ELSEVIER

A note on the comparison of some randomized response procedures

Massimo Guerriero[a], Marco F. Sandri[b],*

[a]*E.S.I. Department, University of Verona, Via dell'Artigliere 19, 37129 Verona, Italy*
[b]*Department of Quantitative Methods, University of Brescia, Contrada Santa Chiara 50, 25122 Brescia, Italy*

Abstract

The paper reconsiders and compares some randomized response models in the light of the efficiency/protection criterion. The results show that, despite the assertion of the authors about the superiority of their methods, some procedures are perfectly equivalent to the Warner's or the Simmons' models. Our work extends and generalizes the results of Bhargava and Singh [2002. On the efficiency comparison of certain randomized response strategies. Metrika 55, 191–197] and other authors.
© 2006 Elsevier B.V. All rights reserved.

Keywords: Randomized response; Privacy protection; Efficiency

1. Introduction

Many randomized response (RR) procedures have been proposed in the past decades as a valuable way for conducting surveys on sensitive topics and often these methods have been presented without an adequate comparison of their performances. As correctly pointed out by many authors (Leysieffer and Warner, 1976; Lanke, 1976; etc.), for a valid comparison, both efficiency and respondents' protection should be taken into account. In other words, one must compare the variances of the estimates of the different methods under the condition of equal levels of protection of the individuals' privacy.

The aim of this note is to present a comparison of some RR methods based on the efficiency/protection criterion, showing that the results are sometimes in contrast with the authors' assertions regarding the superiority of their methods.

Our work extends the results of Bhargava and Singh (2002), Nayak (1994) and Zou (1997) in two directions: (1) we analyze and compare a larger set of models and (2) we prove a stronger form of equivalence between these models and the Warner's or Simmons' models.

2. Equivalence between RR models

We are interested in knowing the proportion π of individuals in a population belonging to the class A, which is thought to be socially unacceptable. In the following, we only consider RR models with two possible answers, 'Yes'(Y) and 'No'(N). Moreover, we assume that all the responses of the interviewed subjects are truthful.

A RR model is a procedure for conducting surveys on sensitive topics and to produce trustworthy data for estimating the proportion π. A RR model is provided with m ($m \geqslant 1$) randomization devices (RDs) and a set of rules for determining and communicating the answer: for each RD the respondent randomly selects one of the s_k ($s_k \geqslant 1$, $k = 1, \ldots, m$) statements and, following the rules, reports 'Yes' or 'No' without revealing which questions he/she is answering. The kth RD of the RR model M is described by a vector of $s_k - 1$ parameters (probabilities) $\theta_k = (p_{k1}, \ldots, p_{k(s_k-1)})$, where $p_{ki} \in S_{ki} \subseteq [0, 1]$, S_{ki} is the set of admissible values for p_{ki} and $\theta_k \in \Theta_k = (S_{k1} \times \cdots \times S_{k(s_k-1)})$. Therefore, the RR model M can be characterized by the vector $\theta = (\theta_1, \ldots, \theta_m)$, where $\theta \in \Theta = (\Theta_1 \times \cdots \times \Theta_m)$.

We write $M(\theta)$ to indicate a RR model with a specified set θ of parameters (e.g. a Warner's model with probability $p = 0.7$) and by $\mathscr{M}(\Theta)$ we indicate the associated family of RR models (e.g. the family of Warner's models, with $\Theta = S_1 = \{p \in [0, 1], p \neq 1/2\}$).

Let $P(R|A)$, $R \in \{$'Yes','No'$\}$ be the conditional probabilities that a response R comes from an individual of group A. These probabilities are called *design probabilities* and can be controlled by setting the elements of the vector θ.

