CrossMark

ORIGINAL PAPER

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

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

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 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 2:58–77, 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

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

    María del Mar García Rueda
    mrueda@ugr.es

    Beatriz Rodríguez Cobo
    beacr@ugr.es

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

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

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, 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.

Keywords Auxiliary information · Calibration estimator · Domain estimator · Item count technique · Horvitz–Thompson estimator · Randomized response · Sensitive characteristic

Mathematics Subject Classification 62D05 · 62P25

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, colored 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 colored 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 and Larson 2002), the nonrandomized 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 Sect. 4, we discuss methodological advances for IST estimation under a generic sampling design. Specifically, a Horvitz–Thompson-type estimator is examined in Sect. 4.1, a calibration-type estimator is proposed in Sect. 4.2, and in Sect. 4.3 the calibration approach is employed for domain estimation. The results of various simulation experiments are presented and commented in Sect. 5. In particular, Sect. 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 nonstigmatizing 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

$$\hat{\bar{Z}}_{\mathrm{HT}} = \frac{1}{N} \sum_{i \in s_1} d_i z_i \quad \text{and} \quad \hat{\bar{T}}_{\mathrm{HT}} = \frac{1}{N} \sum_{i \in s_2} d_i t_i$$

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:

$$\hat{\bar{Y}}_{\mathrm{HT}} = \hat{\bar{Z}}_{\mathrm{HT}} - \hat{\bar{T}}_{\mathrm{HT}}. \tag{2}$$

From the unbiasedness of $\hat{\bar{Z}}_{\mathrm{HT}}$ and $\hat{\bar{T}}_{\mathrm{HT}}$, it readily follows that the estimator $\hat{\bar{Y}}_{\mathrm{HT}}$ is unbiased for $\bar{Y}$. In fact

$$\mathbb{E}_p(\hat{\bar{Y}}_{\mathrm{HT}}) = \mathbb{E}_p(\hat{\bar{Z}}_{\mathrm{HT}}) - \mathbb{E}_p(\hat{\bar{T}}_{\mathrm{HT}}) = \frac{1}{N} \sum_{i \in U} z_i - \frac{1}{N} \sum_{i \in U} t_i$$

$$= \frac{1}{N} \sum_{i \in U} (z_i - t_i) = \frac{1}{N} \sum_{i \in U} y_i.$$

The variance of $\hat{\bar{Y}}_{\mathrm{HT}}$, as long as the two samples s_1 and s_2 are independent, can be expressed as:

$$\mathbb{V}_p(\hat{\bar{Y}}_{\mathrm{HT}}) = \mathbb{V}_p(\hat{\bar{Z}}_{\mathrm{HT}}) + \mathbb{V}_p(\hat{\bar{T}}_{\mathrm{HT}})$$

$$= \frac{1}{N^2} \left(\sum_{i \in U} \sum_{j \in U} \Delta_{ij}(d_i z_i)(d_j z_j) + \sum_{i \in U} \sum_{j \in U} \Delta_{ij}(d_i t_i)(d_j t_j) \right),$$

where $\Delta_{ij} = \pi_{ij} - \pi_i \pi_j$. Finally, an unbiased estimator of $\mathbb{V}(\hat{\bar{Y}}_{\mathrm{HT}})$ is achieved by means of

$$\hat{\mathbb{V}}_p(\hat{\bar{Y}}_{\mathrm{HT}}) = \frac{1}{N^2} \left(\sum_{i \in s_1} \sum_{j \in s_1} \check{\Delta}_{ij}(d_i z_i)(d_j z_j) + \sum_{i \in s_2} \sum_{j \in s_2} \check{\Delta}_{ij}(d_i t_i)(d_j t_j) \right),$$

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 χ^2 distance function

$$\Phi_{s_j}(d_i, \omega_{ij}) = \sum_{i \in s_j} \frac{(\omega_{ij} - d_i)^2}{d_i q_i}, \quad j = 1, 2 \tag{3}$$

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 $\boldsymbol{\lambda} = (\lambda_1 \dots, \lambda_Q)^t$ is the vector of the Lagrange multipliers given by:

$$\boldsymbol{\lambda} = N^2 \mathbf{F}_{s_j}^{-1} (\bar{\mathbf{X}} - \hat{\bar{\mathbf{X}}}_{\mathrm{HT}}),$$

with $\mathbf{F}_{s_j} = \sum_{i \in s_j} d_i q_i \mathbf{x}_i \mathbf{x}_i^t$ and where $\hat{\bar{\mathbf{X}}}_{\mathrm{HT}}$ 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

$$\hat{\bar{Z}}_C = \frac{1}{N} \sum_{i \in s_1} \omega_{i1} z_i = \hat{\bar{Z}}_{\mathrm{HT}} + (\bar{\mathbf{X}} - \hat{\bar{\mathbf{X}}}_{\mathrm{HT}})^t \hat{\mathbf{B}}_{s_1}$$

