Revista Colombiana de Estadística
Junio 2012, volumen 35, no. 1, pp. 39 a 54

An Alternative Item Count Technique in Sensitive Surveys

Una técnica alternativa de conteo de ítems en encuestas sensitivas

Zawar Hussain[1,2,a], Ejaz Ali Shah[2,b], Javid Shabbir[2,c]

[1]Department of Statistics, Faculty of Sciences, King Abdulaziz University, Jeddah, Kingdom of Saudi Arabia

[2]Department of Statistics, Quaid-I-Azam University, Islamabad, Pakistan

Abstract

The present study is basically meant to propose an improved item count technique which will mainly have an impact on sensitive fields such as health care. It is attempted to highlight the scope of the proposal relative to the usual and existing methods serving the same purpose. The proposed improved Item Count Technique (ICT) has the major advantage that it does not require two subsamples (as is the case in usual ICT) and there is no need of finding optimum subsample sizes. The proposed ICT has been observed performing well, as compared to the usual ICT, in terms of relative efficiency. The innovative method of Randomized Response (RR) technique has also been compared with the proposed ICT and it is found that the proposed technique uniformly performs better when the number of innocuous items is greater than 3.

***Key words*:** Health surveys, Privacy, Proportion estimation, Randomized response, Sensitive question.

Resumen

El presente articulo propone una técnica de conteo de items con aplicaciones principalmente en el campo de la salud. Se muestran las ventajas de nuestra propuesta y de otros métodos que sirven con el mismo fin. La técnica de conteo de ítems propuesta (ICT, por su sigla en inglés) tiene la ventaja de que no requiere dos submuestras (como es el caso en el ICT clásico) y no es necesario de encontrar los tamaños de las submuestras óptimos. El ICT propuesto tiene un mejor comportamiento en términos de eficiencia relativa. El método de la técnica de respuesta aleatorizada (RR, por su sigla en inglés) es también comparado con el ICT propuesto y se encuentra que la técnica

[a]Professor. E-mail: zhlangah@yahoo.com
[b]Professor. E-mail: alishah_ejaz@yahoo.com
[c]Professor. E-mail: jsqau@yahoo.com

propuesta se desempeña mejor cuando el número de ítems inocuos es mayor de 3.

***Palabras clave*:** encuestas de salud, estimación de la proporción, preguntas sensibles, privacidad, respuesta al azar.

1. Introduction

In estimating the population proportion of a sensitive characteristic (induced abortion, shoplifting, tax evasion) through direct questioning, truthfulness of the answers may be suspected due to various reasons, namely, social stigma, embarrassment, monetary penalty, and many others. These and similar other factors are directly related to the health issues and some improved/alternative techniques to hit these areas are indispensable to address the complications involved in them. There are a number of papers showing such concerns. Some literature in this regard may be seen in Bjorner, Kosinski & Ware (2003) and Martin, Kosinski, Bjorner, Ware & MacLean (2007), and the references therein.

An ingenious alternative to direct questioning introduced by Warner (1965), known as Randomized Response Technique (RRT), has been developed rapidly. For a good review of developments on RRTs we would refer the reader to Tracy & Mangat (1996) and Chaudhuri & Mukherjee (1988). The RRT has been used in many studies including Liu & Chow (1976), Reinmuth & Geurts (1975), Geurts (1980), Larkins, Hume & Garcha (1997), etc. Geurts (1980) reported that RRT had financial limitations since it requires larger sample sizes to obtain the confidence intervals comparable to the direct questioning technique. More time is needed to administer and explain the procedure to the survey respondents. In addition, tabulation and calculation of the results are comparatively laborious. Larkins et al. (1997) found that RRT was not a good alternative for estimating the proportion of tax payers/non-payers. Dalton & Metzger (1992) were of the view that RRT might not be effective through a mailed or telephonic survey. Hubbard, Casper & Lessler (1989) stated that the main technical problem for RRTs is making the decision about what kind of the randomization device would be the best in a given situation, and that the most crucial aspect of the RRT is about the respondent's acceptance of the technique. Chaudhuri & Christofides (2007) also gave a criticism on the RRT in the sense that it demands the respondent's skill of handling the device and also asks respondents to report the information which may be useless or tricky. A clever respondent may also think that his/her reported response can be traced back to his/her actual status if he/she does not understand the mathematical logic behind the randomization device. Some of the alternatives to the RR technique include the Item Count Technique (Droitcour, Caspar, Hubbard, Parsley, Visscher & Ezzati 1991), the Three card method (Droitcour, Larson & Scheuren 2001), and the Nominative technique (Miller 1985). These alternatives are designed because, in general, respondent evade sensitive questions especially regarding personal issues, socially deviant behaviors or illegal acts. Chaudhuri & Christofides (2007) also added that in these three alternatives to RRT respondents know that what they are revealing about themselves and they do not need to know about any special estimation technique. Also respondents provide answers which make sense to them.

2. Item Count Techniques

In order to estimate the proportion of people with a stigmatizing attribute a promising indirect questioning technique called Item Count Technique (ICT), was introduced by Droitcour et al. (1991). It consists of taking two subsamples of sizes n_1 and n_2. The ith respondent in the first subsample is given a list of g innocuous items and asked to report the number, say X_i of items that are applicable to them ($X_i \leq g$). Similarly, the jth respondent in the second subsample is provided another list of $(g+1)$ items including the sensitive item and asked to report a number, say Y_j of the items that are applicable to them ($Y_j \leq g+1$). The g innocuous items may or may not be the same in both subsamples. An unbiased estimator of the proportion of sensitive item in the population say π is given by:

$$\widehat{\pi}_I = \overline{Y} - \overline{X} \tag{1}$$

where $\overline{Y}$ and $\overline{X}$ represent the sample mean from the second and first subsamples, respectively.

To our knowledge, no author has given the variance expression of the estimator given in (1). We have derived the variance of the estimator in (1), and it is given by:

$$V\left(\widehat{\pi}_I\right) = \frac{\pi\left(1-\pi\right)}{n_2} + \frac{n \sum\limits_{j=1}^{g} \theta_j \left(1 - \sum\limits_{j=1}^{g} \theta_j\right)}{n_1 n_2} + \frac{n \sum\limits_{\substack{j,k=1 \\ j \neq k}}^{g} \theta_j \theta_k}{n_1 n_2} \tag{2}$$

where θ_j is the known proportion of the item j in the population. More details about ICT can be found in Droitcour et al. (1991) and Droitcour & Larson (2002). Dalton, Wimbush & Daily (1994) named ICT as the unmatched count technique and applied it to study the illicit behaviors of the auctioneers, and as compared to the direct questioning they obtained higher estimates of six stigmatized items. Wimbush & Dalton (1997) applied this technique in estimating the employee theft rate in high-theft-exposure business and found higher theft rates. Tsuchiya (2005) extended the ICT to domain estimators by the stratified method, the cross-based method, and the double cross-based method. More recently, Tsuchiya, Hirai & Ono (2007) studied the properties of the ICT through an experimental web survey and found that ICT yielded higher estimates of the proportions of the shoplifters by nearly 10% as that of yielded by direct questioning. They also found that the cross-based method was the most appropriate one.