There are a multitude of measures of privacy protection (henceforth PP-measures) proposed in the statistical literature (for a brief but almost exhaustive summary, see Ljungqvist, 1993). Most of them can be expressed as a function of the design probabilities. For example, the measure proposed by Lanke (1976) is based on the idea that, given a certain answer, the larger the conditional probability of being classified in the sensitive group A, the greater is the embarrassment caused by giving that answer. Therefore, according to this measure (henceforth, the L-measure of protection), a RR model is considered more protective than another if

$$f(A) = \max\{P(A|Y), P(A|N)\}$$

is smaller for the former method, where $P(A|R)$ is the conditional probability of A given the answer R. The design probabilities can be used to compute this probability:

$$P(A|R) = \frac{\pi P(R|A)}{\pi P(R|A) + (1-\pi)P(R|A^c)},$$

where A^c means 'Do not belong to the sensitive group A'.

Another measure of protection is the level of jeopardy proposed by Leysieffer and Warner (1976) (henceforth, the LW-measure of protection) given by

$$g(Y, A) = \frac{P(Y|A)}{P(Y|A^c)}, \quad g(N, A^c) = \frac{P(N|A^c)}{P(N|A)} = \frac{1 - P(Y|A^c)}{1 - P(Y|A)}.$$

The conditions $g(Y, A) > 1$ and $g(N, A^c) > 1$ require that the 'Yes' and the 'No' answers are jeopardizing with respect to A and A^c, respectively.

Let $PP(M(\theta))$ be the PP-measure of model $M(\theta)$, $\hat{\pi}(\theta)$ the estimator of π according to model $M(\theta)$ and $V(\hat{\pi}(\theta))$ its variance. We give now some definitions of equivalence between two RR models and two families of RR models. We start considering the equivalence between two RR models:

Definition 1. Two RR models $M_A(\theta_1) \in \mathscr{M}_A$ and $M_B(\theta_2) \in \mathscr{M}_B$ are said to be PP-equivalent according to the efficiency/protection criterion and for a given measure PP of privacy protection if the variances of their estimator and their levels of respondents' privacy protection are equal, that is if $V(\hat{\pi}_A(\theta_1)) = V(\hat{\pi}_B(\theta_2))$ and $PP(M_A(\theta_1)) = PP(M_B(\theta_2))$.

The definition of (weak) *PP-equivalence* between two families of RR models follows:

Definition 2. Two families of RR models $\mathscr{M}_A$ and $\mathscr{M}_B$ are said to be PP-equivalent according to the efficiency/ protection criterion and for a given measure of protection PP if for every model $M_A(\theta_1) \in \mathscr{M}_A$ there is at least one PP-equivalent model $M_B(\theta_2) \in \mathscr{M}_B$, and vice versa.

Finally, we define (strong) $\mathscr{PP}$-equivalence between two families of RR models:

Definition 3. Two families of RR models $\mathscr{M}_A$ and $\mathscr{M}_B$ are said to be $\mathscr{PP}$-equivalent according to the efficiency/ protection criterion and for a given set $\mathscr{PP}$ of privacy protection measures if for every model $M_A(\theta_1) \in \mathscr{M}_A$ and for every measure $PP \in \mathscr{PP}$ there is at least one PP-equivalent model $M_B(\theta_2) \in \mathscr{M}_B$, and vice versa.

From now on, we restrict our attention to the set of PP-measures which can be expressed as functions of the design probabilities $P(R|A)$ only. This set $\mathscr{PP}$ contains most of the PP-measures proposed in the literature. In our work, $\mathscr{PP}$-equivalences between RR models have been proved for this set of measures. The utilitarian approach to respondent's privacy proposed by Ljungqvist (1993) has not been considered in our work.

3. Reference models

In our comparisons we choose two well known RR models as benchmarks: the symmetric procedure of Warner (1965) and the unrelated question (Simmons') procedure of Greenberg et al. (1969).

