{ll} + c_{sl}n_{sl}$ is achieved by choosing*

$$n_{ll} = (C - c_0)\frac{\sqrt{A_z/c_{ll}}}{\sqrt{A_z c_{ll}} + \sqrt{A_t c_{sl}}} \quad , \quad n_{sl} = (C - c_0)\frac{\sqrt{A_t/c_{sl}}}{\sqrt{A_z c_{ll}} + \sqrt{A_t c_{sl}}}. \tag{5}$$

The minimum variance of the estimator $\hat{\hat{Y}}$ is

$$\mathbb{V}(\hat{\hat{Y}}) = \frac{1}{C - c_o}\left(\sqrt{A_z c_{ll}} + \sqrt{A_t c_{sl}}\right)^2 + B. \tag{6}$$

Proof
As in Särndal et al. (1992; Section 3.7.3), determining n_{ll} and n_{sl} to minimize $\mathbb{V}(\hat{\hat{Y}})$ for fixed C is equivalent to minimizing the product

$$(\mathbb{V}(\hat{\hat{Y}}) - B)(C - c_0) = \left(\frac{A_z}{n_{ll}} + \frac{A_t}{n_{sl}}\right)(c_{ll}n_{ll} + c_{sl}n_{sl}).$$

From the Cauchy-Schwarz inequality, we obtain:

$$(\mathbb{V}(\hat{\hat{Y}}) - B)(C - c_0) \geq \left(\sqrt{A_z c_{ll}} + \sqrt{A_t c_{sl}}\right)^2,$$

where the equality holds if and only if:

$$\sqrt{\frac{c_{ll}n_{ll}}{\frac{A_z}{n_{ll}}}} = \sqrt{\frac{c_{sl}n_{sl}}{\frac{A_t}{n_{sl}}}} = K.$$

From the previous equality, it follows that

$$n_{ll} = K\sqrt{\frac{A_z}{c_{ll}}} \quad , \quad n_{sl} = K\sqrt{\frac{A_t}{c_{sl}}}. \tag{7}$$

By replacing these quantities in the budget constraint (4), we obtain the value of K as:

$$K = \frac{c_{ll} n_{ll} + c_{sl} n_{sl}}{\sqrt{A_z c_{ll}} + \sqrt{A_t c_{sl}}} = \frac{C - c_0}{\sqrt{A_z c_{ll}} + \sqrt{A_t c_{sl}}},$$

which, when replaced in (7), yields (5). Hence, with this optimal choice of n_{ll} and n_{sl}, the quantity $(\mathbb{V}(\hat{\hat{Y}}_{HT}) - B)(C - c_0)$ attains its minimum value $\left(\sqrt{A_z c_{ll}} + \sqrt{A_t c_{sl}}\right)^2$ or, equivalently, $\mathbb{V}(\hat{\hat{Y}})$ achieves the minimum variance bound given in (6). Hence the proof. $\square$

In terms of the sample size $n = n_{ll} + n_{sl}$, from (5) we have

$$n = (C - c_0) \frac{\sqrt{A_z / c_{ll}} + \sqrt{A_t / c_{sl}}}{\sqrt{A_z c_{ll}} + \sqrt{A_t c_{sl}}},$$

from which it easily follows that:

$$n_{ll} = n \frac{\sqrt{A_z / c_{ll}}}{\sqrt{A_z / c_{ll}} + \sqrt{A_t / c_{sl}}} \quad , \quad n_{sl} = n \frac{\sqrt{A_t / c_{sl}}}{\sqrt{A_z / c_{ll}} + \sqrt{A_t / c_{sl}}}.$$

Hence, the following result is proved:

Corollary 1. *If the sample costs c_{ll} and c_{sl} are equal, the optimal sample size allocation is given by:*

$$n_{ll} = n \frac{\sqrt{A_z}}{\sqrt{A_z} + \sqrt{A_t}} \quad , \quad n_{sl} = n \frac{\sqrt{A_t}}{\sqrt{A_z} + \sqrt{A_t}}. \tag{8}$$

We observe that the calculation of n_{ll} and n_{sl} given in (8) requires the knowledge of A_z and A_t. These quantities generally depend on the population variances which are usually unknown. When such values are unknown and cannot be properly guessed on the basis of previous data or experts opinion, they must be estimated making use, for instance, of a pilot survey (Sukhatme et al., 1984).

4.2 Allocation under simple random sampling without replacement

Let us suppose that the two samples s_{ll} and s_{sl} are selected according to simple random sampling without replacement (SRSWOR) and that all costs are equal. Hence, from (2), the variance of $\hat{\hat{Y}}$ can be reformulated as in (3):

$$\begin{aligned}
\mathbb{V}(\hat{\hat{Y}}) &= \mathbb{V}(\hat{\hat{Z}}) + \mathbb{V}(\hat{\hat{T}}) \\
&= \left(1 - \frac{n_{ll}}{N}\right) \frac{S_z^2}{n_{ll}} + \left(1 - \frac{n_{sl}}{N}\right) \frac{S_t^2}{n_{sl}} \\
&= \frac{S_z^2}{n_{ll}} + \frac{S_t^2}{n_{sl}} - \frac{1}{N}(S_z^2 + S_t^2),
\end{aligned}$$

where $S_{\cdot}^2$ denotes the population variance of the variables in the subscript. Note that $S_z^2 = S_y^2 + S_t^2 + S_{yt}$, where S_{yt} denotes the covariance. By replacing these population quantities by their sampling counterpart, we obtain an unbiased estimator of $\mathbb{V}(\hat{\hat{Y}})$ as:

$$\hat{\mathbb{V}}(\hat{\hat{Y}}) = \frac{s_z^2}{n_{ll}} + \frac{s_t^2}{n_{sl}} - \frac{1}{N}(s_z^2 + s_t^2)$$

where $s_{\cdot}^2$ denotes the sample variance.

Finally, from (8), we have:

$$\gamma = \frac{n_{ll}}{n} = \frac{S_z}{S_z + S_t} = \frac{\sqrt{S_y^2 + S_t^2 + 2S_{yt}}}{\sqrt{S_y^2 + S_t^2 + 2S_{yt}} + S_t},$$

$$1 - \gamma = \frac{n_{sl}}{n} = \frac{S_t}{S_z + S_t} = \frac{S_t}{\sqrt{S_y^2 + S_t^2 + 2S_{yt}} + S_t}.$$

Clearly, if the correlation between the sensitive and the innocuous variables is positive, the LL-sample will be larger than the SL-sample. This is because the responses given in the LL-sample are expected to have a larger variance, which must be compensated with a larger sample size. Moreover, the function γ is: (i) an increasing function of S_y; (ii) a decreasing function of S_t; (iii) an increasing function of S_{yt}. Figure 1 shows the behaviour of γ as a function of $S_y =$ 10, 20,...,1000 and $S_t =$ 10, 20,...,1000 for $\rho_{yt} = S_{yt}/S_y S_t =$ 0.5.

4.3 Allocation under a stratified sampling design

In the case of a stratified design, let the population U be divided into H strata. Let N_h denote the size of the h-th stratum, say U_h, and $W_h = N_h/N$ be the weight of U_h in the population, $h = 1, \ldots, H$. From the stratum U_h, two samples $s_{h(ll)}$ and $s_{h(sl)}$ of sizes $n_{h(ll)}$ and $n_{h(sl)}$ are selected according to SRSWOR. The sampled elements in $s_{h(ll)}$ are confronted with the LL of items while those in $s_{h(ll)}$ are

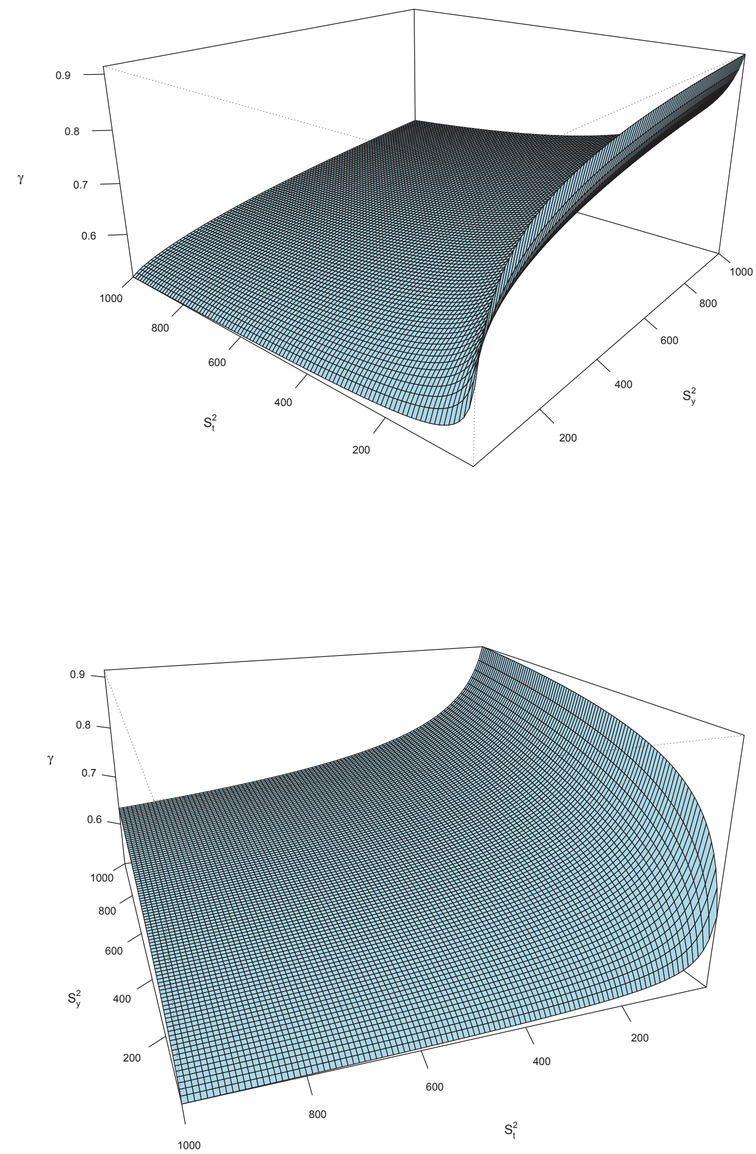


Figure 1: Behavior of γ for different values of S_y^2 and S_t^2 and $\rho_{yt} = 0.5$.

confronted with the SL of items. Under stratified SRSWOR, expression (2) takes the form:

$$\mathbb{V}(\hat{\hat{Y}}_{str}) = \sum_{h=1}^{H} W_h^2 \left(1 - \frac{n_{h(ll)}}{N_h}\right) \frac{S_{h,z}^2}{n_{h(ll)}} + \sum_{h=1}^{H} W_h^2 \left(1 - \frac{n_{h(sl)}}{N_h}\right) \frac{S_{h,t}^2}{n_{h(sl)}}, \tag{9}$$

where $S^2_{h,\cdot}$ is the variance in the stratum h.

As in Theorem 1, minimizing (9) subject to $\sum_{h=1}^{H}(n_{h(ll)} + n_{h(sl)}) = n$ with equal cost gives the following optimal sample size allocation for the stratum U_h:

$$n_{h(ll)} = n\frac{S_{h,z}W_h}{\sum_{h=1}^{H} S_{h,z}W_h + \sum_{h=1}^{H} S_{h,t}W_h} \quad , \quad n_{h(sl)} = n\frac{S_{h,t}W_h}{\sum_{h=1}^{H} S_{h,z}W_h + \sum_{h=1}^{H} S_{h,t}W_h}.$$

Consequently:

$$\gamma_h = \frac{n_{h(ll)}}{n} = \sum_{h=1}^{H} W_h \frac{\sqrt{S^2_{h,y} + S^2_{h,t} + 2S_{h,yt}}}{\sum_{h=1}^{H} W_h\sqrt{S^2_{h,y} + S^2_{h,t} + 2S_{h,yt}} + \sum_{h=1}^{H} W_h S_{h,t}}$$

and

$$1 - \gamma_h = \frac{n_{h(sl)}}{n} = \sum_{h=1}^{H} W_h \frac{S_{h,t}}{\sum_{h=1}^{H} W_h\sqrt{S^2_{h,y} + S^2_{h,t} + 2S_{h,yt}} + \sum_{h=1}^{H} W_h S_{h,t}}.$$

4.4 Allocation in multiple IST estimation

Determining optimal sample size allocation is of particular importance in the multiple IST estimation introduced in Section 3 where, under the separate and mixed approaches, more than two samples will be selected. Optimal allocation is easily achieved under the separate approach by applying the results of the previous sections to each sensitive variable under study. In other words, optimal sample size allocation is obtained for each IST survey by minimizing the variance of the estimator of the sensitive mean corresponding to the variable referred to by the IST survey. For the other approaches, the problem is slightly different but can be solved by extending the results of the previous sections after having specified the expression of the variance to be minimized. Let us first discuss the all-in-one procedure. In this case, just one sample is selected for the entire survey on the Q sensitive questions. This sample must then be optimally split into the LL-sample and SL-sample, and so the initial question is to decide how this optimality is to be achieved. One possibility is to focus on one of the Q sensitive variables, perhaps the most relevant variable - if any - for the survey, and then to minimize the variance of the estimator of its mean. Obviously, however, obtaining the optimal sample size allocation for the variable considered does not ensure variance reduction in estimating the mean of the remaining variables. To overcome this limitation, a more general solution that involves all the study variables might be considered. Since multiple estimation leads to Q estimators of the Q population means of the sensitive variables under investigation, we may opt to minimize the variance of a convex combination of the Q variances of the estimators. Without loss of generality, let $\hat{\hat{Y}}_k = \hat{\hat{Z}}_k - \hat{\hat{T}}_k$ denote the estimator of the population mean $\bar{Y}_k$ for the sensitive variable $\mathcal{Y}_k$, $k = 1, \ldots, Q$. The meaning of $\hat{\hat{Z}}_k$ and $\hat{\hat{T}}_k$ follows accordingly. Hence, the optimal sample sizes n_{ll} and n_{sl} are obtained by minimizing:

$$\mathbb{V}_\alpha = \sum_{k=1}^{Q} \alpha_i \mathbb{V}(\hat{\hat{Y}}_k),$$

with $\sum_{k=1}^{Q} \alpha_i = 1$. For instance, under SRSWOR, for $Q = 2$ sensitive variables, say $\mathcal{Y}_1$ and $\mathcal{Y}_2$, and $G = 2$ innocuous variables, say $\mathcal{T}_1$ and $\mathcal{T}_2$, we have:

$$\mathbb{V}_\alpha = \frac{1}{n_{ll}} \left(\alpha_1 S^2_{z_1} + \alpha_2 S^2_{z_2}\right) + \frac{1}{n_{sl}} \left(\alpha_1 S^2_{t_1} + \alpha_2 S^2_{t_2}\right) - \frac{1}{N} \left[\alpha_1 \left(S^2_{z_1} + S^2_{t_1}\right) + \alpha_2 \left(S^2_{z_2} + S^2_{t_2}\right)\right]. \quad (10)$$

For the mixed approach, finding the optimal sample size allocation by minimizing the variance of one estimator is unfeasible since this will allocate the entire total size n between two samples, leaving a zero size for the remaining $Q - 1$ samples. The only solution to this problem is to minimize the convex combination of the Q variances of the estimators:

$$\begin{aligned} \mathbb{V}_\alpha &= \alpha_1 \mathbb{V}(\hat{\hat{Y}}_1^*) + \alpha_2 \mathbb{V}(\hat{\hat{Y}}_2^*) \\ &= \alpha_1 \mathbb{V}(\hat{\hat{Z}}_0) + \mathbb{V}(\hat{\hat{Z}}_1) + \alpha_2 \mathbb{V}(\hat{\hat{Z}}_2). \end{aligned}$$

For instance, if the samples are selected according to SRSWOR, $Q = 2$ and $G = 1$, we have:

$$\mathbb{V}_\alpha = \frac{1}{n_0} \alpha_1 S^2_t + \frac{1}{n_1} S^2_{z_1} + \frac{1}{n_2} \alpha_2 S^2_{z_2} - \frac{1}{N} \left(\alpha_1 S^2_t + S^2_{z_1} + \alpha_2 S^2_{z_2}\right). \quad (11)$$

Note that the choice $\alpha = 0.5$ is equivalent to minimizing $\mathbb{V}(\hat{\hat{Y}}_1^*) + \mathbb{V}(\hat{\hat{Y}}_2^*)$.