Besides its fruitful applications ICT has not been found fruitful in many studies; for example, Droitcour et al. (1991), Biemer & Wright (2004) and Ahart & Sackett (2004) failed to get higher estimates in their studies of different stigmatized traits. We have focused on the issue of the need of two subsamples in the usual application of ICT and have proposed an alternative ICT which does not need two subsamples. Avoiding the need of two subsamples for our proposed ICT makes it more attractive in terms of cost and statistical efficiency. The following section provides a description of the proposed methodology.

2.1. Proposed Item Count Technique

Each respondent in a sample of size n is provided a questionnaire (list of questions) consisting of $g\,(\geq 2)$ questions. The jth question consists of queries about an unrelated item (F_j), and a sensitive characteristic (S). The respondent is requested to count 1 if he/she possesses at least one of the characteristics F_j and S, otherwise, count 0, as a response to the jth question, and to report the total count based on entire questionnaire.

The list of items is given to the respondents and they are sent to another room so that they are unseen to the interviewer. To illustrate, suppose the sensitive study item (S) be the cheating in exams and the unrelated items $(F_j, j = 1, 2.)$ are: (i) "Do you live in the hostel?" and (ii) "Is the last digit of your registration number odd?" It is obvious that there are almost (if not exactly) 50% (known) of the students having an odd registration number and proportion of the students living in hostel is easily available from the warden office. Let Z_i denote the total count of ith respondent, and then mathematically we can write it as:

$$Z_i = \sum_{j=1}^{g} \alpha_j \tag{3}$$

where α_j can assume values "1" and "0" with probabilities $(\pi + \theta_j - \pi\theta_j)$ and $(1 - \pi - \theta_j + \pi\theta_j)$, respectively.

Taking expectation on (3) we have:

$$\begin{aligned} E\left(Z_i\right) &= \sum_{j=1}^{g} E\left(\alpha_j\right) = g\pi + \sum_{j=1}^{g} \theta_j - \pi \sum_{j=1}^{g} \theta_j \\ &= \left(g - \sum_{j=1}^{g} \theta_j\right) \pi + \sum_{j=1}^{g} \theta_j \end{aligned}$$

This suggests defining an unbiased estimator of π as:

$$\widehat{\pi}_P = \frac{\overline{Z} - \sum_{j=1}^{g} \theta_j}{g - \sum_{j=1}^{g} \theta_j} \tag{4}$$

The estimator given in (4) serves the purpose of estimating π as is done by $\widehat{\pi}_I$ in (1). The estimator $\widehat{\pi}_P$ obtained through our proposed ICT does not demand two subsamples which are needed by $\widehat{\pi}_I$ based on the usual ICT. This property (avoiding the need of two subsamples) makes our proposal more attractive and practicable.

The variance of the estimator $\widehat{\pi}_P$ is given by (see Appendix)

$$V\left(\widehat{\pi}_P\right) = \frac{\pi\left(1-\pi\right)}{n} + \frac{\left(1-\pi\right)}{n\left(g-\sum_{j=1}^{g}\theta_j\right)^2}\left\{\left(\sum_{j=1}^{g}\theta_j\right)\left(1-\sum_{j=1}^{g}\theta_j\right)+\sum_{\substack{j,k=1\\ j\neq k}}^{g}\theta_j\theta_k\right\} \tag{5}$$

Some comments are in order. It is to be noted that in some surveys it may be possible to have unrelated traits $(F_j, j=1,2,\ldots,g)$ with equal proportions $(\theta_j, j=1,2,\ldots,g)$. In these situations we have $\theta_j=\frac{1}{g}$ for all j and consequently the variance of the proposed estimator $\widehat{\pi}_P$ reduces to

$$V\left(\widehat{\pi}_P\right) = \frac{\pi\left(1-\pi\right)}{n} + \frac{\left(1-\pi\right)}{ng\left(g-1\right)} \tag{6}$$

As pointed by the two referees, it is just possible that the actual status of the respondents about one (or all) the unrelated item(s) may be known to the interviewer by any means, then the response of 0 or g would disclose his/her status about the sensitive item. In this case privacy protection provided to the respondents will be limited. Thus, the unrelated items should be chosen in such a way that the actual status of the respondents about at least one of the unrelated items must be impossible to know by any means. To fix the idea, suppose the unrelated items are (i) and (ii) as we discussed above, then knowing the residential status of a particular student is difficult while actually conducting the survey but the proportion of students living in hostel may be readily available from the warden office. Similar is the case with the unrelated item of registration number. If it is possible to exactly guess or know about the particular item(s) for a given individual then such item(s) must not be included in the group of items. In this way, respondents would feel more protected and be motivated to answer truthfully. And, of course, the interviewer's ethical responsibility of being honest is more apparent, in the sense that he would be asking about those items about which he knows nothing of a particular respondent. The item count technique surveys are conducted in the hope that the respondents will be motivated more to reveal truthful answers rather than trapping them in mathematical tricks to trace their actual responses on the sensitive items. It will essentially be a direct questioning situation if surveyor is able to know the status of each respondent on each unrelated item. So, respondents must be assured that it is impossible to know the status of individual about an item but, of course, its population proportion is known somehow. It is easy to understand now that knowing the population proportion of an unrelated item is not harmful but knowing the individuals' status is. Moreover, another characteristic of such indirect survey methods is the anonymity. The identity (in terms of name or registration number, etc.) of the respondent is not required. The respondents may just write their answers on a sheet of paper and drop them in a box making it impossible to know the response of a particular respondent even the interviewer is able to know the status of a particular respondent on a given item. For example, in our situation, if the surveyor is able enough to guess or know the residential status (*hostelite* or *non-hostelite*) of a student, due to anonymity, he/she is not able to know reported response of a given respondent. Thus, any unrelated item whose population proportion is known may be used in this technique.

The acceptance of the unrelated question by the respondents, as pointed by the two learned referees, is another key issue of concern. In some cases, it would be needed to explain the working of whole the technique to the respondents. But it depends on the nature and composition of the population. In such cases survey must be conducted under the supervision of a trained statistician. More specifically, if the studied population is composed of illiterate individuals the technique must be explained to them prior to actually conducting the survey. The explanation of the technique would possibly decrease the suspicion among the respondents of being tricked. Further, the suspicion depends upon the anonymity provided by the survey method. If the respondents are explained about the working of the survey in such a way that their anonymity is assured and they are giving meaningful answers in the sense that only population proportion of study item is estimated and individual's status can not be known through their reported response. With this explanation and provision of anonymity it is anticipated that any unrelated item with known population proportion of prevalence may be fairly used. One more thing about the acceptance of unrelated items by the respondents is the simplicity of the question. The unrelated question must not be an open ended or having multiple answers, that is, it must be a binary item.