The Warner's model requires the respondent to give a 'Yes' or 'No' answer to either the statements 'I am a member of the sensitive group A' or to its converse 'I am not a member of the sensitive group A' depending on the outcome of the RD (with probability p^W and $1 - p^W$, respectively). Privacy is ensured since the interviewer does not know which question any respondent has answered. The only parameter of this model is $\theta^W = p^W$ and the design probabilities are $P(Y|A) = p^W$ and $P(Y|A^c) = 1 - p^W$. The L-measure and the LW-measure of protection are, respectively, given by

$$f_W(A) = \begin{cases} \pi(1 - p^W)/(1 - \lambda_W) & \text{if } p^W < 1/2, \\ \pi p^W/\lambda_W & \text{if } p^W > 1/2, \end{cases}$$

$$g_W(Y, A) = g_W(N, A^c) = \frac{p^W}{1 - p^W},$$

where $\lambda_W = \pi p^W + (1 - \pi)(1 - p^W)$. Moreover, the efficiency of the Warner's estimator $\hat{\pi}_W = [\hat{\lambda} - (1 - p^W)]/(2p^W - 1)$ is $V_W = \lambda_W(1 - \lambda_W)/[n(2p^W - 1)^2]$, where $\hat{\lambda} = n'/n$ is the observed proportion of 'Yes' answers in the sample, n is the number of respondents and n' the number of 'Yes' answers.

The Simmons' model requires the respondent to answer the question 'I am a member of the sensitive group A' with probability p_1^S and 'I am a member of the group Y' with probability $1 - p_1^S$, where the two questions are unrelated and the second one is a completely harmless question. The parameters of this model are $\theta^S = (p_1^S, p_2^S)$, where p_2^S is the relative frequency of Y-individuals in the population. The design probabilities are $P(Y|A) = p_2^S(1 - p_1^S) + p_1^S$ and $P(Y|A^c) = p_2^S(1 - p_1^S)$. The L-measure and the LW-measure of protection are, respectively, given by

$$f_S(A) = \pi[p_2^S(1 - p_1^S) + p_1^S]/\lambda_S, \quad g_S(Y, A) = \frac{p_2^S(1 - p_1^S) + p_1^S}{p_2^S(1 - p_1^S)},$$

where $\lambda_S = \pi p_1^S + (1 - p_1^S)p_2^S$. Moreover, the efficiency of the Simmons' estimator $\hat{\pi}_S = [\hat{\lambda} - (1 - p_1^S)p_2^S]/p_1^S$ is $V_S = \lambda_S(1 - \lambda_S)/[n(p_1^S)^2]$.

The works of Lanke (1976) and Leysieffer and Warner (1976) compare these models under the condition of equal L- and LW-protection, respectively, and find that the Simmons' procedure is more efficient than the Warner's model for $p_2^S > 1/2$, for all $\pi \in (0, 1)$. Equivalence between the two models with respect to every PP-measure $\in \mathscr{PP}$ can be found only for $p_2^S = 1/2$ and $p^W = (1 + p_1^S)/2$, for all $\pi \in (0, 1)$.

4. Efficiency/protection model comparison

4.1. The Mangat and Singh (1990) model

Mangat and Singh (1990) proposed a two-stage RR procedure where each of the n respondents is provided with two random devices: the first RD consists of two statements with probabilities p_1^{MS} and $1 - p_1^{MS}$, namely (a) 'I am a member of the sensitive group A' and (b) 'Go to the second random device'. The second RD is the same used by Warner (1965) with probability p_2^{MS}.

By simple algebra, one can prove that the L-level of PP of this model is

$$f_{MS}(A) = \begin{cases} \pi(1 - \phi)/(1 - \lambda_{MS}) & \text{if } \phi < 1/2, \\ \pi\phi/\lambda_{MS} & \text{if } \phi \geqslant 1/2. \end{cases}$$

and the LW-measure is given by

$$g_{MS}(Y, A) = g_{MS}(N, A^c) = \frac{\phi}{(1 - \phi)},$$

where $\lambda_{MS} = \pi\phi + (1 - \pi)(1 - \phi)$ and $\phi = p_1^{MS} + (1 - p_1^{MS})p_2^{MS}$. The efficiency of the estimator $\hat{\pi}_{MS} = [\hat{\lambda} - (1 - \phi)]/(2\phi - 1)$ is $V_{MS} = \lambda_{MS}(1 - \lambda_{MS})/[n(2\phi - 1)^2]$.