5 Simulation

5.1 Simulation design

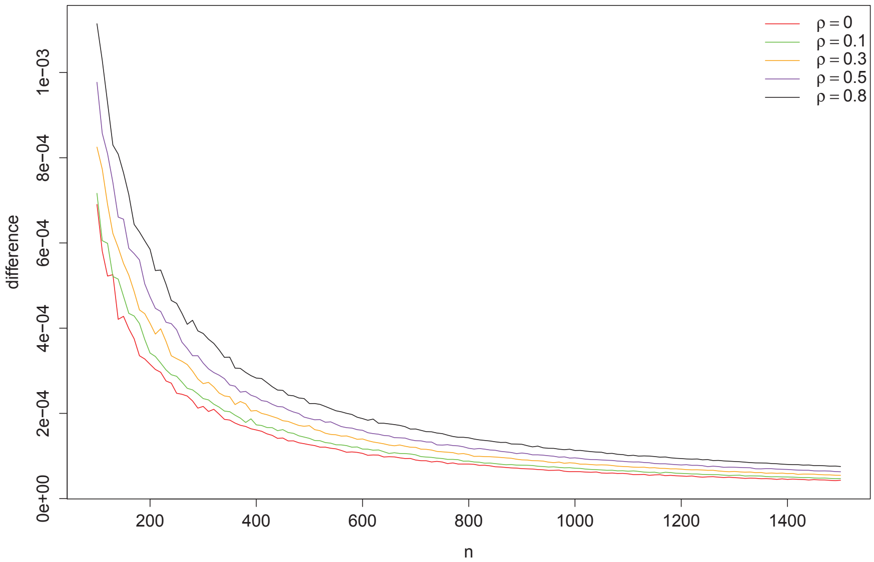
In this section, we run a number of simulation studies to evaluate the performance of the optimal allocation discussed above. To do so, $N = 52409$ artificial observations are generated for the sensitive variable $\mathcal{Y}$ and the innocuous one $\mathcal{T}$. It is assumed that $(\mathcal{Y}, \mathcal{T})$ are observed from a bivariate normal distribution with different values of the correlation coefficient $\rho_{yt} = \rho$, and with mean and standard error vectors $\boldsymbol{\mu} = (3.114, 7.446)$ and $\boldsymbol{\sigma} = (0.604, 0.049)$, respectively. The values generated are then used to define the total score variable $\mathcal{Z} = \mathcal{Y} + \mathcal{T}$ and to obtain an estimate of $\bar{Y}$ using: (i) the values of $\mathcal{Y}$ in a standard HT-estimator as obtained by direct questioning; (ii) the values of $\mathcal{T}$ and $\mathcal{Z}$ in the HT-estimator as defined in (1). Hence, for each simulation study, we evaluate the estimated variance of the estimators for $B = 1000$ runs and for different sample sizes. Throughout the simulation, the costs are assumed to be constant.

The values for $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ are taken from a real sensitive research conducted at the University of Granada in the academic year 2015/2016 to investigate the *consumption of cannabis*, using the IST. During the class time break, a sample of students were invited to participate in the study and to fill in a questionnaire. Some of these students (492) were directly posed the sensitive question: "*How many cannabis cigarettes did you consume last year?*". The remaining students (1293) were asked to provide data using the IST. In the IST survey, 773 students were arbitrarily allocated to the LL-sample and 520 to the SL-sample. The values $\mu = 3.114$ and $\sigma = 0.604$ represent the estimated mean and the estimated standard error of the sample mean for the number of cannabis cigarettes smoked. Similarly, the values $\mu = 7.446$ and $\sigma = 0.049$ refer to the estimates of the innocuous variable in the SL-sample. The innocuous variable $\mathcal{T}$ is represented by the students score in the university entrance examinations (general stage score, ranging from 0 to 10). As a referee noted, the choice of this innocuous variable may not have sufficiently protected respondents' privacy especially when the number of cannabis cigarettes smoked is "large", for instance more than 50 cigarettes. Indeed, from the collected data, we observed that students who released IST responses (total scores) higher than 10 and 50 were 24.5% and 6.5%, respectively, and that nonresponse rate was very low (1.93%).

It is worthy noting that, according to the IST, 14.931 cannabis cigarettes were smoked on average, a value significantly higher than that obtained by direct questioning (one-tailed *t*-test, *p*-value < 0.001).

5.2 Direct questioning vs optimal allocation IST estimates

In this first study, the samples are selected according to SRSWOR and the variance of the sample mean estimator $\bar{y} = \sum_{i\in s} y_i/n$ is compared with that of the IST estimator with optimal sample size allocation, performed on the same sample size n, as described in Section 4.2. Figure 2 illustrates the difference and the ratio between the estimated variances of the two estimators. Both the difference and the ratio are presented as mean values computed over $B = 1000$ replications. As expected, the variance of the IST estimator is higher than that of the sample mean estimator under direct questioning. The difference becomes negligible as the sample size increases, while the ratio highlights the fact that the loss of efficiency remains within acceptable limits especially when ρ is low. Moreover, for a fixed sample size, the difference (ratio) increases with ρ. The fact that the difference and the loss of efficiency are fairly modest values makes it clear that the optimal IST could provide estimates which are nearly as accurate as those obtained by direct questioning, and without jeopardizing respondents' confidentiality. This finding is of major importance in appraising the use of the IST in real surveys.



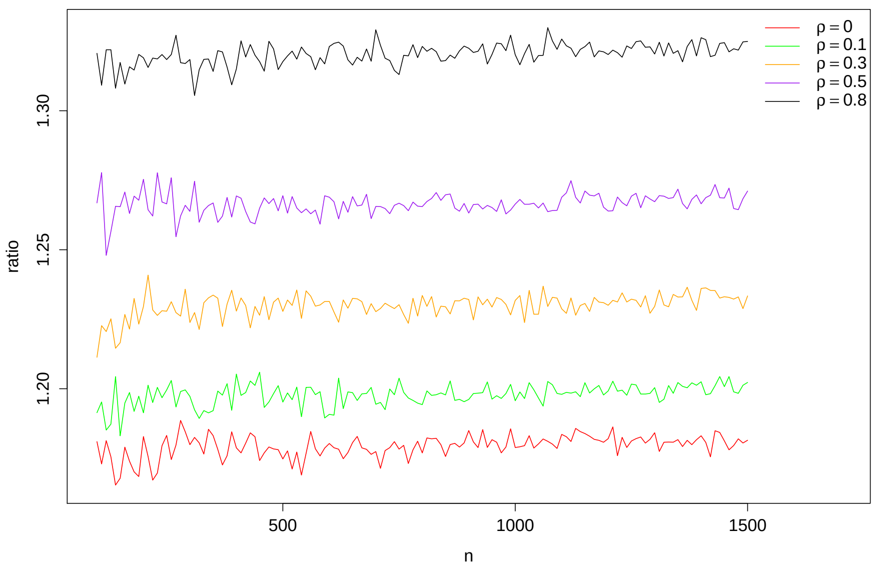


Figure 2: Difference and ratio between the variance of direct questioning estimates and optimal allocation IST estimates.

5.3 Optimal vs arbitrary IST allocation

In SRSWOR, we now examine the efficiency gains that can be obtained when the IST allocation is optimal. To illustrate the magnitude of the increased efficiency, we consider the ratio between the variance of the optimal allocation IST estimator and that of the IST estimator arbitrarily obtained assuming $n_{ll} = \lambda n$ and $n_{sl} = (1 - \lambda)n$, $\lambda = 0.5, 0.6$. The results are shown in Figure 3. The improved efficiency is evident in both situations. As also shown in Figure 2, the correlation coefficient does not appear to significantly affect the variance of the IST estimators and, consequently, the efficiency gain from the optimal allocation.

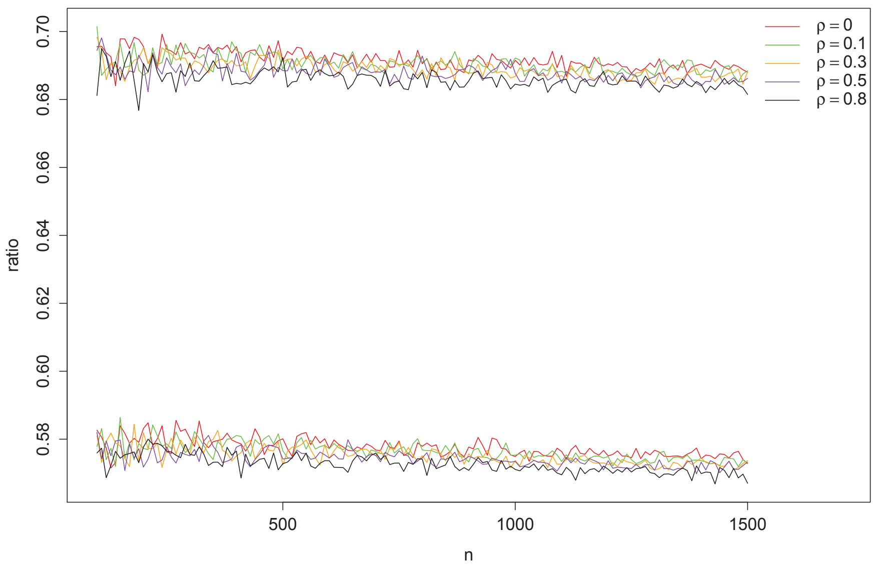


Figure 3: Ratio between the variance of the optimal allocation IST estimator and the variance of the IST estimator with $n_{ll} = \lambda n$ under arbitrary allocation. The upper plots refers to $\lambda = 0.6$ and the lower to $\lambda = 0.5$.

5.4 Optimal IST allocation in stratified SRSWOR

We now examine the case in which stratified SRSWOR is adopted. We assume that the $N = 52409$ students of the University of Granada (see Section 5.1) are stratified into two groups - male (*M*) and female (*F*) - with weights $W_M = 0.442$ and $W_F = 0.558$ known from administrative sources. Under the same framework as in Section 5.1, for the male group we generate $N_M = 23151$ observations from the bivariate normal distribution $(\mathcal{Y}, \mathcal{T})$ with different values of ρ, $\boldsymbol{\mu}_M = (6.340, 7.507)$ and $\boldsymbol{\sigma}_M = (1.431, 0.072)$. Similarly, for the female stratum ($N_F = 29258$), we assume $\boldsymbol{\mu}_F = (0.240, 7.408)$ and $\boldsymbol{\sigma}_F = (0.121, 0.067)$. As in Section 5.1, the entries of the vectors $\boldsymbol{\mu}_.$ and $\boldsymbol{\sigma}_.$ represent the estimated means and the estimated standard errors of the unknown population means of the sensitive variable and the innocuous variable computed from the male/female direct questioning samples and for the male/female SL-samples, respectively.

The minimum variance estimator of the stratified IST estimator is achieved by using the optimal sample size allocation given in Section 4.3. The variance of the estimates under optimal allocation is then compared using two different forms of allocation:

(i) **Arbitrary allocation.** In stratified IST with two strata (U_M and U_F), four samples are considered. From the U_M stratum, the LL-sample and the SL-sample are selected. Similarly, for the U_F stratum. Let $n_{ll|.}$ and $n_{sl|.}$ be the sample sizes in the respective groups. Hence, we trivially assume: $n_{ll|M} = n_{sl|M} = n_{ll|F} = n_{sl|F} = n/4$.

(ii) **Naive two-step optimal allocation.** Allocation is conducted in two steps, separately determining the optimal IST allocation in one sample of men and in another of women. In the first step, a stratified sample of male and female students is selected with proportional allocation (see, e.g, Särndal et al., 1992). In the second step, each of the two first-step samples is optimally allocated in the LL-sample and SL-sample according to (8).

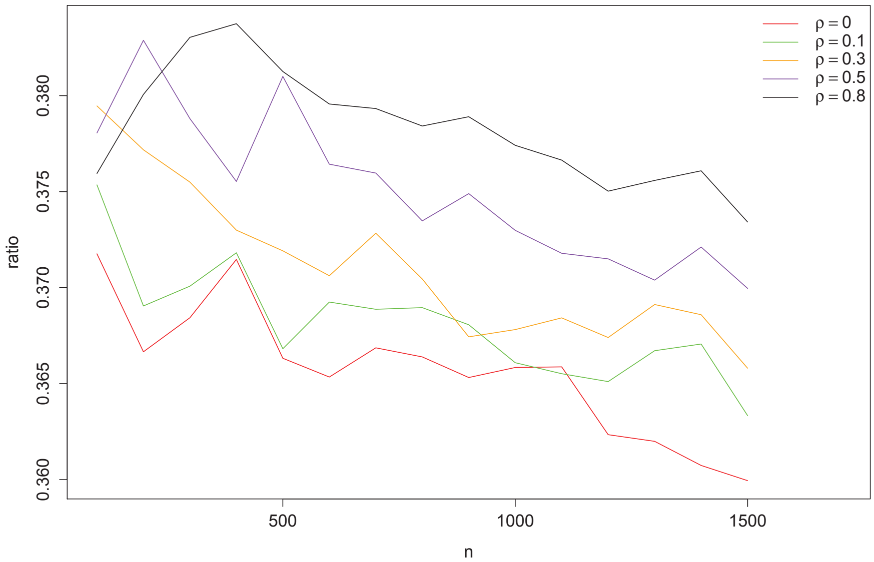
Figure 4 shows the ratio between the variances of the optimal and non-optimal allocation stratified IST estimators. It can be seen that arbitrary allocation is not at all efficient, while the results obtained with two-step allocation are almost identical to those attainable with the theoretical optimal allocation.

Finally, we compared the efficiency of stratified and SRSWOR IST estimates under optimal allocation. The results shown in Figure 5 reflect the considerable gain in efficiency achieved by stratifying the population.