3. Performance Evaluations and Comparison

In this section, we provide efficiency comparisons of the estimator $\widehat{\pi}_P$ of the proposed ICT with the $\widehat{\pi}_I$ of the usual ICT and another obtained through RRT of Warner (1965). As we have discussed, that ICT has been developed as an alternative to RRT, so we have also compared our technique with RR technique proposed by Warner (1965).

3.1. Proposed versus Usual ICT

We compare the proposed estimator $\widehat{\pi}_P$ with the usual ICT estimator $\widehat{\pi}_I$ in both the situations of having and not having unequal $\theta_j = \frac{1}{g}$. In case of having unequal $\theta_j^{\prime}s$ the proposed estimator $\widehat{\pi}_P$ would be more efficient than the estimator $\widehat{\pi}_I$ if

$$V\left(\widehat{\pi}_I\right) - V\left(\widehat{\pi}_P\right) \geq 0,$$

$$\frac{\pi\left(1-\pi\right)}{n_2} + \frac{n\sum\limits_{j=1}^{g}\theta_j\left(1-\sum\limits_{j=1}^{g}\theta_j\right)}{n_1 n_2} + \frac{n\sum\limits_{\substack{j,k=1\\ j\neq k}}^{g}\theta_j\theta_k}{n_1 n_2}$$

$$-\frac{\pi(1-\pi)}{n}-\frac{(1-\pi)}{n\left(g-\sum\limits_{j=1}^{g}\theta_j\right)^2}\left\{\left(\sum_{j=1}^{g}\theta_j\right)\left(1-\sum_{j=1}^{g}\theta_j\right)+\sum_{\substack{j,k=1\\ j\neq k}}^{g}\theta_j\theta_k\right\}\geq 0$$

$$\frac{\pi(1-\pi)n_1}{nn_2}+\left\{\sum_{j=1}^{g}\theta_j\left(1-\sum_{j=1}^{g}\theta_j\right)+\sum_{\substack{j,k=1\\ j\neq k}}^{g}\theta_j\theta_k\right\}\times\left[\frac{n^2\left(g-\sum\limits_{j=1}^{g}\theta_j\right)^2-(1-\pi)n_1n_2}{nn_1n_2\left(g-\sum\limits_{j=1}^{g}\theta_j\right)^2}\right]\geq 0$$

Moreover, in case of having $\theta_j=\frac{1}{g}\forall\, j$, such that $\sum_{j=1}^{g}\theta_j=1$, the proposed estimator $\widehat{\pi}_P$would be more efficient than the estimator $\widehat{\pi}_I$ if

$$\left[\frac{\pi(1-\pi)n_1}{nn_2}+\frac{n^2(g-1)^2-(1-\pi)n_1n_2}{nn_1n_2g(g-1)}\right]\geq 0 \tag{7}$$

which is always true for every value of g (≥ 2) (i.e., the number of innocuous items).

3.2. Proposed versus Warner's RRT

To have an efficiency comparison, we first give a short description of Warner (1965) RRT. Warner (1965) introduced this method to decrease the biasedness in the estimators and to increase the response rate. Warner's technique consists of two complimentary questions A (Do you belong to the sensitive group?) and A^c (Do you not belong to the sensitive group?) to be answered on a probability basis. Assuming a simple random sampling with replacement (SRSWR), the ith selected respondent is asked to select a question (A or A^c) and report "yes" if his/her actual status matches with selected question, and "no" otherwise. Assuming that p is the probability of selecting question A, and π is the population proportion of individuals with sensitive group, the probability of "yes" for a particular respondent, denoted by θ, is given by:

$$P(\text{yes})=\theta=p\pi+(1-p)(1-\pi) \tag{8}$$

From (8), we have

$$\pi=\frac{\theta-(1-p)}{2p-1} \tag{9}$$

An unbiased estimator of π, by the methods of moment and maximum likelihood estimation, is given as:

$$\widehat{\pi}_W = \frac{\widehat{\theta} - (1-p)}{2p-1} \tag{10}$$

where $\widehat{\theta} = \frac{n'}{n}$ and n' is the number of "yes" responses in the sample of size n.

The variance of the estimator $\widehat{\pi}_W$ is given by:

$$Var\left(\widehat{\pi}_W\right) = \frac{\pi\left(1-\pi\right)}{n} + \frac{p\left(1-p\right)}{n\left(2p-1\right)^2} \tag{11}$$

Comparing (5) and (11) we can see that the proposed estimator $\widehat{\pi}_P$ will be more precise than $\widehat{\pi}_W$if

$$Var\left(\widehat{\pi}_W\right) - Var\left(\widehat{\pi}_P\right) \geq 0$$

$$\frac{p\left(1-p\right)}{n\left(2p-1\right)^2} - \frac{\left(1-\pi\right)}{n\left(g - \sum_{j=1}^{g}\theta_j\right)^2}\left\{\left(\sum_{j=1}^{g}\theta_j\right)\left(1-\sum_{j=1}^{g}\theta_j\right) + \sum_{\substack{j,k=1\\ j\neq k}}^{g}\theta_j\theta_k\right\} \geq 0$$

Further comparing (6) and (11) we can see that the proposed estimator $\widehat{\pi}_P$ will be more precise than $\widehat{\pi}_W$ if

$$\frac{p\left(1-p\right)}{n\left(2p-1\right)^2} - \frac{\left(1-\pi\right)}{ng\left(g-1\right)} \geq 0$$

We have calculated the Relative Efficiency (RE) of the proposed estimator $\widehat{\pi}_P$ relative to $\widehat{\pi}_I$ when it is difficult/impossible to have $\theta_j = \frac{1}{g}$, and results are provided in Tables 1–9. The RE of the proposed estimator $\widehat{\pi}_P$ relative to $\widehat{\pi}_W$ for $\theta_j \neq \frac{1}{g}$ is presented in Tables 10–12. For $\theta_j = \frac{1}{g}$ the RE of $\widehat{\pi}_P$ relative to $\widehat{\pi}_W$ is arranged in Table 13.

TABLE 1: RE of proposed estimator $\widehat{\pi}_P$ relative to $\widehat{\pi}_I$ for $n=20$, $n_1=10$, $n_2=10$.