The following proposition states that, for every Mangat and Singh (1990) model, there exists at least a Warner's procedure with equal privacy protection and equal efficiency and vice versa:

Proposition 1. *For every $\pi \in (0, 1)$, the family of Mangat and Singh (1990) models is $\mathscr{PP}$-equivalent to the family of Warner's models with $p^W = p_1^{MS} + (1 - p_1^{MS})p_2^{MS}$.*

Proof. Substituting p^W with $p_1^{MS} + (1 - p_1^{MS})p_2^{MS}$, the design probabilities of the two families of models become equal for all $p_1^{MS}, p_2^{MS} \in (0, 1)$, that is $P_W(Y|A) = P_{MS}(Y|A)$, and $P_W(Y|A^c) = P_{MS}(Y|A^c)$. Therefore $\lambda_W = \lambda_{MS}$, $\hat{\pi}_W = \hat{\pi}_{MS}$, $V_W = V_{MS}$ and the levels of privacy protection, for every PP measure based on design probabilities, are equal (e.g. $f_W(A) = f_{MS}(A)$, $g_W(Y, A) = g_{MS}(Y, A)$ and $g_W(N, A^c) = g_{MS}(N, A^c)$) This proves that for every Mangat and Singh (1990) model there exists at least one Warner's model with equal efficiency and respondent's protection. The converse is also true: for every Warner's model of parameter p^W there exists at least one $\mathscr{PP}$-equivalent Mangat and Singh (1990) model of parameters $(p_1^{MS}, p_2^{MS}) = (p_1^{MS}, (p^W - p_1^{MS})/(1 - p_1^{MS})), p_1^{MS} < p^W$. $\square$

4.2. The Chang and Liang (1996) model

The method of Chang and Liang (1996) is analogous to the two-stage procedure of Mangat and Singh (1990) with the difference that the second RD is a Simmons' model with parameters (p_2^{CL}, p_3^{CL}) instead of a Warner's device. Therefore, the vector of parameters characterizing this family of RR models is the vector $\theta^{CL} = (p_1^{CL}, p_2^{CL}, p_3^{CL})$.

In this case the L-measure and the LW-measure of protection are, respectively,

$$f_C(A) = \pi[p_1^{CL} + (1 - p_1^{CL})p_2^{CL} + (1 - p_1^{CL})(1 - p_2^{CL})p_3^{CL}]/\lambda_C$$

$$g_C(Y, A) = \frac{p_1^{CL} + (1 - p_1^{CL})p_2^{CL} + (1 - p_1^{CL})(1 - p_2^{CL})p_3^{CL}}{(1 - p_1^{CL})(1 - p_2^{CL})p_3^{CL}},$$

where $\lambda_C = \pi[p_1^{CL} + (1 - p_1^{CL})p_2^{CL}] + (1 - p_1^{CL})(1 - p_2^{CL})p_3^{CL}$. Moreover, the efficiency of the estimator $\hat{\pi} = [\hat{\lambda} - (1 - p_1^{CL})(1 - p_2^{CL})p_3^{CL}]/(p_2^{CL} + p_1^{CL}(1 - p_2^{CL}))$ is $V_C = \lambda_C(1 - \lambda_C)/[n(p_1^{CL} + (1 - p_1^{CL})p_2^{CL})^2]$.

We state the following equivalence:

Proposition 2. *For every proportion $\pi \in (0, 1)$, the family of Chang and Liang (1996) models is $\mathscr{PP}$-equivalent to the family of Simmons' models with $p_1^S = p_1^{CL} + (1 - p_1^{CL})p_2^{CL}$ and $p_2^S = p_3^{CL}$.*

Proof. Analogous to the Proof of Proposition 1. $\square$

4.3. The Chang et al. (2004) model

The RR procedure proposed by Chang et al. (2004) has one RD which consists of four statements: "I belong to group A", "I do not belong to group A", 'Yes', 'No', with probabilities $p_1^{CH}, p_2^{CH}, p_3^{CH}$ and p_4^{CH} ($\sum_{i=1}^4 p_i^{CH} = 1$), respectively.