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

$$\hat{\bar{T}}_C = \frac{1}{N} \sum_{i \in s_2} \omega_{i2} t_i = \hat{\bar{T}}_{\mathrm{HT}} + (\bar{\mathbf{X}} - \hat{\bar{\mathbf{X}}}_{\mathrm{HT}})^t \hat{\mathbf{B}}_{s_2}$$

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:

$$\mathbb{V}_p(\hat{\bar{Y}}_C) = \mathbb{V}_p(\hat{\bar{Z}}_C) + V_p(\hat{\bar{T}}_C)$$

$$= \frac{1}{N^2} \left(\sum_{i \in U} \sum_{j \in U} \Delta_{ij} (d_i E_i)(d_j E_j) + \sum_{i \in U} \sum_{j \in U} \Delta_{ij} (d_i G_i)(d_j G_j) \right),$$

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:

$$\hat{\mathbb{V}}_p(\hat{\bar{Y}}_C) = \frac{1}{N^2} \left(\sum_{i \in s_1} \sum_{j \in s_1} \check{\Delta}_{ij} (d_i e_i)(d_j e_j) + \sum_{i \in s_2} \sum_{j \in s_2} \check{\Delta}_{ij} (d_i g_i)(d_j g_j) \right), \quad (7)$$

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 Sect. 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:

$$\hat{\bar{Y}}_{\text{HT},d} = \frac{1}{N_d} \sum_{i \in s_{1d}} d_i z_i - \frac{1}{N_d} \sum_{i \in s_{2d}} d_i t_i.$$

The estimator $\hat{\bar{Y}}_{\text{HT},d}$ is design-unbiased. In fact,

$$\mathbb{E}_p(\hat{\bar{Y}}_{\text{HT},d}) = \frac{1}{N_d} \mathbb{E}_p \left(\sum_{i \in s_{1d}} d_i z_i \right) - \frac{1}{N_d} \mathbb{E}_p \left(\sum_{i \in s_{2d}} d_i t_i \right)$$

$$= \frac{1}{N_d} \mathbb{E}_p \left(\sum_{i \in s_1} d_i z_i \delta_{di} \right) - \frac{1}{N_d} \mathbb{E}_p \left(\sum_{i \in s_2} d_i t_i \delta_{di} \right)$$

$$= \frac{1}{N_d} \sum_U z_i \delta_{di} - \frac{1}{N_d} \sum_U t_i \delta_{di}$$

$$= \frac{1}{N_d} \sum_{i \in U_d} z_i - \frac{1}{N_d} \sum_{i \in U_d} t_i$$

$$= \frac{1}{N_d} \sum_{i \in U_d} (z_i - t_i) = \frac{1}{N_d} \sum_{i \in U_d} y_i.$$

The variance of $\hat{\bar{Y}}_{HT,d}$ is given by:

$$\mathbb{V}_p(\hat{\bar{Y}}_{HT,d}) = \frac{1}{N_d^2}\left(\sum_{i \in U_d}\sum_{j \in U_d}\Delta_{ij}(d_i z_i)(d_j z_j) + \sum_{i \in U_d}\sum_{j \in U_d}\Delta_{ij}(d_i t_i)(d_j t_j)\right),$$

which can be unbiasedly estimated with

$$\hat{\mathbb{V}}_p(\hat{\bar{Y}}_{HT,d}) = \frac{1}{N_d^2}\left(\sum_{i \in s_{1d}}\sum_{j \in s_{1d}}\check{\Delta}_{ij}(d_i z_i)(d_j z_j) + \sum_{i \in s_{2d}}\sum_{j \in s_{2d}}\check{\Delta}_{ij}(d_i t_i)(d_j t_j)\right).$$

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 Sect. 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 Sect. 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 3 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 $\hat{\bar{Y}}_{\mathrm{HT}_y}$ 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:

$$\hat{\bar{Y}}_{\mathrm{SRM}} = \frac{1}{N} \sum_{i \in s} d_i r_i,$$

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 $\hat{\bar{Y}}^* = \hat{\bar{Y}}_{HT_y}, \hat{\bar{Y}}_{C_y}, \hat{\bar{Y}}_{HT}, \hat{\bar{Y}}_{C}, \hat{\bar{Y}}_{SRM_1}, \hat{\bar{Y}}_{SRM_2}, \hat{\bar{Y}}_{C_1}, \hat{\bar{Y}}_{C_2}$:

$$|\text{RB}(\hat{\bar{Y}}^*)| = \left| \frac{\mathbb{E}_M(\hat{\bar{Y}}^*) - \bar{Y}}{\bar{Y}} \right| \quad \text{and} \quad \text{RMSE}(\hat{\bar{Y}}^*) = \frac{\mathbb{E}_M(\hat{\bar{Y}}^* - \bar{Y})^2}{\bar{Y}^2},$$

where $\hat{\bar{Y}}_{C_i}$ 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 Fig. 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 $\hat{\bar{Y}}_{HT}$. This is particularly true for the estimates obtained using SRM_2, which are generally less efficient than those provided by $\hat{\bar{Y}}_{HT}$. As regards the estimates under SRM_1, in some cases across the three sampling designs, they appear to be slightly more efficient than $\hat{\bar{Y}}_{HT}$ 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.
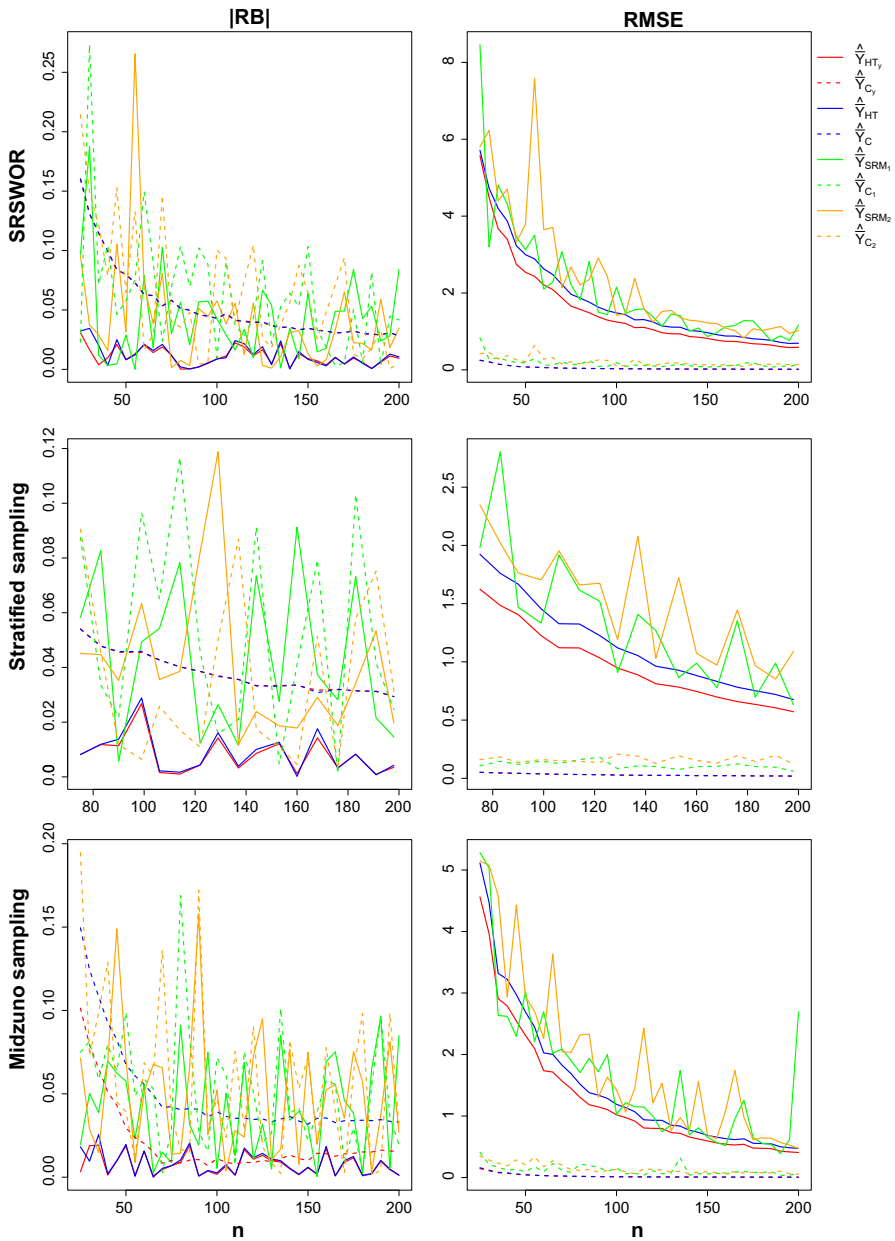
Fig. 1 Performance of the HT and calibration estimators under DQ, IST and RRT survey modes

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.

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 Fig. 2.

We observe that two HT-type estimators $\hat{\bar{Y}}_{HT_1}$ and $\hat{\bar{Y}}_{HT_2}$, 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, Fig. 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.