π	$g=2$		$g=3$		$g=4$		$g=5$	
	$\sum_{j=1}^{g}\theta_j = 0.3$	$\sum_{j=1}^{g}\theta_j = 1.7$	$\sum_{j=1}^{g}\theta_j = 0.6$	$\sum_{j=1}^{g}\theta_j = 2.4$	$\sum_{j=1}^{g}\theta_j = 1$	$\sum_{j=1}^{g}\theta_j = 3$	$\sum_{j=1}^{g}\theta_j = 1.5$	$\sum_{j=1}^{g}\theta_j = 3.5$
0.1	7.0298	0.4556	12.4787	1.6290	18.6250	4.1389	24.9068	8.4680
0.2	5.7590	0.5541	9.6476	1.8271	14.0400	4.3333	18.5551	8.2768
0.3	5.2484	0.6591	8.4993	2.0463	12.1764	4.6000	15.9676	8.3472
0.4	5.0701	0.7762	8.0579	2.3046	11.4419	4.9697	14.9373	8.6756
0.5	5.1150	0.9152	8.0701	2.6325	11.4231	5.500	14.8905	9.3253
0.6	5.3896	1.0953	8.5311	3.0887	12.0984	6.3076	15.7921	10.4674
0.7	6.0181	1.3610	9.6598	3.8089	13.8000	7.6667	18.0910	12.5347
0.8	7.4450	1.8447	12.2746	5.1979	17.7721	10.400	23.4744	16.8545
0.9	11.9614	3.2084	20.6151	9.2755	30.4773	18.6250	40.7140	30.100

From the above tables 1–13 it is advocated that

1. For larger values of $\sum_{j=1}^{g} \theta_j$ the proposed estimator $\widehat{\pi}_P$ is less efficient than $\widehat{\pi}_I$ when g and $\pi$are smaller, but when g increases it becomes more efficient even for smaller values of π.
2. For smaller values of $\sum_{j=1}^{g} \theta_j$ the proposed estimator $\widehat{\pi}_P$ is more efficient than $\widehat{\pi}_I$ even when g and $\pi$are smaller.
3. n, n_1 and n_2 do not have a significant effect on the RE of the proposed estimator relative to $\widehat{\pi}_I$ except the case when n and $\sum_{j=1}^{g} \theta_j$ are larger and $g = 2$.
4. When $\sum_{j=1}^{g} \theta_j = 1$ the proposed estimator is always more efficient.
5. For smaller p the proposed estimator is less efficient than $\widehat{\pi}_W$ but as g and π are increased the RE of the proposed estimator is increased.
6. When $\sum_{j=1}^{g} \theta_j$is smaller the proposed estimator is more efficient than $\widehat{\pi}_W$ when $\pi > 0.1$ and $g > 2$.
7. Compared to $\widehat{\pi}_W$ proposed estimator $\widehat{\pi}_P$ is more efficient than $\widehat{\pi}_W$ for $g > 3$ under the given condition of $\theta_j = \frac{1}{g}$.
8. The RE of the proposed estimator $\widehat{\pi}_P$ relative to $\widehat{\pi}_W$ increases with an increase in p for a given value of g and $\pi$and it increases, for a given value of p, if g increases.

In the application scenario all the disciplines which are of sensitive nature and need extreme care in taking responses may take benefit out of the proposal, e.g., having more concern on time sensitivity (cf. Bonetti, Waeckerlin, Schuepfer & Frutiger 2000).

TABLE 2: RE of proposed estimator $\widehat{\pi}_P$ relative to $\widehat{\pi}_I$ for $n = 20$, $n_1 = 12$, $n_2 = 8$.

π	$g = 2$		$g = 3$		$g = 4$		$g = 5$	
	$\sum_{j=1}^{g} \theta_j = 0.3$	$\sum_{j=1}^{g} \theta_j = 1.7$	$\sum_{j=1}^{g} \theta_j = 0.6$	$\sum_{j=1}^{g} \theta_j = 2.4$	$\sum_{j=1}^{g} \theta_j = 1$	$\sum_{j=1}^{g} \theta_j = 3$	$\sum_{j=1}^{g} \theta_j = 1.5$	$\sum_{j=1}^{g} \theta_j = 3.5$
0.1	7.5462	0.4891	13.2303	1.7271	19.6354	4.3634	26.1792	8.9007
0.2	6.2910	0.6051	10.3474	1.9596	14.9250	4.6064	19.6285	8.7555
0.3	5.7906	0.7271	9.1825	2.2107	13.0147	4.9167	16.9640	8.8681
0.4	5.6240	0.8610	8.7410	2.5000	12.2674	5.3282	15.9087	9.2398
0.5	5.6834	1.0169	8.7665	2.8594	12.2596	5.9028	15.8716	9.9397
0.6	5.9783	1.2150	9.2843	3.3505	12.9713	6.7628	16.8191	11.1481
0.7	6.6398	1.5015	10.4363	4.1152	14.7500	8.1944	19.2198	13.3168
0.8	8.1312	2.0147	13.1650	5.5748	18.8924	11.0556	24.8323	17.8295
0.9	12.8399	3.4441	21.8568	9.8341	32.1307	19.6354	42.7940	31.6386

Table 3: *RE* of proposed estimator $\widehat{\pi}_P$ relative to $\widehat{\pi}_I$ for $n = 20$, $n_1 = 8$, $n_2 = 12$.

π	$g = 2$		$g = 3$		$g = 4$		$g = 5$	
	$\sum_{j=1}^{g} \theta_j = 0.3$	$\sum_{j=1}^{g} \theta_j = 1.7$	$\sum_{j=1}^{g} \theta_j = 0.6$	$\sum_{j=1}^{g} \theta_j = 2.4$	$\sum_{j=1}^{g} \theta_j = 1$	$\sum_{j=1}^{g} \theta_j = 3$	$\sum_{j=1}^{g} \theta_j = 1.5$	$\sum_{j=1}^{g} \theta_j = 3.5$
0.1	7.0994	0.4601	12.7671	1.6667	19.1667	4.2592	25.7098	8.7411
0.2	5.7082	0.5492	9.7519	1.8468	14.3250	4.4213	19.0280	8.4877
0.3	5.1438	0.6459	8.5244	2.0523	12.3530	4.6667	16.3018	8.5219
0.4	4.9388	0.7561	8.0463	2.3013	11.5698	5.0252	15.2107	8.8344
0.5	4.9730	0.8898	8.0479	2.6250	11.5348	5.5556	15.1502	9.4879
0.6	5.2500	1.0670	8.5189	3.0843	12.2336	6.3782	16.0812	10.6589
0.7	5.8981	1.3340	9.6888	3.8202	14.0000	7.7778	18.4696	12.7970
0.8	7.3791	1.8284	12.4072	5.2540	18.1329	10.6111	24.0726	17.2841
0.9	12.0796	3.2401	21.0914	9.4897	31.3636	19.1667	42.0268	31.0714

Table 4: *RE* of proposed estimator $\widehat{\pi}_P$ relative to $\widehat{\pi}_I$ for $n = 50$, $n_1 = 25$, $n_2 = 25$.