When $p_3^{CH} = 0$ and $p_4^{CH} = 1 - p_1^{CH} - p_2^{CH}$, the above model reduces to the Mangat et al. (1990) model. When $p_4^{CH} = 0$ and $p_3^{CH} = 1 - p_1^{CH} - p_2^{CH}$, the model reduces to the Bhargava and Singh (2000) model.

The L-level of PP of the Chang et al. (2004) model is

$$f_{CH}(A) = \begin{cases} \pi(p_1^{CH} + p_3^{CH})/\lambda_{CH} & \text{if } \theta^{CH} \in \Theta^{CH}, \\ \pi(1 - p_1^{CH} - p_3^{CH})/(1 - \lambda_{CH}) & \text{if } \theta^{CH} \notin \Theta^{CH}. \end{cases}$$

and the LW-measure is given by

$$g_{CH}(Y, A) = \frac{p_1^{CH} + p_3^{CH}}{p_2^{CH} + p_3^{CH}},$$

where $\Theta^{CH} = \{(p_1, p_2, p_3) | p_2 < 1/2 - |1/2 - p_1|, p_3 < 1 - p_1 - p_2\}$, $\theta^{CH} = (p_1^{CH}, p_2^{CH}, p_3^{CH})$ and $\lambda_{CH} = \pi p_1^{CH} + (1 - \pi)p_2^{CH} + p_3^{CH}$. The efficiency of the estimator $\hat{\pi}_{CH} = (\hat{\lambda} - p_2^{CH})/(p_1^{CH} - p_2^{CH})$ is $V_{CH} = \lambda_{CH}(1 - \lambda_{CH})/[n(p_1^{CH} - p_2^{CH})^2]$.

We state the following equivalence:

Proposition 3. *For every $\pi \in (0, 1)$, the family of RR models proposed by Chang et al. (2004) with $\theta^{CH} \in \Theta^{CH}$ is $\mathscr{PP}$-equivalent to the family of Simmons' models, with $p_1^S = p_1^{CH} - p_2^{CH}$ and $p_2^S = (p_2^{CH} + p_3^{CH})/(1 - p_1^{CH} + p_2^{CH})$.*

Proof. Analogous to the Proof of Proposition 1. $\square$

The procedures proposed by Mangat et al. (1990) and Bhargava and Singh (2000) with parameters $\theta^{CH} \in \Theta^{CH}$ are $\mathscr{PP}$-equivalent to the family of Simmons' models since they are special cases of the Chang et al. (2004) model.

4.4. The Kuk (1990) model

In the procedure proposed by Kuk (1990) each respondent generates two binary outcomes according to two Bernoulli distributions with known parameters p_1^K and p_2^K, respectively. If the respondent has attribute A, he/she reports the outcome of the first random device. Otherwise, the respondent is required to report the second outcome.

The crucial difference between this model and the others here considered lies in the type of response: generic 'R1' or 'R2' responses (the outcomes of two Bernoulli distributions) instead of 'Yes' or 'No'. The idea is that the embarrassment of a respondent giving an 'R1'/'R2' answer is less than giving 'Yes'/'No'. Unfortunately, the family of PP-measures here considered is not able to quantify this additional level of protection because all these measures are based on the hypothesis that $R \in \{\text{'Yes', 'No'}\}$. The analysis of this important question goes beyond the scope of the present work. Therefore we consider $R \in \{\text{'R1', 'R2'}\}$ equivalent in terms of perceived protection to $R \in \{\text{'Yes', 'No'}\}$. Under this hypothesis, the L-level of respondent's protection of this model is