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 $\hat{\bar{Y}}_{C,G}$ denote the IST calibration estimator for the list of items which includes G innocuous variables, $G = 1, \ldots, 5$. Hence, for each $\hat{\bar{Y}}_{C,G}$, we computed the absolute RB and RMSE as in Sect. 5.1.
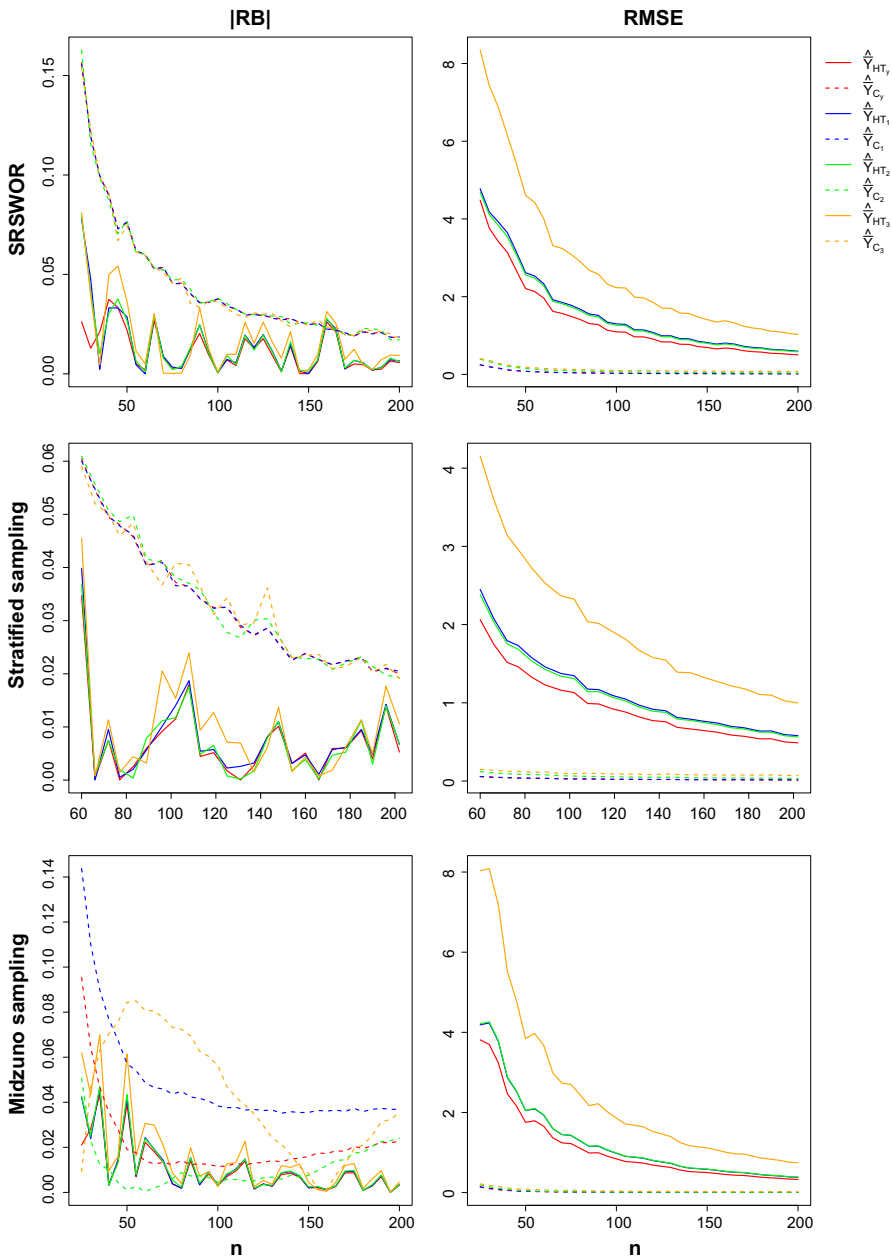
Fig. 2 Performance of the HT and calibration estimators under DQ and IST survey modes and for different correlations