5.5 Optimal allocation in multiple IST estimation

In this section, we investigate multiple IST estimation under each of the approaches discussed in Section 3. The simulation study is based on real data from the *Survey of Household Income and Wealth* (SHIW) conducted by the Bank of Italy (2014). The survey covers 8156 households composed of 19366 individuals. We assume the 8156 households as the target population and focus on two sensitive variables: (i) *net disposable income* ($\mathcal{Y}_1$), and (ii) *net wealth* ($\mathcal{Y}_2$). For all the households surveyed, the values of these and other variables are known.

The aim of this simulation study is to compare the IST estimates of $\bar{Y}_1$ and $\bar{Y}_2$ under the separate, all-in-one and mixed approaches by assuming that $\mathcal{T}$ =*consumption* is the innocuous variable for implementing the IST. From the available data, we know that $\bar{Y}_1 = 31248$ euro, $\bar{Y}_2 = 236097$ euro and these values are used as benchmarks. Under the separate approach, the optimal sample allocation for n_{ll} and n_{sl} is separately considered for each of the two variables in such a way that the estimates for $\bar{Y}_1$ and $\bar{Y}_2$ both attain their minimum variance bound. The all-in-one estimates are obtained assuming that data on both the variables are collected by performing the IST twice on the same units belonging to the only sample selected. The optimal sample sizes n_{ll} and n_{sl}, which minimize (10) with $\alpha = 0.5$, are used to



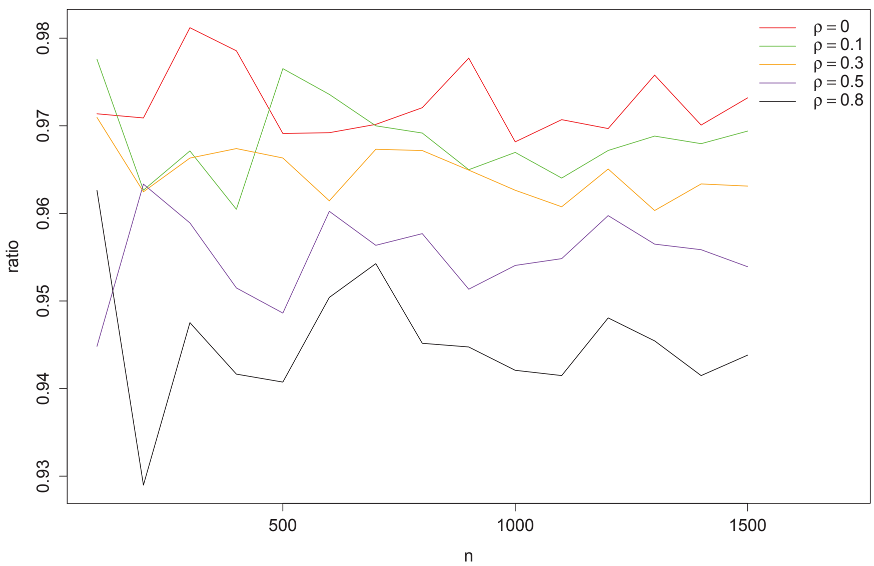


Figure 4: Ratio between variance under optimal allocation and under: (i) arbitrary allocation (upper plot), (ii) naive two-step optimal allocation (lower plot).

obtain the estimates of $\bar{Y}_1$ and $\bar{Y}_2$. Obviously, using n_{ll} and n_{sl} does not ensure that minimum variance is achieved for $\hat{\bar{Y}}_1$ and $\hat{\bar{Y}}_2$. A similar procedure is employed for the mixed approach. In this case, the three sample sizes n_0, n_1 and n_2 are optimally determined to minimize (11) with $\alpha = 0.5$ and then used in the

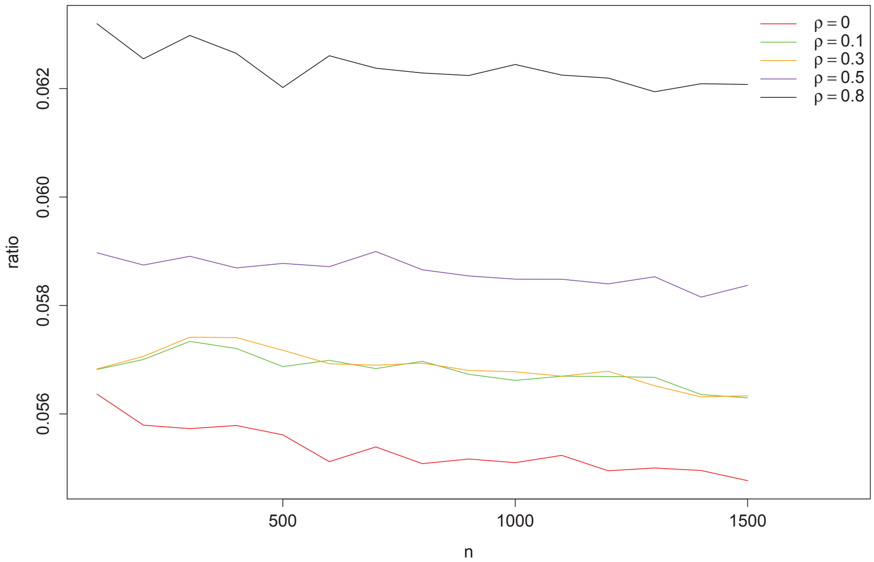


Figure 5: Ratio of optimal allocation variances under stratified IST and SRSWOR IST

single estimators $\hat{\bar{Y}}_1^*$ and $\hat{\bar{Y}}_2^*$. We specify that in all situations the optimal allocation has been achieved by minimizing the estimated variance.

For different sample sizes and $B = 1000$ replications, we investigate the performance of the estimators under the three approaches by means of the absolute relative bias (ARB) and the relative variance (RV):

$$\text{ARB}(\hat{\theta}_i) = \frac{\sum_{k=1}^{B} |\hat{\theta}_i^{(k)} - \bar{Y}_i|}{B\bar{Y}_i} \quad , \quad \text{RV}(\hat{\theta}_i) = \frac{\sum_{k=1}^{B} (\hat{\theta}_i^{(k)} - \bar{Y}_i)^2}{B\bar{Y}_i^2}$$

with $\hat{\theta}_i^{(k)}$ denoting the estimate of $\bar{Y}_i$ on the k-th sample selected from the SHIW target population according to SRSWOR.

The outcomes of the simulation are summarized in Figure 6. It is immediately apparent that both the ARB and the RV decrease as the sample size increases, which is a clear indication of the consistency of the estimates under the three approaches. The three approaches produce equivalent results in estimating the mean of $\mathcal{Y}_2 =$ *wealth*, while for $\mathcal{Y}_1 =$ *income* the separate approach seems to slightly outperform the others, especially for usual sample sizes. As the sample size increases, the difference between the methods decreases. However, on the whole there are no striking differences and for the situations considered in this analysis, the mixed approach seems to be competitive in terms of efficiency while clearly reducing the statistical burden on the respondents. We then replicated the simulation study by directly comparing the theoretical estimated variances of the estimators of $\bar{Y}_1$ and $\bar{Y}_2$ under the three approaches. Figure 7 shows the behaviour of the estimated relative variance (ERV), obtained by dividing the estimated variance of $\hat{\theta}_i$ by $\bar{Y}_i^2$, $i = 1, 2$. The results obtained confirm those for the RV reported in Figure 6. We conclude, therefore, that multiple estimation may be profitably pursued via different approaches and that a useful trade-off between efficiency in the estimates and reducing the statistical burden may be achieved by using the mixed approach with optimal allocation. The findings of this study may therefore be of major significance to survey statisticians and practitioners to support the use of the IST in real-world studies.

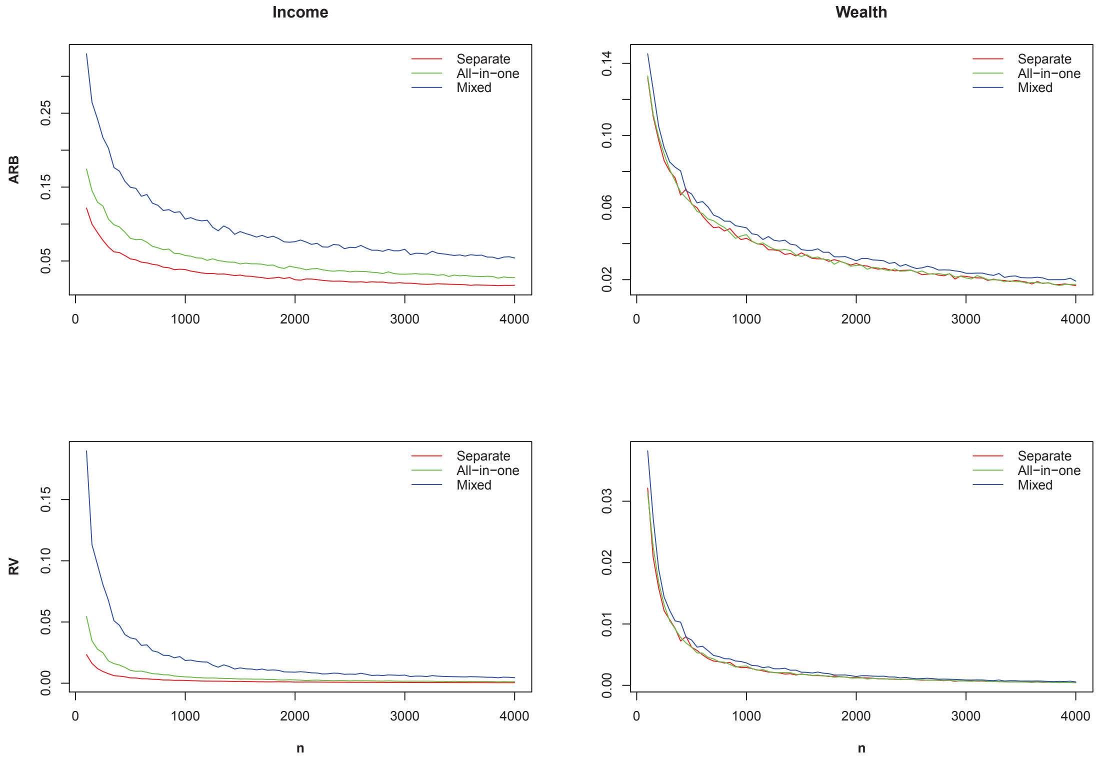


Figure 6: Performance of the estimates under the three IST approaches for multiple estimates purposes. Results based on Monte Carlo simulated ARB and RV.

6 Conclusions

The IST enables us to estimate the mean (or the total) of stigmatizing quantitative variables using an indirect questioning approach, thus reducing nonresponse rates and social desirability response bias. This method is closely related to the ICT, which was developed to measure the proportion of dichotomous

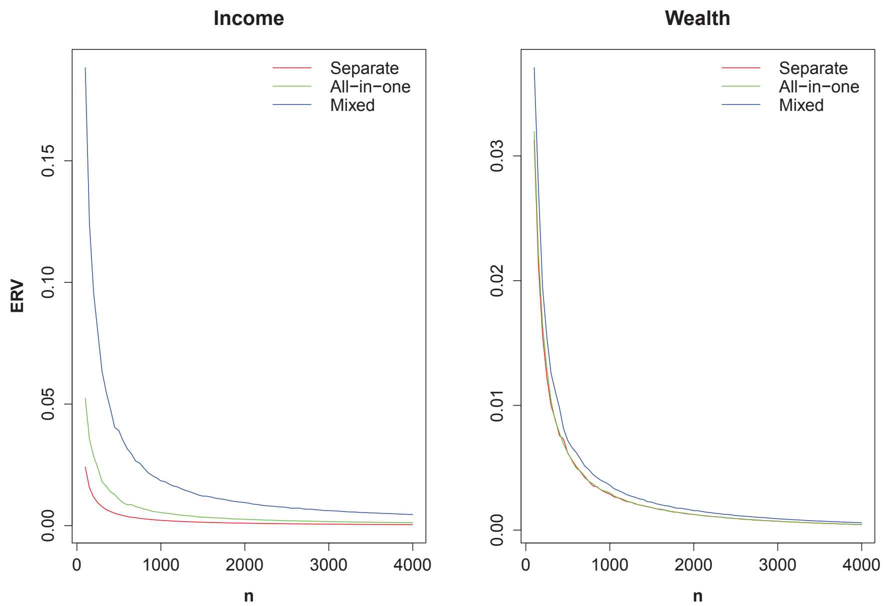


Figure 7: Performance of the estimates under the three IST approaches for multiple estimates purposes. Results based on estimated theoretical variances.

sensitive items in human population surveys.

In this article, we presented certain methodological advances in the use of the IST, and discussed two open questions. First, we considered the problem of how to reduce the statistical burden on respondents when $Q \geq 2$ sensitive variables are surveyed and the population means need to be estimated. Three ways of applying the IST have been discussed. The first of these, the separate approach, requires that for each sensitive item one LL-sample and one SL-sample be selected, i.e. in total, $2Q$ samples are used. In the second approach, termed all-in-one, one LL-sample and one SL-sample are selected and the respondents are asked to participate in Q distinct IST surveys, one for each sensitive item. The separate approach is time-consuming and costly, while the all-in-one approach places an excessive burden on the survey participants that could even induce them to break the rules of the IST or to drop out of the survey. Given the weaknesses of these two approaches, a viable alternative providing a possible trade-off could be pursued. A mixed approach, which requires the selection of $Q + 1$ independent samples, has been therefore proposed, and its performance investigated through a number of simulation experiments based on optimal sample size allocation.

The optimal allocation of the total sample size into the LL-sample and the SL-sample is the second, but no less important, issue discussed in this article. First, we considered a method of allocation based on minimizing the variance of the IST estimator of the mean of one sensitive variable which is valid under a budget constraint and for a general sampling design. Thus, explicit expressions for the sampling fractions have been worked out when SRSWOR and stratified sampling are used. The allocation method has been then extended to the case of Q sensitive variables under the all-in-one and mixed approaches.

An extensive simulation study has been conducted to investigate the performance of the proposed techniques and the related estimators under different sampling designs and for different sample sizes. All the situations examined reflect the benefits of determining the optimal sample size, which can significantly increase the efficiency of the estimates with respect to any arbitrary allocation of the sample units.

A very interesting result has been achieved when optimal allocation is used for multiple IST estimation purposes under the mixed approach. In this case, in relation to the marked reduction obtained in the statistical burden placed on respondents and in survey costs, the loss of efficiency with respect to the all-in-one and separate approaches may be considered very modest or even negligible. Hence, from a theoretical standpoint, the mixed approach appears to be a viable alternative for the purposes of multiple IST estimation. That said, final users interested in experiencing multiple IST have enough elements to critically evaluate the feasibility of the different procedures and to weight between pros and cons with regards to costs, time effort, respondents' burden, and accuracy.