π	$g = 2$		$g = 3$		$g = 4$		$g = 5$	
	$\sum_{j=1}^{g} \theta_j = 0.3$	$\sum_{j=1}^{g} \theta_j = 1.7$	$\sum\limits_{j=1}^{g} \theta_j =$ 0.6	$\sum_{j=1}^{g} \theta_j = 2.4$	$\sum_{j=1}^{g} \theta_j = 1$	$\sum_{j=1}^{g} \theta_j = 3$	$\sum_{j=1}^{g} \theta_j = 1.5$	$\sum_{j=1}^{g} \theta_j = 3.5$
0.1	7.0299	0.4556	12.4788	1.6290	18.6250	4.1389	24.9067	8.4680
0.2	5.7590	0.5541	9.6477	1.8271	14.0400	4.3333	18.5551	8.2768
0.3	5.2484	0.6591	8.4993	2.0463	12.1765	4.6000	15.9675	8.3472
0.4	5.0701	0.7762	8.0579	2.3046	11.4419	4.9697	14.9373	8.6756
0.5	5.1150	0.9152	8.0709	2.6325	11.4231	5.5000	14.8904	9.3253
0.6	5.3896	1.0953	8.5311	3.0887	12.0984	6.3076	15.7912	10.4674
0.7	6.0182	1.3610	9.6598	3.8089	13.8000	7.6667	18.0910	12.5347
0.8	7.4450	1.8447	12.2747	5.1979	17.7722	10.40	23.4744	16.8545
0.9	11.9614	3.2084	20.6152	9.2755	30.4773	18.6250	40.7140	30.1008

Table 5: *RE* of proposed estimator $\widehat{\pi}_P$ relative to $\widehat{\pi}_I$ for $n = 50$, $n_1 = 30$, $n_2 = 20$.

π	$g = 2$		$g = 3$		$g = 4$		$g = 5$	
	$\sum_{j=1}^{g} \theta_j = 0.3$	$\sum_{j=1}^{g} \theta_j = 1.7$	$\sum_{j=1}^{g} \theta_j = 0.6$	$\sum_{j=1}^{g} \theta_j = 2.4$	$\sum_{j=1}^{g} \theta_j = 1$	$\sum_{j=1}^{g} \theta_j = 3$	$\sum_{j=1}^{g} \theta_j = 1.5$	$\sum_{j=1}^{g} \theta_j = 3.5$
0.1	7.5462	0.4891	13.2303	1.7271	19.6354	4.3634	26.1792	8.9007
0.2	6.2910	0.6051	10.3474	1.9596	14.9250	4.6064	19.6285	8.7555
0.3	5.7906	0.7271	9.1825	2.2107	13.0147	4.9167	16.9640	8.8681
0.4	5.6240	0.8610	8.7410	2.5000	12.2674	5.3282	15.9087	9.2398
0.5	5.6834	1.0169	8.7665	2.8594	12.2596	5.9028	15.8716	9.9397
0.6	5.9783	1.2150	9.2843	3.3505	12.9713	6.7628	16.8191	11.1481
0.7	6.6398	1.5015	10.4363	4.1152	14.7500	8.1944	19.2198	13.3168
0.8	8.1312	2.0147	13.1650	5.5748	18.8924	11.0556	24.8323	17.8295
0.9	12.8399	3.4441	21.8568	9.8341	32.1307	19.6354	42.7940	31.6386

Table 6: *RE* of proposed estimator $\widehat{\pi}_P$ relative to $\widehat{\pi}_I$ for $n = 50$, $n_1 = 20$, $n_2 = 30$.

π	$g = 2$		$g = 3$		$g = 4$		$g = 5$	
	$\sum_{j=1}^{g} \theta_j = 0.3$	$\sum_{j=1}^{g} \theta_j = 1.7$	$\sum_{j=1}^{g} \theta_j = 0.6$	$\sum_{j=1}^{g} \theta_j = 2.4$	$\sum_{j=1}^{g} \theta_j = 1$	$\sum_{j=1}^{g} \theta_j = 3$	$\sum_{j=1}^{g} \theta_j = 1.5$	$\sum_{j=1}^{g} \theta_j = 3.5$
0.1	7.0994	0.4601	12.7671	1.6667	19.1667	4.2592	25.7098	8.7411
0.2	5.7082	0.5492	9.7519	1.8468	14.3250	4.4213	19.0280	8.4877
0.3	5.1438	0.6459	8.5244	2.0523	12.3530	4.6667	16.3018	8.5219
0.4	4.9388	0.7561	8.0463	2.3013	11.5698	5.0252	15.2107	8.8344
0.5	4.9730	0.8898	8.0479	2.6250	11.5348	5.5556	15.1502	9.4879
0.6	5.2500	1.0670	8.5189	3.0843	12.2336	6.3782	16.0812	10.6589
0.7	5.8981	1.3340	9.6888	3.8202	14.0000	7.7778	18.4696	12.7970
0.8	7.3791	1.8284	12.4072	5.2540	18.1329	10.6111	24.0726	17.2841
0.9	12.0796	3.2401	21.0914	9.4897	31.3636	19.1667	42.0268	31.0714

Table 7: RE of proposed estimator $\widehat{\pi}_P$ relative to $\widehat{\pi}_I$ for $n = 100$, $n_1 = 50$, $n_2 = 50$.

π	$g=2$		$g=3$		$g=4$		$g=5$	
	$\sum_{j=1}^{g}\theta_j = 0.3$	$\sum_{j=1}^{g}\theta_j = 1.7$	$\sum_{j=1}^{g}\theta_j = 0.6$	$\sum_{j=1}^{g}\theta_j = 2.4$	$\sum_{j=1}^{g}\theta_j = 1$	$\sum_{j=1}^{g}\theta_j = 3$	$\sum_{j=1}^{g}\theta_j = 1.5$	$\sum_{j=1}^{g}\theta_j = 3.5$
0.1	7.0298	0.4556	12.4787	1.6290	18.6250	4.1389	24.9068	8.4680
0.2	5.7590	0.5441	9.6476	1.8271	14.0400	4.3333	18.5555	8.2768
0.3	5.2484	0.6591	8.4993	2.0463	12.1764	4.6000	15.9675	8.3472
0.4	5.0701	0.7762	8.0579	2.3046	11.4419	4.9697	14.9373	8.6756
0.5	5.1150	0.9152	8.0701	2.6325	11.4231	5.5000	14.8904	9.3253
0.6	5.3896	1.0954	8.5311	3.0887	12.0936	6.3076	15.7922	10.4674
0.7	6.0181	1.3610	9.6598	3.8089	13.8000	7.6667	18.0910	12.5347
0.8	7.4450	1.8447	12.2746	5.1979	17.7721	10.40	23.4744	16.8545
0.9	11.9614	3.2084	20.6151	9.2755	30.4773	18.6250	40.7140	30.1008

Table 8: RE of proposed estimator $\widehat{\pi}_P$ relative to $\widehat{\pi}_I$ for $n = 100$, $n_1 = 80$, $n_2 = 20$.