The results obtained are shown in Fig. 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
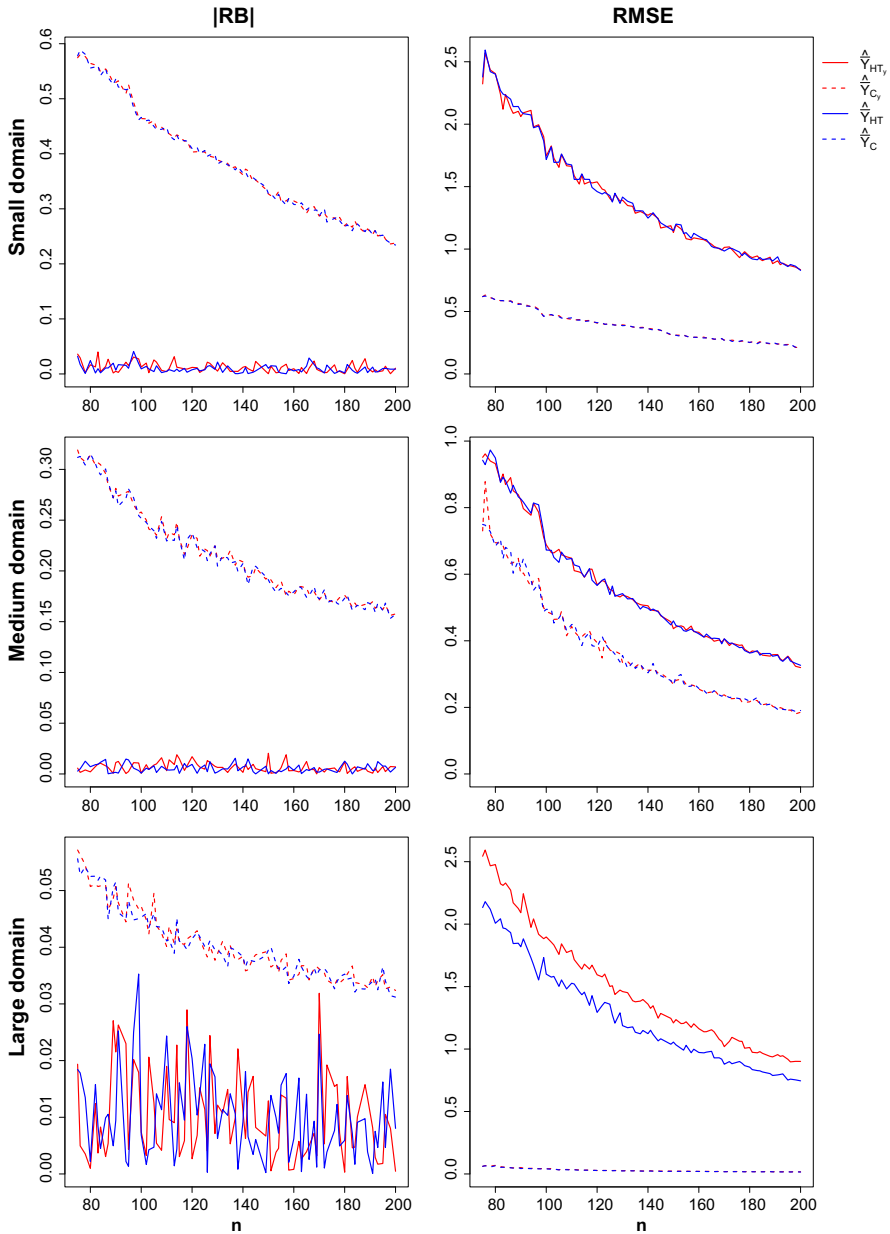
Fig. 3 Stratified domain estimates by the HT and calibration estimators under DQ and IST survey modes

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
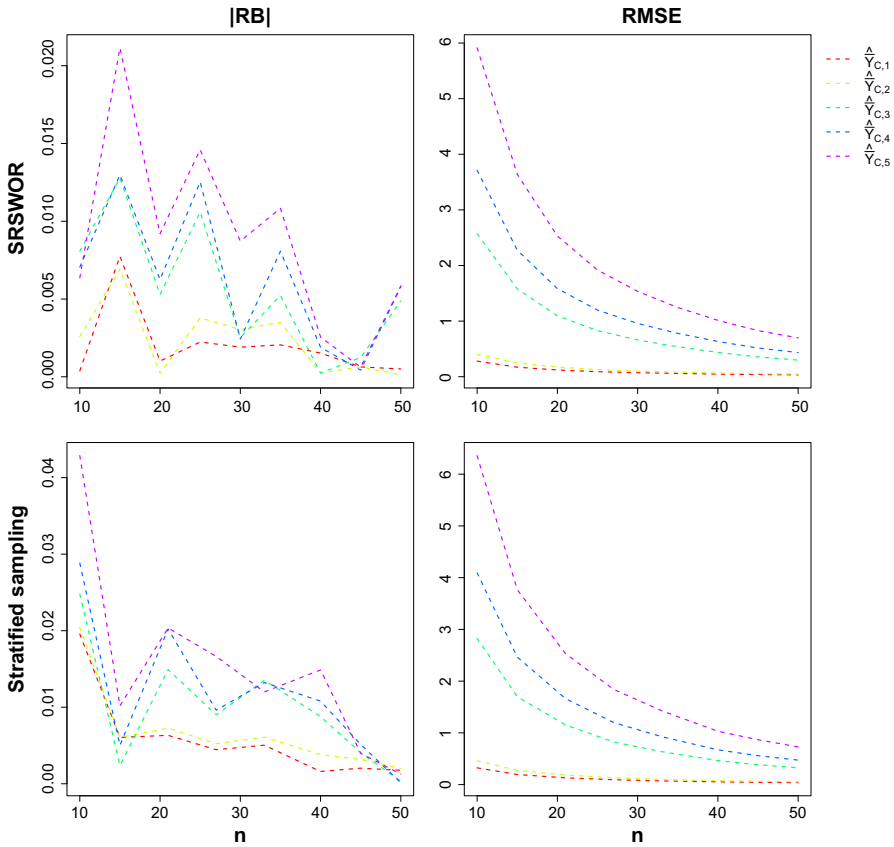
Fig. 4 Performance of the IST calibration estimators with an increasing number of innocuous items