We conclude by observing that all the ideas, the methodological advances and the results presented in this article regarding the IST may be easily extended to its forerunner, the ICT, which, although it is a more widespread and long-established technique, suffers from the same drawbacks that are discussed in this article with respect to the IST and that, to our knowledge, have not yet been addressed. Hence, the value of this article is twofold.

Acknowledgement

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).

Conflict of Interest

The authors have declared no conflict of interest.

References

[1] Arcos, A., Rueda, M., Singh, S. (2015). Generalized approach to randomized response for quantitative variables. *Quality & Quantity*, **49**, 1239–1256.

[2] Arentoft, A., Van Dyk, K., Thames, A.D., Sayegh, P., Thaler, N., Schonfeld, D., LaBrie, J., Hinkin,

C.H. (2016). Comparing the unmatched count technique and direct self-report for sensitive health-risk behaviors in HIV+ adults. *AIDS Care*, **28**, 370–375.

[3] Arias, A. and Sutton, S.G. (2013). Understanding recreational fishers compliance with no-take zones in the Great Barrier Reef Marine Park. *Ecology and Society*, **18**. Available at http://dx.doi.org/10.5751/ES-05872-180418.

[4] Arnab, R. and Singh, S. (2010). Randomized response techniques: An application to the Botswana AIDS impact survey. *Journal of Statistical Planning and Inference*, **140**, 941–953.

[5] Bank of Italy. (2014). Survey of household income and wealth. Available at https://www.bancaditalia.it/statistiche/tematiche/indagini-famiglie-imprese/bilanci-famiglie/index.html?com.dotmarketing.htmlpage.language=1.

[6] Barabesi, L., Diana, G., Perri, P.F. (2013). Design-based distribution function estimation for stigmatized populations. *Metrika*, **76**, 919–935.

[7] Barabesi, L., Diana, G., Perri, P.F. (2015). Gini index estimation in randomized response surveys. *Asta-Advances in Statistical Analysis*, **99**, 45–62.

[8] Blank, S.G. and Gavin, M.C. (2009). The randomized response technique as a tool for estimating non-compliance rates in fisheries: A case study of illegal red abalone (Haliotis rufescens) fishing in Northern California. *Environmental Conservation*, **36**, 112–119.

[9] Chaudhuri, A. (2011). *Randomized Response and Indirect Questioning Techniques in Surveys.* Chapman & Hall/CRC, Boca Raton, FL.

[10] Chaudhuri, A. and Christofides, T.C. (2013). *Indirect Questioning in Sample Surveys.* Springer-Verlag, Berlin, Heidelberg.

[11] Chaudhuri, A. and Mukerjee, R. (1988). *Randomized Response: Theory and Techniques.* Marcel Dekker, INC., New York.

[12] Chen, X., DU, Q., Jin, Z., Xu, T., Shi, J., Gao, G. (2014). The randomized response technique application in the survey of homosexual commercial sex among men in Beijing. *Iran Journal of Public Health*, **43**, 416–422.

[13] Cobo, B., Rueda, M.D.M., López-Torrecillas, F. (2016). Application of randomized response techniques for investigating cannabis use by Spanish university students. *International Journal of Methods in Psychiatric Research*, first published online, doi: 10.1002/mpr.1517.

[14] Conteh, A., Gavin M.C., Solomon, J. (2015). Quantifying illegal hunting: A novel application of the randomised response technique. *Biological Conservation*, **189**, 16–23.

[15] Cross, P., Edwards-Jones, G., Omed, H., Williams. A.P. (2010). Use of a randomized response technique to obtain sensitive information on animal disease prevalence. *Preventive Veterinary Medicine*, **96**, 252–262.

[16] 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**, 543–64.

[17] Diana, G. and Perri, P.F. (2011). A class of estimators for quantitative sensitive data. *Statistical Papers*, **52**, 633–650.

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

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

[20] Fox, J.A. and Tracy, P.E. (1986). *Randomized Response: A Method for Sensitive Survey*. Sage Publication, Inc., Newbury Park.

[22] Fox, J.P., Avetisyan. M., van der Palen, J. (2013). Mixture randomized item-response modeling: A smoking behavior validation study. *Statistics in Medicine*, **32**, 4821–4837.

[22] Fox, J.P., Entink, R.K., Avetisyan, M. (2014). Compensatory and non-compensatory multidimensional randomized item response models. *British Journal of Mathematical and Statistical Psychology*, **67**, 133–152.

[23] Franke, A.G., Bagusat, C., Dietz, P., Hoffmann, I., Simon, P., Ulrich, R., Lieb, K. (2013). Use of illicit and prescription drugs for cognitive or mood enhancement among surgeons. *BMC Medicine*, **11**:102.

[24] 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**, 523–529.

[25] Glynn, A.N. (2013). What can we learn with statistical truth serum? Design and analysis of the list experiment. *Public Opinion Quarterly*, **77**, 159–172.

[26] Groenitz, H.(2014). Improvements and extensions of the item count technique. *Electronic Journal of Statistics*, **8**, 2321–2351.

[27] Gunarathne, A., Kubota, S., Kumarawadu, P., Karunagoda, K., Kon, H. (2016). Is hiding foot and mouth disease sensitive behavior for farmers? A survey study in Sri Lanka. *Asian-Australasian Journal of Animal Sciences*, **29**, 280–287.

[28] Harrison, M., Baker, J., Twinamatsiko, M., Milner-Gulland, E.J. (2015). Profiling unauthorized natural resource users for better targeting of conservation interventions. *Conservation Biology*, **29**, 1636–1646.

[29] Hoffmann, A. and Musch, J. (2016). Assessing the validity of two indirect questioning techniques: A stochastic lie detector versus the crosswise model. *Behavior Research Methods*, **48**, 1032–1046.

[30] Hoffmann, A., Diedenhofen, B., Verschuere, B., Musch, J. (2016). A strong validation of the crosswise model using experimentally-induced cheating behavior *Experimental Psychology*, **62**, 403–414.

[31] Hussain, Z. Shabbir, N., Shabbir J. (2015). An alternative item sum technique for improved estimators of population mean in sensitive surveys. *Hacettepe Journal of Mathematics and Statistics*, first published online, doi: 10.15672/HJMS.20159113160.

[32] Ibrahim, F. (2016). An alternative modified item count technique in sampling survey. *International Journal of Statistics and Applications*, **6**, 177–187.

[33] Imai, K. (2011). Multivariate regression analysis for the item count technique. *Journal of the American Statistical Association*, **106**, 407–416.

[34] Imai, K., Park, B., Greene, K.F. (2015). Using the predicted responses from list experiments as explanatory variables in regression models. *Political Analysis*, **23**, 180–196.

[35] James, R.A., Nepusz, T., Naughton, D.P., Petróczi, A. (2013). A potential inflating effect in estimation models: Cautionary evidence from comparing performance enhancing drug and herbal hormonal supplement use estimates. *Psychology of Sports and Exercise*, **14**, 84–96.

[36] Kazemzadeh, Y., Shokoohi, M., Baneshi, M.R., Haghdoost, A.A. (2016). The frequency of high-risk behaviors among Iranian college students using indirect methods: network scale-up and crosswise model. *International Journal of High Risk Behaviors & Addictions*, first published online, doi: 10.5812/ijhrba.25130.

[37] Khosravi, A., Mousavi, S.A., Chaman, R., Khosravi, F., Amiri, M., Shamsipour, M. (2015). Crosswise model to assess sensitive issues: A study on prevalence of drug abuse among university students of Iran. *International Journal of High Risk Behaviors & Addictions*, first published online, doi: 10.5812/ijhrba.24388v2

[38] 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**, 219–235.

[39] Liu, Y. and Tian, G.L. (2013). Multi-category parallel models in the design of surveys with sensitive questions. *Statistics and Its Interface*, **6**, 137–142

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

[41] Moshagen, M., Hilbig, B.E., Erdfelder, E., Moritz, A. (2014). An experimental validation method for questioning techniques that assess sensitive issues. *Experimental Psychology*, **61**, 48–54.

[42] Moshagen, M., Musch, J., Ostapczuk, M., Zhao, Z. (2010). Reducing socially desirable responses in epidemiologic surveys: An extension of the randomized-response technique. *Epidemiology*, **21**, 379–82.

[43] Moseson, H., Massaquoi, M., Dehlendorf, C., Bawo, L., Dahn, B., Zolia, Y, Vittinghoff, E., Hiatt, R.A., Gerdts, C. (2015). Reducing under-reporting of stigmatized health events using the list experiment: Results from a randomized, population-based study of abortion in Liberia. *International Journal of Epidemiology*, **44**, 1951–1958.

[44] Nakhaee, M.R., Pakravan, F., Nakhaee, N. (2013). Prevalence of use of anabolic steroids by bodybuilders using three methods in a city of Iran. *Addiction & Health*, **5**, 77–82.

[45] Nepusz, T., Petróczi, A., Naughton, D.P., Epton, T., Norman, P. (2014). Estimating the prevalence of socially sensitive behaviors: attributing guilty and innocent noncompliance with the single sample count method. *Psychological Methods,* **19**, 334–355.

[46] Nuno, A., Bunnefeld, N., Naiman, L.C., Milner-Gulland, E.J. (2013). A novel approach to assessing the prevalence and drivers of illegal bushmeat hunting in the Serengeti. *Conservation Biology*, **2**, 1355–1365.

[47] Oliveras, E. and Letamo, G. (2010). Examples of methods to address underreporting of induced abortion: preceding birth technique and randomized response technique. *Methodologies for Estimat-*

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

[48] Perri, P.F. and van der Heijden, P.G.M. (2012). A property of the CHAID partitioning method for dichotomous randomized response data and categorical predictors. *Journal of Classification*, **29**, 76–90.

[49] Perri, P.F., Pelle, E., Stranges, M. (2016). Estimating induced abortion and foreign irregular presence using the randomized response crossed model. *Social Indicators Research*, **129**, 601–618.

[50] Petróczi, A., Nepusz, T., Cross, P., Taft, H., Shah, S., Deshmukh, N., Schaffer, J., Shane, M., Adesanwo, C., Barker, J., Naughton, D.P. (2011). New non-randomised model to assess the prevalence of discriminating behaviour: A pilot study on mephedrone. *Substance Abuse Treatment, Prevention, and Policy*, **6**:20.

[51] Randrianantoandro, T.N., Kono, H. Kubota, S. (2015). Knowledge and behavior in an animal disease outbreak: Evidence from the item count technique in a case of African swine fever in Madagascar. *Preventive Veterinary Medicine*, **118**, 483–487.

[52] Rueda, M., Cobo, B., Arcos, A. (2016). An improved class of estimators in RR surveys. *Mathematical Methods in the Applied Sciences*, first published online, doi:10.1002/mma.4256.

[53] Safiri, S. (2016). Knowledge, attitude, self-efficacy and estimation of frequency of condom use among Iranian students based on a crosswise model: more explanation is needed for the crosswise model. *International Journal of Adolescent Medicine and Health*, first published online, doi:10.1515/ijamh-2016-0110.

[54] Särndal, C.E., Swensson, B, Wretman. J. (1992). *Model Assisted Survey Sampling*. Springer-Verlag, New York.

[55] 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.

[56] Starosta, A.J. and Earleywine, M. (2014). Assessing base rates of sexual behavior using the unmatched count technique. *Health Psychology and Behavioral Medicine*, **2**, 198–210.

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

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

[59] Sukhatme, P.V., Sukhatme, B.V., Sukhatme, S., Asok, C. (1984) *Sampling Theory of Survey With Applications.* Iowa State University Press, Ames.

[60] Tian, G.L. and Tang, M.L. (2014). *Incomplete Categorical Data Design: Non-randomized Response Techniques for Sensitive Questions in Surveys.* Chapman & Hall/CRC, Boca Raton, FL.

[61] Trappmann, M., Krumpal, I., Kirchner, A., Jann, B. (2014). Item sum: A new technique for asking quantitative sensitive questions. *Journal of Survey Statistics and Methodology*, **2**, 58–77.

[62] Tsuchiya, T. (2005). Domain estimators for ithe item count technique. *Survey Methodology*, **31**, 41–51.

[63] Tu, S.H. and Hsieh, S.H. (2017). Estimates of lifetime extradyadic sex Using a hybrid of randomized response technique and crosswise design. *Archives of Sexual Behavior*, **46**, 373–384.

[64] Ulrich, R., Schörter, H., Striegel, H., Simon, P. (2012). Asking sensitive questions: a statistical power analysis of randomized response models. *Psychological Methods*, **17**, 623–641.

[65] Vakilian, K., Mousavi, S.A., Keramat, A., Chaman, R. (2016). Knowledge, attitude, self-efficacy and estimation of frequency of condom use among Iranian students based on a crosswise model. *International Journal of Adolescent Medicine and Health*, first published online, doi: 10.1515/ijamh-2016-0010.

[66] Wu, Q. and Tang, M.L. (2016). Non-randomized response model for sensitive survey with noncompliance. *Statistical Methods in Medical Research*, **25**, 2827–2839.

Appendix A6

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

Perri, Pier Francesco; Cobo, Beatriz, Rueda, María del Mar. (2017)
A mixed-mode sensitive research on cannabis use and sexual addiction: improving self-reporting by means of indirect questioning techniques.
Quality & Quantity
DOI: 10.1007/s11135-017-0537-0

STATISTICS & PROBABILITY			
JCR Year	Impact factor	Rank	Quartile
2016	1.094	51/124	Q2

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.

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 nonsampling 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) 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 nonsensitive 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 QD 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 (page 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 (page 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 Section 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 (Section 3.1) and the IST (Section 3.2) used in the study. Section 4 is devoted exclusively to the description of our research. In particular, Section 4.1 outlines the main features and the field work conducted in the survey, while Section 4.2 comments the results obtained for the sensitive characteristics investigated. In Section 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).

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

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.

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 use 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 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. De 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 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 $\widehat{\bar{Y}}_{\text{HT}}$ denote the well-known Horvitz-Thompson estimator (hereafter HT-estimator; Horvitz and Thompson, 1952) of $\bar{Y}$

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

The estimator is unbiased and has variance

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

which can be unbiasedly estimated by

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

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, $\mathbb{E}_{\text{R}}(r_i) = y_i$, $\forall i \in s$ (see Chaudhuri and Christofides, 2013). Then, in this RR framework, the HT-estimator for the $\bar{Y}$ takes the form

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

with variance

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

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

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

where $\widehat{v}_{\text{R}}(r_i)$ 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 $\widehat{\widehat{Y}}_{\text{RRT}}$ 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

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

which is estimated by

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

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

We note that computing $\widehat{\mathbb{V}}(\widehat{\widehat{Y}}_{\text{HT}})$ and $\widehat{\mathbb{V}}(\widehat{\widehat{Y}}_{\text{RRT}})$ 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 Cristofides (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 $\mathcal{Z} = \mathcal{Y} + \mathcal{T}$ 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 $\hat{\bar{Z}}_{\text{HT}}$ and $\hat{\bar{T}}_{\text{HT}}$ 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

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

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

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

and unbiasedly estimated by

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

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 field work 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 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-sample or the LL-sample. The IST described in Section 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.

[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.

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

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.