π	$g=2$		$g=3$		$g=4$		$g=5$	
	$\sum_{j=1}^{g}\theta_j = 0.3$	$\sum_{j=1}^{g}\theta_j = 1.7$	$\sum_{j=1}^{g}\theta_j = 0.6$	$\sum_{j=1}^{g}\theta_j = 2.4$	$\sum_{j=1}^{g}\theta_j = 1$	$\sum_{j=1}^{g}\theta_j = 3$	$\sum_{j=1}^{g}\theta_j = 1.5$	$\sum_{j=1}^{g}\theta_j = 3.5$
0.1	11.9895	0.7770	20.5405	2.6814	30.1562	6.7013	39.9729	13.5904
0.2	10.3074	0.9912	16.4144	3.1085	23.2875	7.1875	30.3435	13.5351
0.3	9.6561	1.2126	14.7610	3.5538	20.5147	7.7500	26.4391	13.8214
0.4	9.4637	1.4488	14.1534	4.0480	19.4477	8.4470	24.9101	14.4679
0.5	9.5908	1.7161	14.2275	4.6406	19.4711	9.3750	24.8896	15.5873
0.6	10.0600	2.0446	14.9845	5.4253	20.5635	10.7211	26.3356	17.4558
0.7	11.0721	2.5039	16.7765	6.6152	23.2500	12.9167	29.9551	20.7549
0.8	13.3247	3.3016	20.8840	8.8436	29.4778	17.2500	38.3881	27.5625
0.9	20.4003	5.4722	33.9334	15.2679	49.3466	30.1562	65.3419	48.3088

Table 9: RE of proposed estimator $\widehat{\pi}_P$ relative to $\widehat{\pi}_I$ for $n = 100$, $n_1 = 20$, $n_2 = 80$.

π	$g=2$		$g=3$		$g=4$		$g=5$	
	$\sum_{j=1}^{g}\theta_j = 0.3$	$\sum_{j=1}^{g}\theta_j = 1.7$	$\sum_{j=1}^{g}\theta_j = 0.6$	$\sum_{j=1}^{g}\theta_j = 2.4$	$\sum_{j=1}^{g}\theta_j = 1$	$\sum_{j=1}^{g}\theta_j = 3$	$\sum_{j=1}^{g}\theta_j = 1.5$	$\sum_{j=1}^{g}\theta_j = 3.5$
0.1	9.9789	0.6467	18.4556	2.4092	28.04688	6.2326	37.8608	12.8723
0.2	7.6896	0.7398	13.7345	2.6010	20.5875	6.3542	27.6413	12.3289
0.3	6.7454	0.8471	11.7994	2.8407	17.5368	6.6250	23.4595	12.2637
0.4	6.3805	0.9768	11.0275	3.1539	16.3081	7.0833	21.7691	12.6435
0.5	6.3938	1.1441	10.9940	3.5859	16.2260	7.8125	21.6431	13.5542
0.6	6.7825	1.3784	11.6752	4.2271	17.2438	8.9904	23.0149	15.2548
0.7	7.7353	1.7492	13.4105	5.2880	19.8750	11.0417	26.5792	18.4158
0.8	9.9407	2.4631	17.4743	7.3997	26.0601	15.2500	34.9695	25.1079
0.9	16.9791	4.5544	30.4890	13.7181	45.8949	28.0469	61.8893	45.7563

TABLE 10: RE of proposed estimator $\widehat{\pi}_P$ relative to $\widehat{\pi}_W$ for $n = 20$ and larger $\sum_{j=1}^{g} \theta_j$.

		π								
p	g, $\sum_{j=1}^{g} \theta_j$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	2, 1.7	0.09	0.13	0.16	0.20	0.24	0.28	0.34	0.42	0.63
	3, 2.4	0.19	0.25	0.32	0.38	0.44	0.51	0.59	0.72	1.06
	4, 3	0.32	0.42	0.50	0.58	0.65	0.73	0.83	1.00	1.44
	5, 3.5	0.49	0.60	0.69	0.77	0.85	0.93	1.04	1.23	1.74
0.2	2, 1.7	0.21	0.25	0.31	0.36	0.42	0.51	0.63	0.84	1.45
	3, 2.4	0.43	0.51	0.59	0.68	0.78	0.91	1.10	1.45	2.45
	4, 3	0.74	0.84	0.93	1.04	1.16	1.32	1.56	2.01	3.34
	5, 3.5	1.14	1.21	1.29	1.39	1.51	1.67	1.94	2.47	4.04
0.3	2, 1.7	0.54	0.62	0.71	0.81	0.95	1.15	1.46	2.06	3.81
	3, 2.4	1.13	1.25	1.38	1.54	1.76	2.07	2.57	3.54	6.44
	4, 3	1.95	2.05	2.18	2.35	2.60	2.99	3.62	4.91	8.77
	5, 3.5	2.98	2.96	3.01	3.15	3.39	3.80	4.52	6.02	10.61
0.4	2, 1.7	2.35	2.59	2.88	3.27	3.81	4.62	5.95	8.61	16.56
	3, 2.4	4.91	5.21	5.62	6.20	7.03	8.31	10.47	14.82	27.96
	4, 3	8.45	8.56	8.87	9.45	10.42	12.00	14.79	20.53	38.06
	5, 3.5	12.96	12.38	12.28	12.65	13.55	15.26	18.46	25.20	46.06

TABLE 11: RE of proposed estimator $\widehat{\pi}_P$ relative to $\widehat{\pi}_W$ $n = 50$ and larger $\sum_{j=1}^{g} \theta_j$.