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.

Finally, another simulation was run to investigate the behavior of the variance estimator of $\hat{\bar{Y}}_{C,G}$. 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 $\hat{\bar{Y}}_{C,G}$ is achieved by the simulated variance:

$$\mathbb{V}_{\text{sim}}(\hat{\bar{Y}}_{C,G}) = \frac{1}{50{,}000} \sum_{k=1}^{50{,}000} \left(\hat{\bar{Y}}_{C,G}^{(k)} - \bar{Y} \right)^2$$

where $\hat{\bar{Y}}_{C,G}^{(k)}$ is the calibration-type estimate computed on the kth sample and $G = 1, \ldots, 5$;
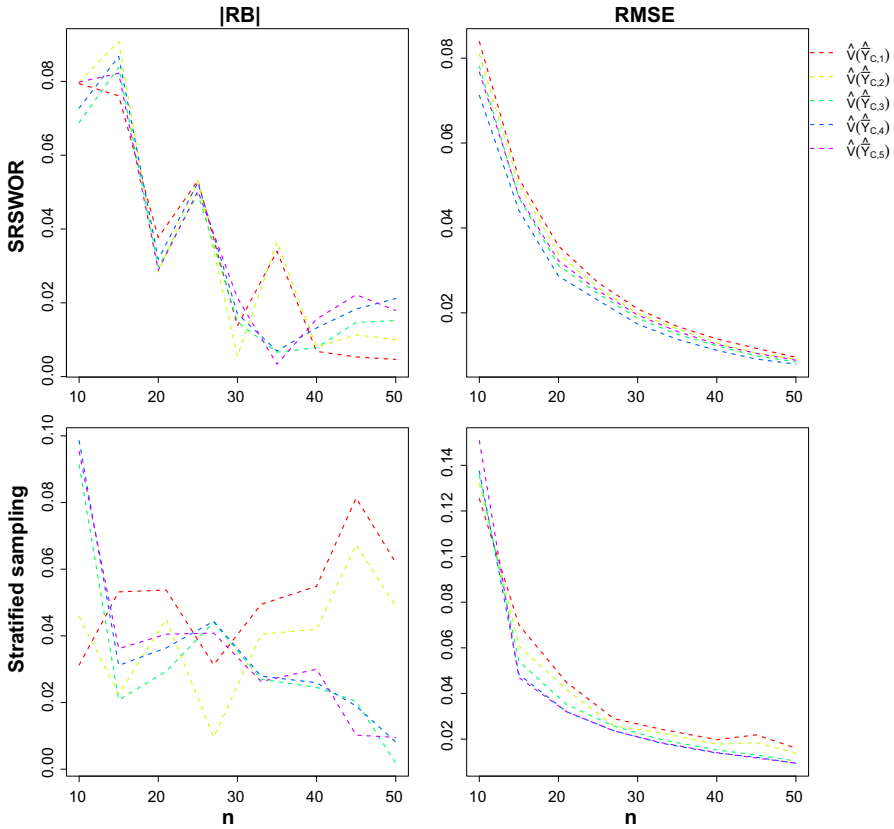
Fig. 5 Performance of the variance estimator for the IST calibration estimators with an increasing number of innocuous items

2. 10,000 Monte Carlo samples are drawn from Shaw's population according to SRSWOR and stratified SRSWOR, and variance estimates $\hat{\mathbb{V}}(\hat{\bar{Y}}_{C,G})$ are computed as reported in (7);