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 ISTs 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 $\widehat{\overline{Y}}_{\mathrm{HT}}$, $\widehat{\overline{Y}}_{\mathrm{RRT}}$ and $\widehat{\overline{Y}}_{\mathrm{IST}}$ 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 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 Section 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 $\widehat{\widehat{Y}}_{\text{IST}}$ 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

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

with $n_1^{\text{opt}} + n_2^{\text{opt}} = n$ while $S_.$ denotes the population standard error of the variables $\mathcal{Z}$ and $\mathcal{T}$ which is

[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.

Table 3: Mean estimates and accuracy measures under the three data-collecting 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	492 (100%)					613 (100%)					1293 (100%)				
Q1		3.11	0.60	[1.93;4.30]	2.37		13.07**	3.38	[6.43;19.70]	13.27		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	211 (42.89%)					252 (41.11%)					597 (46.17%)				
Q1		6.35	1.43	[3.54;9.15]	5.61		21.14*	7.20	[7.03;35.25]	28.22		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	281 (57.11%)					361 (53.83%)					689 (53.58%)				
Q1		0.25	0.12	[0.01;0.49]	0.48		7.91**	3.06	[1.90;13.91]	12.01		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 unknown and has to be estimated, for instance, on the basis of a training sample or a pilot survey.

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 Section 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 $N = 50000$ 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 $z_i = y_i + t_i$, $i = 1, \dots, N$;
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 $\widehat{\bar{Y}}_{\text{IST}}$ under optimal and arbitrary allocations, that is, $\widehat{\mathbb{V}}(\hat{\bar{Y}}^{\text{opt}}_{\text{IST}})$ and $\widehat{\mathbb{V}}(\hat{\bar{Y}}^{\alpha}_{\text{IST}})$;
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

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

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

The outcomes of the simulation study are graphically summarized in Figure 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

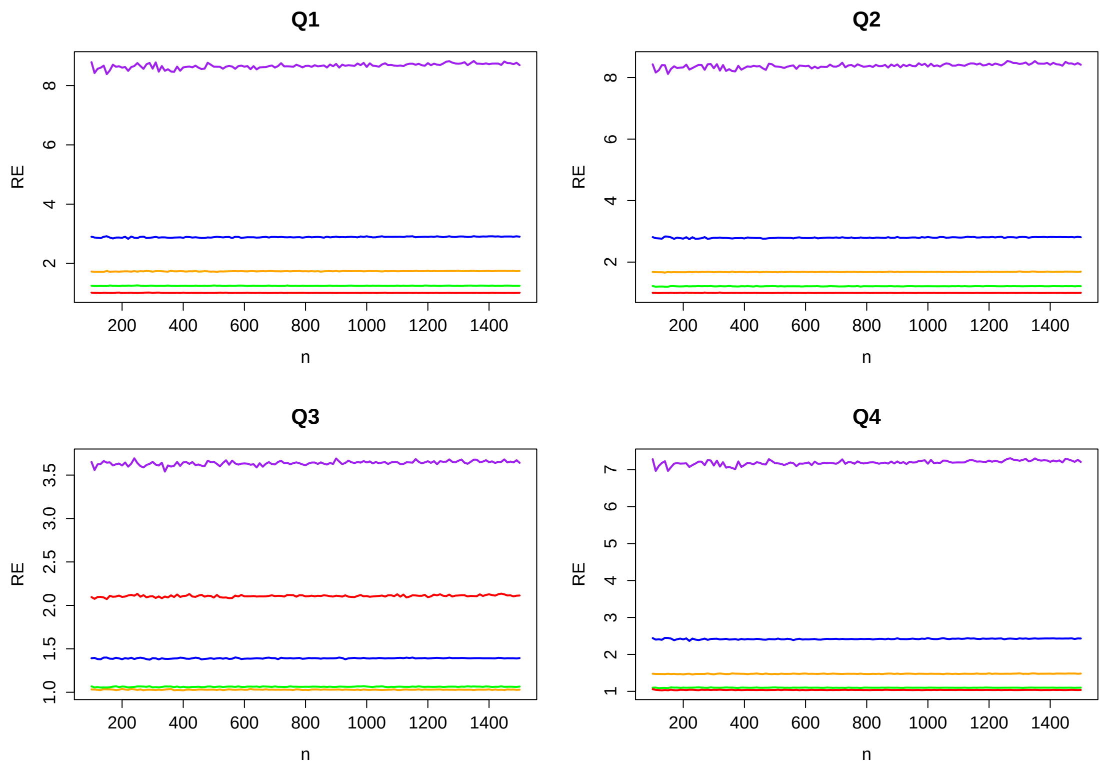


Figure 1: Relative efficiency of optimal allocation w.r.t. arbitrary allocation for different values of α under simulated populations for questions Q1-Q4; $\alpha = 0.1, 0.3, 0.5, 0.7, 0.9$ (violet, blue, orange, green, red).

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 Cristofides, 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.

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

References

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

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

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

[4] Chaudhuri, A., Christofides, T. (2013). *Indirect Questioning in Sample Surveys*. Springer, Berlin.

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

[6] Cochran, W. (1977). *Sampling Techniques*. Wiley & Sons, New York.

[7] 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, 543-64.

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

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

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

[11] 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, 523-529.

[12] Goodstadt, M.S., Gruson, V. (1975). The randomized response technique: A test on drug use. *Journal of the American Statistical Association*, 70, 814-818.

[13] Groves, R.M., Fowler, F.J., Couper, M.O., Lepkowski, J.M., Singer, E., Tourangeau, R. (2004). *Survey Methodology*. John Wiley & Sons, Inc., Hoboken, New Jersey.

[14] Harrison, L., Hughes, A. (1997). *The Validity of Self-Reported Drug Use: Improving the Accuracy of Survey Estimates*. NIDA Research Monograph, 167, 1-16.

[15] Horvitz, D.G., Thompson, D.J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47, 663-685.

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

[17] Kerkvliet, J. (1994). Estimating a logit model with randomized data: The case of cocaine use. *Australian Journal of Statistics*, 36, 9-20.

[18] 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, 219-235.

[19] LaBrie, J.W., Earleywine, M., (2000). Sexual risk behaviors and alcohol: Higher base rates revealed using the unmatched-count technique *The Journal of Sex Research*, 37, 321-326.

[20] Lamarine, R.J. (2012). Marijuana: modern medical chimera. *Journal of Drug Education*, 42, 1-11.

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

[22] Levine, S.B. (2010). What is sexual addiction? *Journal of Sex & Marital Therapy*, 36, 261-275.

[23] Liamputtong, P. (2007). *Researching the Vulnerable. A Guide to Sensitive Research Methods*. Sage Publications, London.

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

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

[26] Miner, M.H. (2008). Improving the measurement of criminal sexual behavior. The application of randomized response technique. *Sexual Abuse*, 20, 88-101.

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

[28] Pitsch, W., Emrich, E., Klein, M. (2007). Doping in elite sports in Germany: Results of a www survey. *European Journal for Sport and Society*, 4, 89-102.

[29] 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.

[30] 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.

[31] Srivastava, R., Nigam, A.K., Singh, N. (2015). Application of randomized response techniques in estimation of prevalence of child sexual abuse. *Statistics and Applications*, 13, 37-45.

[32] Striegel, H., Ulrich, R., Simon, P. (2010). Randomized response estimates for doping and illicit drug use in elite athletes. *Drug and Alcohol Dependence*, 106, 230-232.

[33] Stubbe, J.H., Chorus, A.M.J., Frank, L.E., de Hon O., van der Heijden, P.G.M. (2013). Prevalence of use of performance enhancing drugs by fitness center members. *Drug Testing and Analysis*, 6, 434-438.

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

[35] Tourangeau, R., Smith, T. (1996). Asking sensitive questions: The impact of data collection, question format, and question context. *Public Opinion Quarterly*, 60, 275-304.

[36] Tourangeau, R., Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin*, 133, 859-883.

[37] Trappmann, M., Krumpal, I., Kirchner, A., Jann, B. (2014). Item sum: A new technique for asking quantitative sensitive questions. *Journal of Survey Statistics and Methodology*, 2, 58-77.

[38] 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, 49-72.

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

[40] Weissman, A.N., Steer, R.A., Lipton, D.S. (1986). Estimating illicit drug use through telephone interviews and the randomized response technique. *Drug and Alcohol Dependence*, 18, 225-233.

[41] Wolter, K. (2007). *Introduction to Variance Estimation*. Springer-Verlag, New York.

Appendix A7

Indirect questioning methods reveal hidden support for female genital cutting in South Central Ethiopia

Gibson, Mhairi A.; Gurmu, Eshetu; Cobo, Beatriz; Rueda, María del Mar; Scott, Isabel M. (2018)
Indirect questioning methods reveal hidden support for female genital cutting in South Central Ethiopia.
PLOS ONE

MULTIDISCIPLINARY SCIENCES			
JCR Year	Impact factor	Rank	Quartile
2016	2.806	15/64	Q1

Abstract

Female genital cutting (FGC) has major implications for women's physical, sexual and psychological health, and eliminating the practice is a key target for public health policy-makers. To date one of the main barriers to achieving this has been an inability to infer privately-held views on FGC within communities where it is prevalent. As a sensitive (and often illegal) topic, people are anticipated to hide their true support for the practice when questioned directly.

Here we use an indirect questioning method (unmatched count technique) to identify hidden support for FGC in a rural South Central Ethiopian community where the practice is common, but thought to be in decline. Employing a socio-demographic household survey of 1620 Arsi Oromo adults, which incorporated both direct and indirect direct response (unmatched count) techniques we compare directly-stated versus privately-held views in support of FGC, and individual variation in responses by age, gender and education and target female (daughters versus daughters-in-law).

Both genders express low support for FGC when questioned directly, while indirect methods reveal substantially higher acceptance (of cutting both daughters and daughters-in-law). Educated adults (those who have attended school) are privately more supportive of the practice than they are prepared to admit openly to an interviewer, indicating that education may heighten secrecy rather than decrease support for FGC. Older individuals hold the strongest views in favour of FGC (particularly educated older males), but they are also more inclined to conceal their support for FGC when questioned directly. As these elders represent the most influential members of society, their hidden support for FGC may constitute a pivotal barrier to eliminating the practice in this community.

Our results demonstrate the great potential for indirect questioning methods to advance knowledge and inform policy on culturally-sensitive topics like FGC; providing more reliable data and improving understanding of the "true" drivers of FGC.

1 Introduction

Over 200 million of the world's female population currently live with female genital cutting (FGC), which involves the removal of, or injury to, their genital organs for non-medical reasons [1]. Twenty years since the first joint WHO/UN statement against FGC, elimination of the practice remains a key unmet development goal (SDGs, Target 5). The health implications, particularly from the more severe types of FGC, are well-documented and include infection, obstetric, psychological and sexual problems [2]. However, high quality data on FGC behaviour and norms, which are essential to the design of effective intervention programmes, remain elusive. The practice occurs in private and the effects are not externally visible. Furthermore it is a sensitive topic, making people inclined to hide their "true" views on the topic. For example, people may feel pressure to understate the prevalence of the practice or their preference for it (due to its illegality) [3], or to overstate it (due to social pressures, e.g. as a requirement of marriage). The possibility of both understatement and overstatement makes it especially difficult to assess to the prevalence and predictors of support for FGC at present.

To date most studies exploring FGC behaviour have relied on self-report data derived from direct questioning methods, with many indicating that rates of (and interest in) FGC are broadly in decline [4]. For example, the direct questions used in the Ethiopian Demographic and Health Surveys have revealed that FGC prevalence in Ethiopia declined from 80% in 2000 to 74% in 2005, and directly-stated support for the practice almost halved (60% to 31%) over the same 5 year period [5]. FGC status obtained through physical examination (the gold standard for FGC studies) rarely exists to substantiate these claims, and where it does, has revealed discordance between the two measures (indicating both over-reporting and under-reporting) [6, 7]. This disparity, between clinical and self-report data, confirms that people may be inclined to conceal FGC behaviour (and their support for it) in surveys. Yet, physical examination is intrusive and expensive (requiring a health professional), and thus is infeasible as a tool to guide research and policy.

To address these issues we employ an "indirect" questioning method (unmatched count technique or UCT, see Methods) to explore variation in support for FGC within one rural Ethiopian community where FGC is illegal. These kinds of indirect questioning methods can anonymously obtain responses to sensitive questions [8-10], and can be used to gauge the extent and direction of 'true' responses, as well as individual variation in hidden views or behaviours. These methods permit the estimation of population-level support for cutting without revealing the individual preferences to the interviewer. Further, by comparing indirect question responses with those from traditional direct questioning, it possible to identify the extent to which behaviours (or views in support of the practice) are concealed (over or under-reported) using direct questioning. These techniques have recently led to an improved understanding of civic issues such as racial prejudice and poaching [11, 12], but have been relatively under-applied to substantive health issues in low income settings. There is, however, growing interest in the potential of UCT to accurately record sensitive reproductive health-related behaviours, for example, improving abortion statistics [13]. One recent study has used them to explore women's views on female genital cutting among the Afar (a pastoralist community located in the North of Ethiopia: [14]). While ours not the first UCT study to uncover hidden support for FGC, we suggest improvements to study design (e.g. we consider the views of men, as well as women), and develop the UCT data collection methods to make them applicable in low income settings, something missing from the previous study. These are discussed in further details in the Methods section below.

How and why FGC is maintained in some populations despite the health consequences for women and efforts to eliminate the practice has been of long-standing interest for social and medical anthropologists [15]. Recently evolutionary anthropologists have also addressed the question and are providing important and novel insights which help to explain variation in FGC behaviour (and acceptance of it) [16, 17]. For some, the evolutionary origin and persistence of the practice has been linked to controlling women's sexual desires and behaviours before or within marriage, which increases male paternity certainty [18]. In other words, it lessens the chance of pre-marital or extra-marital affairs, and eliminates the risk to men of raising unrelated offspring (rather than their own genetically related progeny). To date, however, these ideas remain largely untested using empirical data. Here, based on a similar evolutionary perspective, and drawing on evolutionary kin selection and sexual conflict theories, we explore the extent to which relatedness is important in explaining individual variation in views in support of the practice. One prediction is that there is likely to be more support to cut daughters-in-law than there is to cut daughters. This is based on the assumption that the adverse health consequences of FGC in closely related kin (e.g. daughters), may be of greater concern than non-biological kin (e.g. daughters-in-law); while paternity certainty (and mate-guarding) may be of greater concern when relating to daughters-in-law than to daughters. An alternative proposal derived from evolutionary theory is that cutting will be endorsed equally for both daughters and daughters-in-law, as parents interests in both are closely tied [19]. Any health risk to either groups of women from the procedure may impact on parents' reproductive fitness (lead to fewer surviving grand-offspring). Further any benefits of cutting may be equivalent too. For example, if cutting signals sexual fidelity, daughters who are cut may have better marriage prospects, and receive greater support from their in-laws.