		π								
p	g, $\sum_{j=1}^{g} \theta_j$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	2, 1.7	0.09	0.13	0.16	0.20	0.24	0.28	0.34	0.42	0.63
	3, 2.4	0.19	0.25	0.32	0.38	0.44	0.51	0.59	0.72	1.06
	4, 3	0.32	0.42	0.50	0.58	0.65	0.73	0.83	1.00	1.44
	5, 3.5	0.49	0.60	0.69	0.77	0.85	0.93	1.04	1.23	1.74
0.2	2, 1.7	0.21	0.25	0.31	0.36	0.42	0.51	0.63	0.84	1.45
	3, 2.4	0.43	0.51	0.59	0.68	0.78	0.91	1.10	1.45	2.45
	4, 3	0.74	0.84	0.93	1.04	1.16	1.32	1.56	2.01	3.34
	5, 3.5	1.14	1.21	1.29	1.39	1.51	1.67	1.94	2.47	4.04
0.3	2, 1.7	0.54	0.62	0.71	0.81	0.95	1.15	1.46	2.06	3.81
	3, 2.4	1.13	1.25	1.38	1.54	1.76	2.07	2.57	3.54	6.44
	4, 3	1.95	2.05	2.18	2.35	2.60	2.99	3.62	4.91	8.77
	5, 3.5	2.98	2.96	3.01	3.15	3.39	3.80	4.52	6.02	10.61
0.4	2, 1.7	2.35	2.59	2.88	3.27	3.81	4.62	5.95	8.61	16.56
	3, 2.4	4.91	5.21	5.62	6.20	7.03	8.31	10.47	14.82	27.96
	4, 3	8.45	8.56	8.87	9.45	10.42	12.00	14.79	20.53	38.06
	5, 3.5	12.96	12.38	12.28	12.65	13.55	15.26	18.46	25.20	46.06

4. Concluding Remarks

An alternative item count technique has been presented in this article. One of the main features of this technique is that it does not require the selection of two subsamples of sizes n_1 and n_2. Therefore, we do not need to worry about the optimum values of n_1 and n_2 (as is the case with usual ICT estimator $\widehat{\pi}_I$). Furthermore, the response from a respondent is bounded to lie between 0 and g, which helps to provide the privacy to the respondent because the response can not be traced back to respondent's actual status about the possession of sensitive item (provided that the actual status of a particular respondent about at least one unrelated characteristic is unknown to the interviewer or anonymity is provided to respondents). To avoid this situation, we recommend conducting the survey in the absence of the interviewer or the whole process must be administered unseen to the interviewer.

TABLE 12: *RE* of proposed estimator $\widehat{\pi}_P$ relative to $\widehat{\pi}_W$ for $n = 20$ and smaller $\sum_{j=1}^{g} \theta_j$.

		π								
p	g, $\sum_{j=1}^{g} \theta_j$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	2, 0.5	0.97	1.03	1.08	1.12	1.18	1.24	1.35	1.56	2.17
	3, 0.6	1.42	1.34	1.32	1.32	1.34	1.39	1.49	1.71	2.35
	4, 1	1.44	1.35	1.33	1.33	1.35	1.40	1.50	1.71	2.36
	5, 1.5	1.44	1.35	1.33	1.33	1.35	1.40	1.50	1.71	2.36
0.2	2, 0.5	2.25	2.07	2.01	2.02	2.09	2.24	2.52	3.13	5.02
	3, 0.6	3.30	2.70	2.46	2.37	2.39	2.51	2.79	3.43	5.45
	4, 1	3.34	2.72	2.47	2.38	2.40	2.52	2.80	3.44	5.47
	5, 1.5	3.35	2.72	2.48	2.39	2.40	2.52	2.81	3.44	5.47
0.3	2, 0.5	5.90	5.05	4.68	4.58	4.70	5.08	5.87	7.63	13.17
	3, 0.6	8.67	6.58	5.72	5.39	5.38	5.71	6.51	8.37	14.33
	4, 1	8.77	6.63	5.76	5.41	5.406	5.73	6.52	8.40	14.34
	5, 1.5	8.78	6.32	5.76	5.42	5.41	5.73	6.53	8.39	14.35
0.4	2, 0.5	25.59	21.13	19.10	18.42	18.81	20.40	23.95	31.93	57.21
	3, 0.6	37.62	27.51	23.35	21.67	21.56	22.94	26.54	35.01	62.15
	4, 1	38.06	27.72	23.48	21.78	21.63	23.01	26.61	35.09	62.28
	5, 1.5	38.11	27.74	23.50	21.78	21.64	23.02	26.62	35.10	62.30

TABLE 13: Relative efficiency of the proposed estimator $\widehat{\pi}_P$ relative to $\widehat{\pi}_W$ for $0.1 \leq \pi \leq 0.9$ and $0.1 \leq p \leq 0.4$.

π \ p	0.1	0.2	0.3	0.4	π \ p	0.1	0.2	0.3	0.4
	$g = 4$					$g = 5$			
0.1	1.397	3.239	8.500	36.909	0.1	1.708	3.958	10.388	45.111
0.3	1.306	2.438	5.673	23.142	0.3	1.4311	2.671	6.214	25.346
0.5	1.339	2.380	5.357	21.428	0.5	1.420	2.525	5.681	22.727
0.7	1.492	2.784	6.478	26.425	0.7	1.558	2.908	6.766	27.600
0.9	2.345	5.435	14.262	61.932	0.9	2.427	5.625	14.763	64.105
	$g = 6$					$g = 7$			
0.1	1.921	4.453	11.687	50.750	0.1	2.069	4.796	12.586	54.653
0.3	1.502	2.804	6.525	26.614	0.3	1.546	2.887	6.716	27.397
0.5	1.464	2.604	5.859	23.437	0.5	1.491	2.651	5.965	23.863
0.7	1.593	2.974	6.920	28.227	0.7	1.614	3.013	7.011	28.598
0.9	2.470	5.726	15.026	65.250	0.9	2.496	5.785	15.181	65.922

It has been observed that the proposed item count technique estimator performs better than the usual item count technique under the conditions that $\theta_j = \frac{1}{g}$ and $\sum_{j=1}^{g} \theta_j = 1$. It may be difficult to select the items in such a way that their proportions in the population are the same and sum to one, but this would be the case if the number of items is large. Thus, in practice, one or two innocuous items with same proportions can be found and included in the item list (e.g., item 1: Were you born in the months from January to June?, and Item 2: Is your gender male?) If the condition to satisfy the inequality (7) is hard to meet we would suggest to look for a large number of innocuous items (4, 5, 6, etc.) such that their prevalence in the population is rare and consequently we have smaller $\sum_{j=1}^{g} \theta_j$, so that inequality (7) is easily satisfied.

In brief, based on the findings of the Section 4 and the concluding discussion above we recommend the use of the proposed ICT in surveys about sensitive items instead of the usual ICT and the Warner's RRT. Preferably, the data collecting phase must be administered unseen to the surveyor.

Acknowledgements

The authors are deeply thankful to the editor and the two learned referees for guiding towards the improvement of the earlier draft of this article.

[Recibido: enero de 2011 — Aceptado: septiembre de 2011]

References

Ahart, A. M. & Sackett, P. R. (2004), 'A new method for examining relationships between individual difference measures and sensitive behavior criteria: Evaluating the unmatched count technique', *Organizational Research Methods* **7**(1), 101–114.