3. The absolute relative bias and relative mean squared error for the variance estimates are computed as:

$$|\text{RB}(\hat{\mathbb{V}}(\hat{\bar{Y}}_{C,G}))| = \left| \frac{\mathbb{E}_M(\hat{\mathbb{V}}(\hat{\bar{Y}}_{C,G})) - \mathbb{V}_{\text{sim}}(\hat{\bar{Y}}_{C,G})}{\mathbb{V}_{\text{sim}}(\hat{\bar{Y}}_{C,G})} \right|$$

and

$$\text{RMSE}(\mathbb{V}(\hat{\bar{Y}}_{C,G})) = \frac{\mathbb{E}_M(\hat{\mathbb{V}}(\hat{\bar{Y}}_{C,G}) - \mathbb{V}_{\text{sim}}(\hat{\bar{Y}}_{C,G}))^2}{(\mathbb{V}_{\text{sim}}(\hat{\bar{Y}}_{C,G}))^2}.$$

Figure 5 shows the behavior of the absolute RB and the RMSE for different sample sizes and under the two sampling designs.

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.

References

Arcos, A., Rueda, M.M., Singh, S.: Generalized approach to randomized response for quantitative variables. Qual. Quant. **49**, 1239–1256 (2015)

Arias, A., Sutton, S.G.: Understanding recreational fishers compliance with no-take zones in the Great Barrier Reef Marine Park. Ecol. Soc. (2013). https://doi.org/10.5751/ES-05872-180418

Arnab, R., Singh, S.: Randomized response techniques: an application to the Botswana AIDS impact survey. J. Stat. Plan. Inference **140**, 941–953 (2010)

Aronow, P.M., Coppock, A., Crawford, F.W., Green, D.P.: Combining list experiment and direct question estimates of sensitive behavior prevalence. J. Surv. Stat. Methodol. **3**, 43–66 (2015)

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

Blair, G., Imai, K.: List: statistical methods for the item count technique and list experiment. http://CRAN.Rproject.org/package=list (2010)

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

Blair, G., Imai, K., Lyall, J.: Comparing and combining list and endorsement experiments: evidence from Afghanistan. Am. J. Polit. Sci. **58**, 1043–1063 (2014)

Blank, S.G., Gavin, M.C.: 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. Environ. Conserv. **36**, 112–119 (2009)

Chaloupka, M.Y.: Application of the randomized response technique to marine park management: an assessment of permit compliance. Environ. Manag. **9**, 393–398 (1985)

Chaudhuri, A.: Randomized Response and Indirect Questioning Techniques in Surveys. Chapman & Hall, Boca Raton (2011)

Chaudhuri, A., Christofides, T.C.: Item count technique in estimating the proportion of people with a sensitive feature. J. Stat. Plan. Inference **137**, 589–593 (2007)

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

Chaudhuri, A., Mukerjee, R.: Randomized Response: Theory and Techniques. Marcel Dekker Inc, New York (1988)

Conteh, A., Gavin, M.C., Solomon, J.: Quantifying illegal hunting: a novel application of the randomised response technique. Biol. Conserv. **189**, 16–23 (2015)

Christofides, T.C.: A new version of the item count technique. Model Assist. Stat. Appl. **10**, 289–297 (2015)

Deville, J.C., Särndal, C.E.: Calibration estimators in survey sampling. J. Am. Stat. Assoc. **87**, 376–382 (1992)

Diana, G., Perri, P.F.: Estimating a sensitive proportion through randomized response procedures based on auxiliary information. Stat. Pap. **50**, 661–672 (2009)

Diana, G., Perri, P.F.: New scrambled response models for estimating the mean of a sensitive quantitative character. J. Appl. Stat. **37**, 1875–1890 (2010)

Diana, G., Perri, P.F.: A class of estimators for quantitative sensitive data. Stat. Pap. **52**, 633–650 (2011)

Diana, G., Perri, P.F.: A calibration-based approach to sensitive data: a simulation study. J. Appl. Stat. **39**, 53–65 (2012)

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

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

Droitcour, J.A., Larson, E.M.: An innovative technique for asking sensitive questions: the three-card method. Bull. Soc. Methodol. **75**, 5–23 (2002)

Eichhorn, B.H., Hayre, L.S.: Scrambled randomized response methods for obtaining sensitive quantitative data. J. Stat. Plan. Inference **7**, 306–316 (1983)

Fox, J.A., Tracy, P.E.: Randomized Response: A Method for Sensitive Survey. Sage Publication Inc, Newbury Park (1986)

Glynn, A.N.: What can we learn with statistical truth serum? Design and analysis of the list experiment. Pub. Opin. Q. **77**, 159–172 (2013)

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

Holbrook, A.L., Krosnick, J.A.: Measuring voter turnout by using the randomized response technique: evidence calling into question the method's validity. Pub. Opin. Q. **74**, 328–343 (2010a)

Holbrook, A.L., Krosnick, J.A.: Social desirability bias in voter turnout reports: tests using the item count technique. Pub. Opin. Q. **74**, 37–67 (2010b)