In this study we combine direct and indirect questioning methods to explore concealed support for FGC according to individual circumstances including gender, age and level of education of the respondent, as well as the characteristics of the target female. Individual variation in level of support for FGC based on gender, age and education are well known in the anthropological literature [20-22]; however the reported effects vary due to contextual differences between populations (as well as differences in methodologies). The extent to which the desirability of FGC varies between categories of female kin (daughters and daughters-in law) has to our knowledge, not previously been tested. Our data are drawn from a rural Ethiopian Arsi Oromo community where household surveys over a five year period have revealed a recent and rapid decline in self-reported FGC prevalence rates (from 90% in 2010 to $<$20% in 2015; Gibson, personal communication). This sudden drop in reporting rates is an indication that women in this community have become more inclined to conceal their FGC status.

2 Methods

In 2016 a socio-demographic household survey was undertaken with 1620 adults living in rural sub-districts of Arsi and East Shewa zones, Southern Oromia. This included an equal and randomly selected sample of adult ($>=$18 years) male and female respondents, married and unmarried respondents (one of each sex, and one of each marital status from alternate households selected from a village plan, and household member lists provided by the district office). The survey was undertaken in the respondent's house (or within their compound) by a trained same-gender interviewer fluent in the local language, Oromiffa. No other adult was present. Prior to the main survey, focus group discussions were undertaken to develop the questionnaire (e.g. choosing the items included in the unmatched count technique list), the survey was then piloted in a neighbouring village, and all interviewers received training in the survey protocols [see Supporting Information S1 Fig and S2 Fig]. Research and Ethical approval to undertake this study was granted by the Ethics Committees at the University of Addis Ababa and the University of Bristol. Informed written consent (or fingerprint consent) was obtained from each participant in the study.

To compare openly-declared and privately-held support for FGC, the survey employed direct questions on the desirability of FGC, as well as the unmatched count technique (UCT), an indirect questioning method designed to mitigate the problems associated with sensitive survey topics. UCT is sometimes referred to at the List Experiment or the Item Count Technique [9, 11]. All respondents were asked about the desirability of FGC for both a hypothetical daughter, and a prospective daughter-in-law. Details of the questions posed in each survey version can be found in S1 Fig, and details of the sampling methods are outlined in S2 Fig.

There were four different versions of the survey which were randomly assigned to respondents, these included direct and indirect questions (Version 1 and 2), a control and treatment condition (lists with and without the sensitive item, FGC; Version A and B). Seventy percent of the sample undertook a survey with the indirect UCT question (n=1112), with participants equally and randomly assigned to either a control or a treatment condition (see Fig 1). Individuals in the control condition (Version 2A) were shown cards with four (non-sensitive) items and asked how many (but not which particular) items are desirable for their daughter or daughters-in-law, while the treatment group (Version 2B) was shown the same set of cards but with the sensitive item added (the five item treatment) [S2 Fig on how UCT item lists were generated, and tested]. The difference between the mean number of items reported in the two conditions provided an estimate of those in favour of cutting (the sensitive item) for the entire population.

The remaining 30% of the sample answered direct questions (n=508), and were also randomly assigned to either a four item control group (Version 1A) or five item treatment group which included the additional FGC card (Version 1B). In this case respondents were asked to identify whether each of the items on the cards were desirable for a hypothetical daughter or daughter-in-law. The proportion of individuals saying "yes" to the FGC card in the five card treatment group (n=331) provided an estimate of directly expressed "popularity" of FGC. The proportion of "yeses" for the four non-FGC cards was compared across control (1A) and treatment (1B) groups to check that adding an FGC item did not influence how people responded to the other four cards (a design effect). This "additional item" test was satisfied [see S2 Fig for more detail on our sampling strategy and data quality tests].

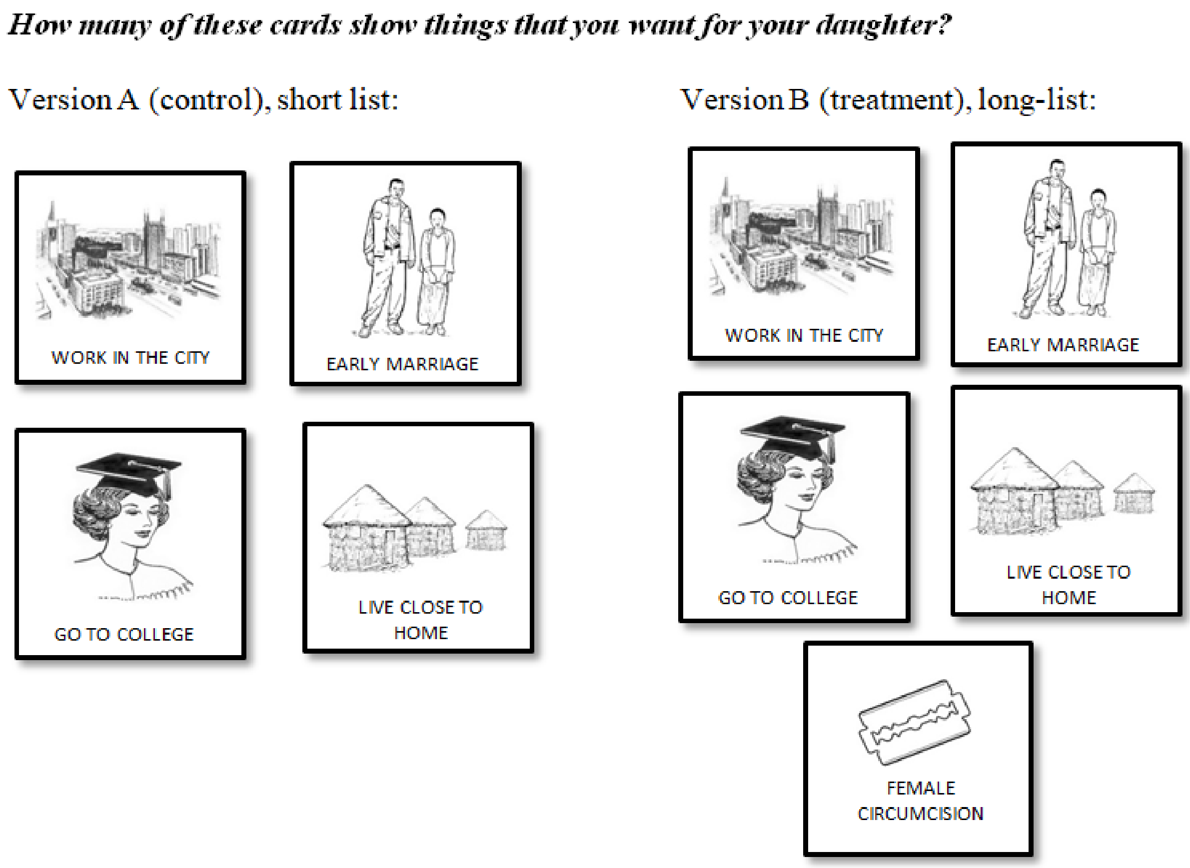


Figure 1: Unmatched count techniques (UCT) question materials. UCT items were presented on illustrated cards, to facilitate comprehension and randomize item presentation order. (Supporting Information S1 Fig includes full details of questions)

A comparison of the estimates obtained from direct and UCT questioning methods provided a measure of the extent to which privately-held views differed from those openly-stated (i.e. FGC support that was over or under-reported when asked directly). To ensure an accurate comparison could be made between direct and UCT estimates we used a single question/measure of FGC support in both direct and UCT surveys "Would you want [item named] for your daughter (or daughter in law)?" This is an improvement on previous studies which have relied on estimates obtained from non-identical measures of the sensitive item (e.g. different questions/scales used to define support for the sensitive behaviour in DT and UCT surveys; [14]), which has the potential to introduce inaccuracies in comparative analyses.

Another novel aspect of this UCT study design was the use of cards with pictures for each item included in the list, allowing randomized presentation of the list items and improved respondent comprehension. The participants were able to handle the cards as each item was read out by the interviewer, who could shuffle the cards between interviews (see Fig 1). Previous UCT studies have required participants to read the list of items, which are less well suited for use in populations with literacy issues, and present challenges for item randomization. All items included in the list were generated from focus group discussions and picture cards quality tested during piloted phases of the study [See S2 Fig].

One challenge for all indirect questioning methods is in minimizing biases in responses. Biased responses to the sensitive time may occur where informants become cognisant of the nature of the survey and/or feel the anonymity of responses may be compromised. For example, in the pioneering study by De Cao and Lutz [14] the sensitive question was embedded in a survey on a related topic (women's reproductive health), and the survey was administered by people known to hold a particular viewpoint (sexual health charity workers trying to eliminate the practice). To minimize potential biases in responses: participants in our study answered either direct questions or indirect questions (not both); the survey included no additional questions related to the sensitive topic; and was administered by professional enumerators of the same gender not known to the informants. Further, the items included in the UCT list were carefully chosen to so as to minimize the chance of floor and ceiling effects that is, of participants preferring either all or none of the items. Such effects can be problematic because they effectively reveal the participant's attitude to the sensitive item [9, 11]. Following practices advocated in prior research to mitigate floor/ceiling effects, [11] four items were selected such that: a) one item was expected to be unpopular (early marriage) b) one item was expected to be popular (go to college) and c) two items were expected to be seen as incompatible (work in the city, and live close to home).

3 Statistical analyses

Analyses were performed using freely available R software, version 3.3.2 (2016-10-31) [23]. To contrast the proportions between the direct question (DQ) method and unmatched count techniques (UCT), and for subgroups (in both DQ and UCT methods) we used a contrast of equal proportions (calculating the value of the statistic and its associated p-value). We also performed multivariate analyses using generalized regression models (with and without iterations of the covariates) developed by [24, 25], and [9] and applied in other UCT analyses including [26, 27] and [14]. These multivariate analyses have not been included in this paper, as none of the tested models fitted well (possibly because some of the sub-groups have a very small sample size (see S1 Table)). This is one important challenge for this methodology: the UCT provides respondents with privacy at the expense of statistical efficiency [25], and large sample sizes are required.

4 Study population

The Arsi Oromo are agropastoralists who combine cattle rearing with maize, wheat and sorghum cultivation in the rural low-lying areas of Arsi and East Shewa administrative zone, in Oromia, South Central Ethiopia. Family sizes are large, but agricultural land is limited [28]; and off-farm employment opportunities are rare [29]. Schooling is limited: a third of adults in our sample had never attended school, the rest attending for on average less than two years. The community has limited access to media and urban exposure: of our sample, over a third (36%) had never listened to radio, 53% had never watched TV, <20% reported never visiting a big town or city. Inheritance patterns are patrilineal, and wealth inequality is relatively low due to a programme of government land redistribution in the late twentieth century [30]. Arranged marriages are central to alliance formation between un-related families, often involve large cash bridewealth payments which are transferred from groom to the bride's family, and post-marital residence is predominantly patrilocal (a daughter moves to join their husband's village and lineage at marriage) [31].

FGC among the Arsi Oromo involves a nick or cut to the clitoris, and is linked closely to marriage. Cutting occurs in adulthood in the months leading up to marriage (typically women marry in their late teens) and is performed in private, by traditional female practitioners. Since 2004 FGC has been illegal in Ethiopia, and this is widely known within the community (98% of our sample were aware that there is a law preventing FGC), although there have been no local incidences of anyone being brought to trial. To date, openly declared rates of FGC from this community suggest that the practice is either in decline, or is being increasingly under-reported. Over a five year period, the number of women directly reporting that they had been cut (in household census surveys) dropped from 90% of women in 2010 to <20% in 2015 [Gibson, personal communication].

5 Results

A total of 1620 adults were included in the survey and analyses, this included equal numbers of males and females (811 males, and 809 females). The non-response rate was zero. Over one third of the sample (n=581) had never attended school. For those who had received formal education (n =1039) the majority had completed less than three years in school. For the purposes of analyses, the sample was divided into two age groups: 18-25 years, and 26+ years. The results are summarized in Table 1 below.

Table 1: Direct question (DQ) and unmatched count technique (UCT) estimates indicating support for FGC in daughters and daughters-in-law by gender, age and education level of the respondent.

Respondents	Relative	DQ estimate (SE)[a]	UCT estimate (SE)[b]	P-values[c]
All	**Daughters**	0.073 (0.014)	0.197 (0.040)	0.003
	In-laws	0.082 (0.014)	0.250 (0.043)	<0.001
	Both[d]	0.077 (0.010)	0.224 (0.030)	<0.001
Male	**Daughters**	0.074 (0.020)	0.142 (0.071)	0.356
	In-laws	0.092 (0.022)	0.228 (0.070)	0.064
	Both	0.083 (0.015)	0.185 (0.050)	0.051
Female	**Daughters**	0.071 (0.019)	0.256 (0.044)	<0.001
	In-laws	0.071 (0.019)	0.278 (0.053)	<0.001
	Both	0.071 (0.013)	0.267 (0.034)	<0.001
18-25 years	**Daughters**	0.072 (0.023)	0.094 (0.074)	0.781
	In-laws	0.072 (0.023)	0.061 (0.079)	0.896
	Both	0.072 (0.017)	0.078 (0.054)	0.923
26+ years	**Daughters**	0.073 (0.017)	0.249 (0.047)	<0.001
	In-laws	0.086 (0.018)	0.344 (0.052)	<0.001
	Both	0.080 (0.012)	0.296 (0.035)	<0.001
No education	**Daughters**	0.116 (0.027)	0.155 (0.068)	0.595
	In-laws	0.124 (0.028)	0.233 (0.073)	0.159
	Both	0.120 (0.019)	0.194 (0.050)	0.165
Some education	**Daughters**	0.045 (0.014)	0.219 (0.049)	<0.001
	In-laws	0.054 (0.015)	0.257 (0.053)	<0.001
	Both	0.050 (0.010)	0.238 (0.036)	<0.001

[a] Derived from direct questions (DQ)
[b] Derived from indirect questions (the unmatched count technique, UCT)
[c] P-value refers to significance of difference between DQ and UCT estimates.
[d] Mean estimates derived from responses to two questions regarding daughters and daughters in law.

5.1 Direct versus indirect response methods

The indirect response method, unmatched count technique (UCT), revealed that people privately had higher levels of acceptance of FGC behaviour than was admitted openly through direct questioning (DQ). Overall, a low proportion (7.7% (95% CI [5.8-9.6]) of directly posed questions about FGC were met with a positive response (in favour of FGC), whereas UCT indicated that "true" support was three times higher, at approximately 22.4% (95% CI [16.6-28.2]) of responses (difference: $p<0.001$, see Table 1 and Fig 2).

5.2 Kinship relationship to women (daughter versus daughter in law)

Respondents reported no difference in level of support for FGC for daughters than daughters-in-law both when asked directly, 7.3% (95% CI [4.6-9.9]) and 8.2% (95% CI [5.4-11.0] respectively, p=0.645), or indirectly using UCT 19.7% (95% CI [11.9-27.6]) and 25% (95% CI [16.5-33.5], respectively, p=0.371); see Table 1 and Fig 2. There is, however, evidence of concealment of FGC support (i.e. greater difference between direct and indirect UCT estimates) for both categories of female relatives. When considering hypothetical daughters, 7.3% (95% CI [4.6-9.9]) of respondents were supportive when asked directly, rising to 19.7% (95% CI [11.9-27.6]) when questioned indirectly (p=0.003). When considering hypothetical daughters-in-law, direct and UCT responses were 8.2% (95% CI [5.4-11.0]) and 25% (95% CI [16.5-33.5]) respectively ($p<0.001$).