Biemer, P. P. & Wright, D. (2004), Estimating cocaine use using the item count methodology, *in* 'Preliminary results form the national Survey on drug use and Health', Annual meeting of the American Association for Public Opinion research, Phoenix, Arizona.

Bjorner, J. B., Kosinski, M. & Ware, J. E. (2003), 'Using item response theory to calibrate the headache impact test (hit^{TM}) to the metric of traditional headache scales', *Quality of Life Research* **12**, 981–1002.

Bonetti, P. O., Waeckerlin, A., Schuepfer, G. & Frutiger, A. (2000), 'Improving time-sensitive processes in the intensive care unit: The example of 'door-to-needle time' in acute myocardial infarction', *International Journal for Quality in Health Care* **12**(4), 311–317.

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**(2), 589–593.

Chaudhuri, A. & Mukherjee, R. (1988), *Randomized Response: Theory and Methods*, Marcel-Decker, New York.

Dalton, D. R. & Metzger, M. (1992), 'Integrity testing for personal selection: An unsparing perspective', *Journal of Business Ethics* **12**, 147–156.

Dalton, D. R., Wimbush, J. C. & Daily, C. M. (1994), 'Using the unmatched count technique (uct) to estimate the base rates for sensitive behavior', *Personnel Psychology* **47**, 817–828.

Droitcour, J. A., Caspar, R. A., Hubbard, M. L., Parsley, T. L., Visscher, W. & Ezzati, T. M. (1991), 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. Lyberg, N. Mathiowetz & S. Sudman, eds, 'Measurement Errors in Serveys', Wiley, New York.

Droitcour, J. A. & Larson, E. M. (2002), 'An innovative technique for asking sensitive questions: The three card method', *Sociological Methodological Bulletin* **75**, 5–23.

Droitcour, J. A., Larson, E. M. & Scheuren, F. J. (2001), The three card method: Estimating sensitive survey items with permanent anonymity of response, *in* 'Proceedings of the Social Statistics Section', American Statistical Association, Alexandria, Virginia.

Geurts, M. D. (1980), 'Using a randomized response design to eliminate nonresponse and response biases in business research', *Journal of the Academy of Marketing Science* **8**, 83–91.

Hubbard, M. L., Casper, R. A. & Lessler, J. T. (1989), Respondents' reactions to item count lists and randomized response, *in* 'Proceedings of the Survey Research Section', American Statistical Association, Washington, D. C., pp. 544–448.

Larkins, E. R., Hume, E. C. & Garcha, B. S. (1997), 'Validity of randomized response method in tax ethics research', *Journal of the Applied Business Research* **13**, 25–32.

Liu, P. T. & Chow, L. P. (1976), 'A new discrete quantitative randomized response model', *Journal of the American Statistical Association* **71**, 72–73.

Martin, M., Kosinski, M., Bjorner, J. B., Ware, J. E. & MacLean, R. (2007), 'Item response theory methods can improve the measurement of physical function by combining the modified health assessment questionnaire and the SF-36 physical function scale', *Quality of Life Research* **16**, 647–660.

Miller, J. D. (1985), 'The nominative technique: A new method of estimating heroin prevalence', *NIDA Research Monograph* **54**, 104–124.

Reinmuth, J. E. & Geurts, M. D. (1975), 'The collection of sensitive information using a two stage randomized response model', *Journal of Marketing Research* **12**, 402–407.

Tracy, D. & Mangat, N. (1996), 'Some development in randomized response sampling during the last decade-a follow up of review by Chaudhuri and Mukerjee', *Journal of Applied Statistical Science* **4**, 533–544.

Tsuchiya, T. (2005), 'Domain estimators for the item count technique', *Survey Methodology* **31**, 41–51.

Tsuchiya, T., Hirai, Y. & Ono, S. (2007), 'A study of the properties of the item count technique', *Public Opinion Quarterly* **71**, 253–272.

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

Wimbush, J. C. & Dalton, D. R. (1997), 'Base rate for employee theft: Convergence of multiple methods', *Journal of Applied Psychology* **82**, 756–763.

Appendix

To find the variance of the estimator $\widehat{\pi}_P$, consider

$$Z_i^2=\sum_{j=1}^{g}\alpha_j^2+\sum_{\substack{j,k=1\\ j\neq k}}^{g}\alpha_j\alpha_k \tag{12}$$

After applying expectation operator on (12), we get:

$$\begin{aligned}
E\left(Z_i^2\right)&=\sum_{j=1}^{g}E\left(\alpha_j^2\right)+\sum_{\substack{j,k=1\\ j\neq k}}^{g}E\left(\alpha_j\alpha_k\right)\\
&=\sum_{j=1}^{g}\left(\pi+\theta_j-\pi\theta_j\right)+\sum_{\substack{j,k=1\\ j\neq k}}^{g}\left\{\pi+\theta_j\theta_k\left(1-\pi\right)\right\}\\
&=\left(1-\pi\right)\sum_{j=1}^{g}\theta_j+g\pi+g\left(g-1\right)\pi+\left(1-\pi\right)\sum_{\substack{j,k=1\\ j\neq k}}^{g}\theta_j\theta_k
\end{aligned} \tag{13}$$

Now by definition of the variance of Z_i we have:

$$V\left(Z_i\right)=E\left(Z_i^2\right)-\left(E\left(Z_i\right)\right)^2 \tag{14}$$

Substituting (13) and $E\left(Z_i\right)=\left(g-\sum_{j=1}^{g}\theta_j\right)\pi+\sum_{j=1}^{g}\theta_j$ in (14), we get

$$\begin{aligned}
V(Z_it)&=(1-\pi)\sum_{j=1}^{g}\theta_j+g\pi+g(g-1)\pi+(1-\pi)\sum_{\substack{j,k=1\\ j\neq k}}^{g}\theta_j\theta_k\\
&-\left\{\left(g-\sum_{j=1}^{g}\theta_j\right)\pi+\sum_{j=1}^{g}\theta_j\right\}^2\\
&=(1-\pi)\left[\left(g-\sum_{j=1}^{g}\theta_j\right)^2\pi+\sum_{j=1}^{g}\theta_j\left(1-\sum_{j=1}^{g}\theta_j\right)+\sum_{\substack{j,k=1\\ j\neq k}}^{g}\theta_j\theta_k\right]
\end{aligned} \tag{15}$$

Now from (4) we have

$$V\left(\widehat{\pi}_P\right)=\frac{n^{-1}V\left(Z_i\right)}{\left(g-\sum_{j=1}^{g}\theta_j\right)^2} \tag{16}$$

Finally, using (15) in (16), we get the result in (5).