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

Houston, J., Tran, A.: A survey of tax evasion using the randomized response technique. Adv. Tax. **13**, 69–94 (2001)

Hussain, Z., Shah, E.A., Shabbir, J.: An alternative item count technique in sensitive surveys. Revista Colombiana de Estadistica **35**, 39–54 (2012)

Hussain, Z., Shabbir, N., Shabbir J.: An alternative item sum technique for improved estimators of population mean in sensitive surveys. *Hacet. J. Math. Stat.*, First published online **46**, 907–934 (2017)

Imai, K.: Multivariate regression analysis for the item count technique. J. Am. Stat. Assoc. **106**, 407–416 (2011)

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

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

Jann, B., Jerke, J., Krumpal, I.: Asking sensitive questions using the crosswise model: an experimental survey measuring plagiarism. Pub. Opin. Q. **76**, 32–49 (2012)

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

Korndörfer, M., Krumpal, I., Schmukle, S.C.: Measuring and explain tax evasion: improving self-reports using the crosswise model. J. Econ. Psychol. **45**, 18–32 (2014)

Kott, P.S.: Developing calibration weights and standard-error estimates for a survey of drug-related emergency-department visits. J. Off. Stat. **30**, 521–532 (2014)

Kott, P.S., Chang, T.: Using calibration weighting to adjust for nonignorable unit nonresponse. J. Am. Stat. Assoc. **105**, 1265–1275 (2010)

Krumpal, I.: Estimating the prevalence of xenophobia and anti-semitism in Germany: a comparison of the randomized response technique and direct questioning. Soc. Sci. Res. **41**, 1387–1403 (2012)

Lara, D., Strickler, J., Olavarrieta, C.D., García, S.G., Ellertson, C.: Measuring induced abortion in Mexico: a comparison of four methodologies. Soc. Methods Res. **32**, 529–558 (2004)

Lara, D., García, S.G., Ellertson, C., Camlin, C., Suaréz, J.: The measure of induced abortion in Mexico using random response technique. Soc. Methods Res. **35**, 279–301 (2006)

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

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

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

Miller, J.D.: The nominative technique: a new method of estimating heroin prevalence. NIDA Res. Monogr. **57**, 104–124 (1985)

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

Ostapczuk, M., Musch, J., Mashagen, M.: A randomized-response investigation of the education effect in attitudes towards foreigners. Eur. J. Soc. Psychol. **39**, 920–931 (2009)

Perri, P.F., Diana, G.: Scrambled response models based on auxiliary variables. In: Torelli, N., Pesarin, F., Bar-Hen, A. (eds.) Advances in Theoretical and Applied Statistics, pp. 281–291. Spriger-Verlag, Berlin (2013)

Perri, P.F., Pelle, E., Stranges, M.: Estimating induced abortion and foreign irregular presence using the randomized response crossed model. Soc. Indic. Res. **129**, 601–618 (2016)

Raghavarao, D., Federer, W.F.: Block total response as an alternative to the randomized response method in survey. J. R. Stat. Soc. B **41**, 40–45 (1979)

Rueda, M., Cobo, B., Arcos, A.: An improved class of estimators in RR surveys. Math. Methods Appl. Sci. (2017). https://doi.org/10.1002/mma.4256

Särndal, C.E.: The calibration approach in survey theory and practice. Surv Methodol. **33**, 99–119 (2007)

Särndal, C.E., Lundström, S.: Estimation in Survey with Nonresponse. Wiley, New York (2005)

Schill, D.J., Kline, P.A.: Use of random response to estimate angler noncompliance with fishing regulations. North Am. J. Fish. Manag. **15**, 721–731 (1995)

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

Shaw, P.: Estimating a finite population mean of a sensitive quantitative variable from a single probability sample by the Item Count Technique. Model Assist. Stat. Appl. **10**, 411–419 (2015)

Shaw, P.: Estimating a finite population proportion bearing a sensitive attribute from a single probability sample by Item Count Technique. In: Chaudhuri, A., Christofides, T.C., Rao, C.R. (eds.) Handbook of Statistics Vol. 34: Data Gathering, Analysis and Protection of Privacy Through Randomized Response Techniques: Qualitative and Quantitative Human Traits, pp. 387–404. Elsevier, Amsterdam (2016)

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

Solomon, J., Jacobson, S.K., Wald, K.D., Gavin, M.: Estimating illegal resources use at the Ugandan park with the randomized response technique. Hum. Dimens. Wildlife **12**, 75–88 (2007)

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

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

Sukhatme, P.V., Sukhatme, B.V., Sukhatme, S., Asok, C.: Sampling Theory of Surveys with Applications. Iowa State University Press, Ames (1984)

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

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

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

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