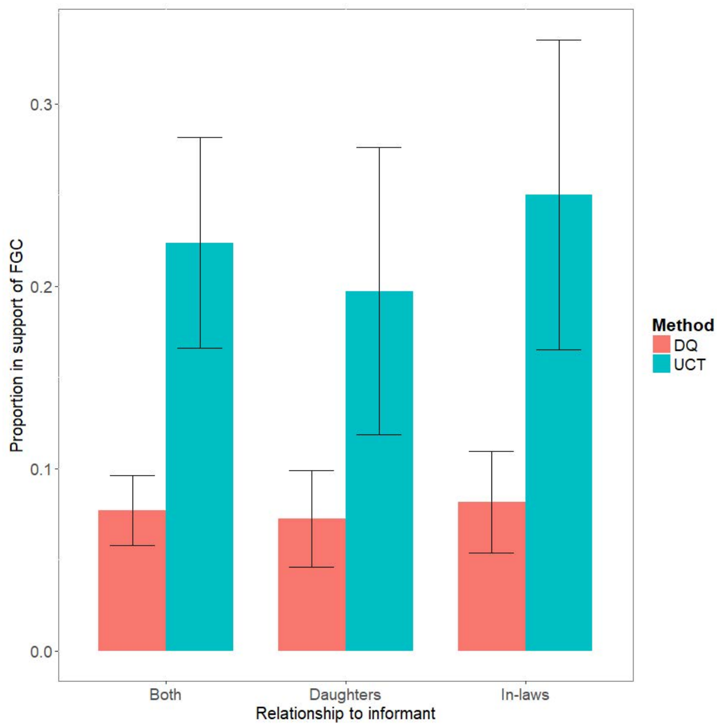


Figure 2: Bar chart comparison of the proportion of people in favour of FGC for a hypothetical daughter or daughter-in-law or both combined, using DQ and UCT responses (estimated proportions) (n=1620). The error bars represent confidence intervals at 95%.

5.3 Individual characteristics of respondents

In addition within the community certain kinds of individual were more likely to hide their views in favour of FGC when questioned directly. We tested whether education, gender or age influenced reported acceptance of FGC, and these results are reported below. The dependent variable in the following paragraphs is the overall proportion of supportive responses regarding FGC, derived from the responses to questions regarding daughters and in-laws combined (unless otherwise stated). Table 1 provides a breakdown of the estimates according to question methodology (direct vs. UCT), individual traits (e.g. gender) and target female (daughter vs daughter-in-law). S1 Table provides more detailed breakdown of the sub group analyses, to explore the relationship between variables, between methodology, and each of the three individual traits (gender, age group and educational level).

5.3.1 Gender

Men and women reported similar and low levels of support for FGC when asked directly (8.3% and 7.1%, 95% CIs [5.4-11.1] and [4.5-9.8], p=0.563; see Fig 3a). Using UCT, women appeared privately more supportive of the practice than men (men: 18.5%, women: 26.7%, 95% CIs [8.7-28.3] vs [20.0-33.5]), but this difference was not significant, (p=0.178). A comparison of direct and UCT estimates indicates that both men and women concealed their true support for FGC to some degree when questioned directly. Men reported low levels of support for FGC in response to the direct question, 8.3% (95% CI [5.4-11.1]) rising to 18.5% (95% CI [8.7-28.3]) in response to the UCT, (at borderline significance, p=0.051). A breakdown of these estimates according to target women (presented in Table 1), however suggests that males were less likely to conceal support for FGC in daughters (p=0.356) than daughters-in-law (although not quite at 5% significance level, p=0.064). For women DQ reveals that FGC support was 7.1% (95% CI [4.5-9.8]), rising to 26.7% (95% CI [20.0-33.5]) using UCT, (p<0.001), with concealment evident for both daughters and daughters-in-law (see Table 1).

5.3.2 Age

When asked directly, individuals in the two age-groups (<26, >=26) reported similar, low support for FGC (7.2%, 8% respectively; 95% CIs [4.0-10.4], [5.6-10.4], p=0.716). Indirect estimates, however, indicate that private support was higher among those aged over 26 (29.6%, 95% CI [22.8-36.5]) than those aged <26 (7.8%, 95% CI [0-18.4]; with p<0.001). There also is a significant discrepancy between DQ and UCT estimates of support among the oldest subgroup (>=26 years), 8% (95% CI [5.6-10.4]) and 29.6% (95% CI [22.8-36.5]) respectively, (p<0.001), but not in the youngest subgroup (<26 years), 7.2% (95% CI [4-10.4]) and 7.8% (95% CI [0-18.4]) respectively, (p=0.923). (See Fig 3b). These results indicate that older individuals were privately more supportive of FGC than younger individuals, but were also more likely to conceal this support when asked directly.

5.3.3 Educational level

When asked directly, uneducated respondents were more likely to admit support for FGC than those who had received formal education (ever attended school) (12.0% compared to 5.0%, 95% CIs [8.2-15.8] vs [2.9-7.0], p=0.001). UCT, however, suggested a reversal of this with uneducated respondents privately being less supportive of FGC than educated individuals (19.4% vs 23.8%, 95% CIs [9.7-29.2] vs [16.7-30.9]); however these differences were not statistically significant (p=0.476). Educated respondents were more likely to admit their support for FGC when questioned indirectly using UCT than using a direct method; 5.0% expressed direct support for FGC, rising to 23.8% using UCT (95% CIs [2.9-7.0]) vs [16.7-30.9]; $p<0.001$) (see Fig 3c). For respondents with no schooling, similar biases were not evident; percentage support was 12% using direct questions (95% CI [8.2-15.8]) and 19.4% for UCT (95% CI [9.7-29.2]); and the difference between direct and UCT responses was not significant (p=0.165). These results indicate that educated and non-educated people did not differ in their privately-held views on the desirability of FGC; however, educated respondents were more likely conceal this support when questioned directly by an interviewer.

5.3.4 Further sub-group analysis

Prior research suggests that FGC practices may persist in certain subsections of society in which they are normative [16]. We therefore tested whether private acceptance of FGC existed at levels around 50% in any sub-group of the Arsi Oromo population. However, it is worth noting that for this population there are small sample sizes in certain sub-groups, (e.g. there are very few older, educated Arsi Oromo females, or uneducated young adults) meaning that many differences are not statistically significant. Our analyses reveal that high levels of private support for FGC are found among older, educated males (+26 years), where estimated acceptance levels reached 45% (39.7% for daughters, and 50.4% for daughters-in-law; 95% CIs [24-55.4] and [34.2-66.5] respectively. This category of individuals was also the least likely to openly admit a preference for FGC (1.4% and 4.1%; 95% CIs [0-3.9] and [0-8.4]); which is reflected in a large discrepancy between directly-expressed and privately-held views, the largest of any subgroup of individuals within this population (both $p<0.001$). (See Fig 3d). Supporting information (S1 Table) includes a breakdown of sub-group analyses, contrasts between question methodology (DQ vs UCT) and each of the respondent's individual traits (gender, age and education level).

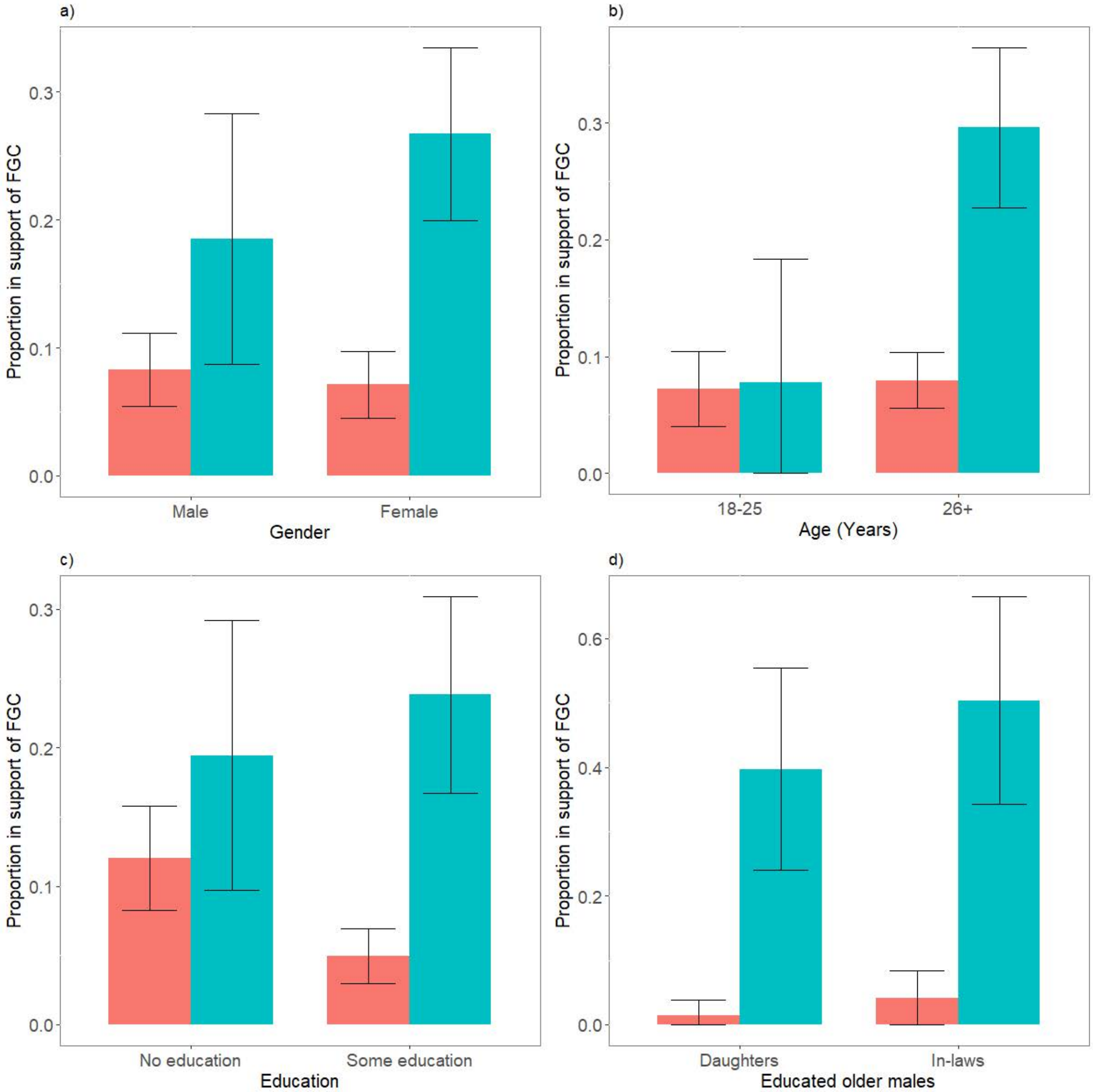


Figure 3: Bar chart comparisons of the proportion of people supporting FGC, using DQ and UCT responses by a) gender, b) age group and c) education level [includes mean estimates of daughter and in-laws combined, n=1620]. Graph d) includes only the subgroup educated, older males [n=408], with separate estimates for daughters and daughters-in-law. DQ=orange bars, UCT=green bars. The error bars represent confidence intervals at 95%.

6 Discussion

Here we demonstrate that traditional direct methods, which rely on direct, face-to-face questioning to determine levels of support for FGC are highly unreliable. Comparing direct and indirect response methods in rural Oromia, South Central Ethiopia, we identify substantial underreporting of support for FGC using direct questioning methods. Across the community, privately-held views in favour of FGC are approximately three times higher than those admitted when asked directly by an interviewer. We identify that older individuals hold the strongest views in favour of FGC, but are also the most likely to hide their 'true' support for the practice when questioned directly. The lowest concealed support for FGC is among the youngest cohort (<26 year olds) which could suggest that social norms favouring FGC are shifting across the whole community overtime or alternatively that individuals become more accepting of the practice with age. Repeated surveys in this community may help to identify the extent to which either or both of these scenarios are true.

Our results also indicate that educated Arsi Oromo give more socially desirable answers than those individuals without schooling, hiding their 'true' FGC intentions when questioned directly. Similar associations linking education and under-reporting of sensitive attitudes (e.g. racist beliefs) have been documented in high-income populations [32], but this link has been harder to establish in low-income settings [14]. The knowledge that the most educated people in the community are inclined to conceal their private views in favour of FGC is highly relevant for public health policy seeking to eradicate the practice. Improved community education on the health risks has long been a major focus for policy. However, to date broad public health information campaigns have had mixed results, and for the most part have not motivated mass abandonment of cutting (see review by Shell-Duncan [15]).

It is however worth noting, that educational attainment for those in our sample attending school is very low (< 3years completed education), and it may be that for acceptance of FGC to decline, much higher levels of education maybe required (e.g. secondary schooling). Lower acceptance of FGC has been identified among the educated Afar pastoralists in Northern Ethiopia using similar UCT methods [14], which may reflect greater schooling and income-generating opportunities available in this population. Among our sample of Arsi Oromo, women's economic security is very strongly linked to securing a good marriage to a wealthy man, arranged on their behalf by their parents and community elders. In this context, the perceived socio-economic advantages with increased education may actually maintain (or even increase) leverage for educated individuals to "demand" FGC in potential spouses/in laws. Our results suggest that educational expansion (at least in its early stages) may not be enough to change FGC norms of an entire community, rather it may simply heighten secrecy. Community-level intervention schemes seeking to promote the abandonment of FGC, have similarly been linked with under-reporting [14], and change in practices to prevent detection, e.g. cutting at an earlier age [33] in other populations.

Some FGC eradication efforts have sought to target key individuals within communities, many with an emphasis on women, who are considered to be 'at the forefront of the perpetuation of FGC'. In this Arsi Oromo community and many others where FGC is common, women lead the rites surrounding the practice, and carry out the cutting. Men are often considered to be less directly involved in the process and (when asked directly) in favour of the abandonment of the practice [34]. Our results reveal that both men and women are equally supportive of FGC in our sample, and attempt to conceal their support in front of interviewers. This finding casts doubt on the potential efficacy of FGC eradication programmes with an exclusive focus on women (in this and similar communities).

We find no clear evidence for weaker support for FGC for daughters over daughters-in-law, in line with an evolutionary prediction that parents will be more concerned with controlling the sexual behaviour of their daughters-in-law. Rather, our results support the proposal that the fitness costs and benefits of cutting, in terms of health risks and paternity certainty respectively, are equivalent for daughter and daughters-in-law. Put simply, parents don't want their sons raising other men's children, but they also want potential spouses (and future in-laws) to have faith in the fidelity of their daughters. In our community, cutting daughters remains the best way of ensuring good marriages, but also in-law support for daughters when leaving their natal home after marriage [16]. We do, however, find an indication that men may be less inclined to conceal their support for cutting daughters, than for daughters-in-law (although at borderline significance, $p=0.064$). Why this should be is currently unclear, but we speculate that this may reflect particular pressures for fathers to openly signal sexual fidelity and hence marriageability of their daughters to potential spouses and in-laws. The suggestion of variation in desirability of FGC based on sex and relatedness supports the evolutionary proposal that sexual conflict in humans is not constant, but may vary across socio-ecological circumstances [19], as well as highlighting a need for further FGC studies exploring the role of kinship and differential kin support.

Finally, our results suggest that it is elders, particularly educated men who hold some of the strongest views in favour of the practice (>45% privately endorse FGC, but these views are hidden when asked directly). This group represents around 12% of the total population, and hold positions of authority in the community, taking on responsibility for village leadership, defence, and key social rites (e.g. arranging marriages). Concealed support and pressure to continue FGC from this powerful and influential group of elders could explain the stubborn persistence of the practice in this and similar communities. For policy-makers, the identification of such sub-groups is important, because individuals are likely to form alliances on the basis of similarity, and because evidence suggests that conformity to normative cultural practices may explain the popularity both of FGC and many other cultural traits [16, 35]. The existence of "pockets" of high support may therefore explain the persistence of FGC in populations in which it is, overall, a minority practice. It is worth noting, however, that the results presented here (and in similar studies) should be interpreted with caution, as there is a statistical requirement for large sample sizes when conducting unmatched count technique analyses [11]. Replication of our methods in larger samples would allow "high support" subgroups to be confidently identified, and key subgroups targeted for interventions in this and in other populations.

7 Conclusion

Our results demonstrate the inadequacy of traditional, yet widely used, direct questioning methods, and the potential for indirect questioning techniques to improve understanding of culturally-sensitive topics, like FGC. Comparing direct and indirect methods we reveal how some individuals (particularly influential older people) are more inclined to hide their "true" support for female genital cutting in rural Ethiopian community. While there is a requirement for large sample sizes to use these techniques, our findings support a growing view that indirect questioning methods can be usefully applied to improve understanding on a range of sensitive health-related behaviours, as well as to improve reliability in monitoring and evaluation of health intervention initiatives across low-income settings.

Acknowledgements

We acknowledge with grateful thanks the participation of the people of Hitosa and Dodota *weredas* of Arsi zone, and Adama *wereda* of East Shewa zone, Oromia region; the contribution of our skilled and dedicated field team in Iteya, and the guidance provided by the Research and Ethics Committee at Addis Ababa University during data collection. Thanks also due to Anne Leaver for her help creating detailed illustrations. In addition we are very grateful for the insightful comments provided by the editor and three academic reviewers, as well as David Lawson, Alejandra Núñez de la Mora, and Caroline Uggla. All of these contributions have considerably improved the manuscript.

References

[1] UNICEF. Female genital mutilation/cutting: A global concern. 2016.

[2] World Health Organization. WHO guidelines on the management of health complications from female genital mutilation. 2016.

[3] Askew I. Methodological issues in measuring the impact of interventions against female genital cutting. Culture Health & Sexuality. 2005;7(5):463-77. doi: 10.1080/13691050410001701939.

[4] Koski A, Heymann J. Thirty-year trends in the prevalence and severity of female genital mutilation: a comparison of 22 countries. BMJ Global Health. 2017;2(4). doi: 10.1136/bmjgh-2017-000467.

[5] Ethiopian Central Statistics Agency [Ethiopia], ORC Macro. Ethiopia Demographic and Health Survey 2005. Addis Ababa, Ethiopia and Calvertion, Maryland, USA: Central Statistics Agency and ORC Macro, 2006.

[6] Elmusharaf S, Elhadi N, Almroth L. Reliability of self reported form of female genital mutilation and WHO classification: cross sectional study. BMJ. 2006;333(7559):124. doi: 10.1136/bmj.38873.649074.55.

[7] Klouman E, Manongi R, Klepp K-I. Self-reported and observed female genital cutting in rural Tanzania: associated demographic factors, HIV and sexually transmitted infections. Trop Med Int Health. 2005;10(1):105-15. doi: 10.1111/j.1365-3156.2004.01350.x.

[8] Blair G, Imai K, Zhou Y-Y. Design and analysis of the randomized response technique. Journal of American Statistical Association. 2015;110(511):1304-19.

[9] Blair G, Imai K. Statistical analysis of list experiments. Political Analysis. 2012;20(1):47-77.

[10] Warner SL. Randomized response: A survey technique for eliminating evasive answer bias. J Am Stat Assoc. 1965;60(309):63-9.

[11] Glynn AN. What can we learn with statistical truth serum? Design and analysis of the list experiment. Public Opinion Quarterly. 2013;77(S1):159-72. doi: 10.1093/poq/nfs070.

[12] Nuno A, St. John FAV. How to ask sensitive questions in conservation: A review of specialized questioning techniques. Biological Conservation. 2015;189:5-15. doi: https://doi.org/10.1016/j.biocon.2014.09.047.

[13] Moseson H, Treleaven E, Gerdts C, Diamond-Smith N. The list experiment for measuring abortion: What we know and what we need. Studies in Family Planning. 2017:n/a-n/a. doi: 10.1111/sifp.12042.

[14] De Cao E, Lutz C. Sensitive survey questions: Measuring attitudes regarding female genital cutting through a list experiment. Oxf Bull Econ Stat. in press.

[15] Shell-Duncan B. From health to human rights: Female genital cutting and the politics of intervention. American Anthropologist. 2008;110(2):225-36. doi: 10.1111/j.1548-1433.2008.00028.x.

[16] Howard JA, Gibson MA. Frequency-dependent female genital cutting behaviour confers evolutionary fitness benefits. 2017;1:0049. doi: 10.1038/s41559-016-0049.

[17] Efferson C, Vogt S, Elhadi A, Ahmed HEF, Fehr E. Female genital cutting is not a social coordination norm. Science. 2015;349(6255):1446-7. doi: 10.1126/science.aaa7978.

[18] Hartung J, Abelson AE, Basu A, Basu MP, Beals KL, Chiarelli B, et al. On Natural Selection and the Inheritance of Wealth [and Comments and Reply]. Current Anthropology. 1976;17(4):607-22. doi: 10.1086/201799.

[19] Moya C, Snopkowski K, Sear R. What do men want? Re-examining whether men benefit from higher fertility than is optimal for women. Philosophical Transactions of the Royal Society B: Biological Sciences. 2016;371(1692). doi: 10.1098/rstb.2015.0149.

[20] Bellemare MF, Novak L, Steinmetz TL. All in the family: Explaining the persistence of female genital cutting in West Africa. Journal of Development Economics. 2015;116:252-65. doi: https://doi.org/10.1016/j.jdeveco.2015.06.001.

[21] Vogt S, Mohmmed Zaid NA, El Fadil Ahmed H, Fehr E, Efferson C. Changing cultural attitudes towards female genital cutting. Nature. 2016;538(7626):506-9. doi: 10.1038/nature20100.

[22] Shell-Duncan B, Hernlund Y. Female "Circumcision" in Africa: Culture, Controversy and Change. Boulder, Colorado, USA: Lynne Rienner; 2000.

[23] R Core Team. A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2016.

[24] Blair G, Imai K. R package: statistical methods for item count technique and list experiment v.8.0. 2010.

[25] Imai K. Multivariate regression analysis for the item count technique. J Am Stat Assoc. 2011;106(494):407-16.

[26] Nuno ANA, Bunnefeld N, Naiman LC, Milner-Gulland EJ. A novel approach to assessing the prevalence and drivers of illegal bushmeat hunting in the Serengeti. Conservation Biology. 2013;27(6):1355-65. doi: 10.1111/cobi.12124.

[27] Hinsley A, Nuno A, Ridout M, John FAVS, Roberts DL. Estimating the extent of CITES noncompliance among traders and end-consumers; Lessons from the global orchid trade. Conservation Letters. 2017;10(5):602-9. doi: 10.1111/conl.12316.

[28] Gibson MA, Mace R. An energy-saving development initiative increases birth rate and childhood malnutrition in rural Ethiopia. PLoS Med. 2006;3(4, e87):0476-83.

[29] Gibson MA, Gurmu E. Rural to urban migration is an unforeseen impact of development intervention in Ethiopia. PLoS ONE. 2012;7(11):e48708. doi: 10.1371/journal.pone.0048708.

[30] Gibson MA, Gurmu E. Land inheritance establishes sibling competition for marriage and reproduction in rural Ethiopia. Proceedings of the National Academy of Sciences. 2011;108(6):2200-4. doi: 10.1073/pnas.1010241108.

[31] Gibson MA, Mace R. Polygyny, reproductive success and child health in rural Ethiopia: Why marry a married man? Journal of Biosocial Science. 2007;39(2):287-300.

[32] Ostapczuk M, Musch J, Moshagen M. A randomized-response investigation of the education effect in attitudes towards foreigners. European Journal of Social Psychology. 2009;39(6):920-31. doi: 10.1002/ejsp.588.

[33] Camilotti G. Interventions to stop female genital cutting and the evolution of the custom: Evidence on age at cutting in Senegal. Journal of African Economies. 2016;25(1):133-58. doi: 10.1093/jae/ejv013.

[34] Varol N, Turkmani S, Black K, Hall J, Dawson A. The role of men in abandonment of female genital mutilation: a systematic review. BMC Public Health. 2015;15(1):1034. doi: 10.1186/s12889-015-2373-2.

[35] Henrich J, Boyd R. The evolution of conformist transmission and the emergence of between-group differences. Evol Hum Behav. 1998;19(4):215-41.

A Supporting Information

A.1 S1 Fig. Sensitive questions in four versions of the survey

I'd like you to imagine that you are the parent of a girl who has not yet become an adult. Family have wishes and aims for their children. There are some things that they hope will happen to them by the time they are adults, or which they will try to ensure happens if possible. There are other things that they do not particularly want or do not care about.
Here I have some cards. On each card is written something that you might (or might not) want for your daughter by the time she is an adult.

- If using survey 1A or 1B, read the following:
  In a moment I'll read out each card, all you have to do is tell me whether you would want this for a daughter.
  1A:

Would you want this for your daughter? (1=Yes, 0=No)			
Early marriage	Work in the city	Go to college	Live close to home

  1B:

Would you want this for your daughter? (1=Yes, 0=No)				
Early marriage	Work in the city	Go to college	FGC	Live close to home

- If using surveys 2A or 2B, read the following:
  We want to know about peoples' views about being a parent, but we also want them to be able to keep their views private so that we get honest answers. So please don't tell me which of the things on these cards you personally would want for your daughter. Instead, I'd like you to tell me how many of these cards show things that you want for your daughter.
  It's important that you don't tell me which individual things you are choosing, just give me a number. You can choose as many or as few as you like. If you'd like to hold or move the cards that is fine, but please don't tell me which particular cards you are choosing.
  OK: how many of these would you want for your daughter?
  2A:

Would you want this for your daughter? (1=Yes, 0=No)			
Early marriage	Work in the city	Go to college	Live close to home

2B:

Would you want this for your daughter? (1=Yes, 0=No)				
Early marriage	Work in the city	Go to college	FGC	To live close to home

Daughter: How many cards were selected?	Total number of cards

OK, we're going to repeat that a second time. This time, I'd like you to imagine that you are the parent of a man who is soon going to be married. I'm going to read out those things again, and this time, please think about which of these things you would want your son's wife to have. Your preferences might be the same as before, or a bit different this time. Either of these is fine; please just be honest.

2A:

Would you want this for your son's wife? (1=Yes, 0=No)			
Early marriage	Work in the city	Go to college	Live close to home

2B:

Would you want this for your son's wife? (1=Yes, 0=No)				
Early marriage	Work in the city	Go to college	FGC	To live close to home

Son's wife: How many cards were selected?	Total number of cards

A.2 S2 Fig. Generating the list, data quality tests and sampling strategy

Generating the list. To generate non-sensitive items for the list, focus groups discussions were conducted with local residents, who were asked to report popular local preferences regarding female relatives' and in-laws' traits. These discussions and a free-listing exercise generated an extended list of potential items, from which four were selected for inclusion in the survey. The final four items were chosen so as to minimise the chance of floor and ceiling effects that is, of participants preferring either all or none of the items. Such effects can be problematic because they effectively reveal the participant's attitude to the sensitive item [9, 11]. Following practices advocated in prior research to mitigate floor/ceiling effects, [11] four items were selected such that: a) one item was expected to be unpopular (early marriage) b) one item was expected to be popular (go to college) and c) two items were expected to be seen as incompatible (work in the city, and live close to home). Expectations regarding the popularity of different items were confirmed in a piloting stage (n=150), and low levels of floor/ceiling effects were observed in the final dataset. The % of respondents selecting all/none of the cards was $<4.5\%$.

A check for independence of responses was also undertaken to ensure that during direct questioning, adding the sensitive item did not change people's tendency to respond "yes" to the other four items. This "additional item" test was passed, there was no statistical difference between in responses to direct questions with and without the FGC card (t=1.04, p=0.297).

Sampling strategy. Our study was designed to ensure to there were adequate numbers and enough statistical power to perform UCT analyses (indirect questions were 70% of the total, n=1117), while reducing the relative number of responses to direct questions without the FGC card (4 card control group), which was included only to test the quality of the UTC list. Performing power calculations we identified we could achieve 80% power to detect an increase/decrease in the proportion of "yeses" in the direct question control and treatment groups we split the sample as follows: 20% (n=331) answered the direct question with the sensitive item, FGC, [5 item treatment], and 8% (n=177) answered the direct question without the sensitive item [4 item control].

Sampling method. All four versions of the survey were randomly assigned across households, and within household by gender and marital status. Each same-sex interviewer received a pile of surveys each morning, which had been randomly sorted by a data editor, including all 4 versions of the survey. Interviewers then travelled house to house, administering surveys to alternate households selected from a village plan supplied by the local district administrators. Accordingly, a random sample of 50% the households in the community were surveyed. Within each household, two surveys were completed by an equal and randomly selected sample of adult male and female, married and unmarried respondents from a household list supplied by the local authorities.

A.3 S1 Table. Sub group analyses between question methodology (direct versus unmatched count technique) and individual traits (gender, education level and age group), with combined estimates for FGC support (for both daughters and daughters in law), n=1620.

Respondent			n	Direct estimate (SE)	UCT estimate (SE)	P-values
male	no education	18-25 years	13	0.5 (0.340)	0.667 (0.819)	0.850
male	no education	26+ years	145	0.196 (0.051)	0 (0)	0.009
male	some education	18-25 years	245	0.085 (0.024)	0 (0)	0.091
male	some education	26+ years	408	0.027 (0.013)	0.450 (0.058)	<0.001
female	no education	18-25 years	35	0 (0)	0.390 (0.162)	0.016
female	no education	26+ years	388	0.1 (0.021)	0.308 (0.052)	<0.001
female	some education	18-25 years	239	0.049 (0.023)	0.199 (0.063)	0.026
female	some education	26+ years	147	0.036 (0.024)	0.246 (0.076)	0.008