$$f_K(A) = \begin{cases} \pi(1 - p_1^K)/(1 - \lambda_K) & \text{if } p_1^K < p_2^K, \\ \pi p_1^K / \lambda_K & \text{if } p_1^K > p_2^K. \end{cases}$$

and the LW-measure is given by $g_K(Y, A) = p_1^K/p_2^K$, where $\lambda_K = \pi p_1^K + (1 - \pi)p_2^K$. The efficiency of the Kuk's estimator $\hat{\pi}_K = (\hat{\lambda} - p_2^K)/(p_1^K - p_2^K)$ is $V_K = \lambda_K(1 - \lambda_K)/[n(p_1^K - p_2^K)^2]$.

Another equivalence follows:

Proposition 4. *For every $\pi \in (0, 1)$, the family of Kuk (1990) models with $p_1^K > p_2^K$ is $\mathscr{PP}$-equivalent to the family of Simmons' models, with $p_1^S = p_1^K - p_2^K$ and $p_2^S = p_2^K/(1 - p_1^K + p_2^K)$.*

Proof. Analogous to the Proof of Proposition 1. $\square$

Proposition 5. *For $p_1^K < p_2^K$, $\pi \in (0, 1)$ and for each Simmons' model, there exists at least one Kuk (1990) model with equal L-level of protection and higher efficiency.*

Proof. By simple algebra, it can be shown that, given a Simmons' model of parameter (p_1^S, p_2^S), the Kuk's model with equal L-level of protection and higher efficiency has parameters $(p_1^K, p_2^K) \in \Theta^K$, where

$$\Theta^K = \left\{ (p_1, p_2) | p_1 < (1 - p_1^S)(1 - p_2^S), \ p_2 = \frac{(1 - p_1^S)(1 - p_1 p_2^S) - 1}{(1 - p_1^S)(1 - p_2^S) - 1} \right\}. \qquad \square$$

4.5. The Christofides (2003) model

In this model each respondent is provided with a RD which produces the integers $1, 2, \ldots, K$ with probabilities $p_1^C, p_2^C, \ldots, p_K^C$ ($\sum_{k=1}^K p_k^C = 1$). Using this RD, the individual produces one of the K integers and he/she reports how far away this number is from $K + 1$ if he/she has the characteristic or from 0 if he/she does not have it.

This model generalizes in some sense the idea of Kuk (1990): if an 'R1'/'R2' response could be perceived more protective than a 'Yes'/'No' response by the interviewed subjects, giving an integer between 1 and K could be perceived even more protective. Like in the model of Kuk (1990), the PP-measures here considered are not able to quantify this additional perceived protection.

The L- and LW-measures of protection are

$$f_C(A) = \max_k \left\{ \frac{\pi p_{K+1-k}^C}{\pi p_{K+1-k}^C + (1 - \pi) p_k^C} \right\}$$

$$g_C(k|A) = p_{K+1-k}^C / p_k^C, \quad k = 1, 2, \ldots, K.$$

The efficiency of the estimator $\hat{\pi} = (\bar{d} - E(y))/(K + 1 - 2E(y))$ is $V_C = \pi(1 - \pi)/n + V(y)/(n(K + 1 - 2E(y))^2)$, where $\bar{d}$ is the sample average of the responses, $E(y) = \sum_{k=1}^K k p_k^C$ is the expected value of the output of the random device and $V(y) = \sum_{k=1}^K (k - E(y))^2 p_k^C$ its variance.

In the special case $K = 2$, we have the following equivalence:

Proposition 6. *For every $\pi \in (0, 1)$, the family of Christofides (2003) models with $K = 2$ is $\mathscr{PP}$-equivalent to the family of Warner's models, with $p^W = p_1^C$.*

For $K = 3$ the comparison between the Simmons' and the Christofides' model gives:

Proposition 7. *For every Simmons' model and for every $\pi \in (0, 1)$ there exists at least one $L = 3$ Christofides (2003) model with higher efficiency and better L-protection.*

Proof. By simple algebra, one can show that Christofides' models of parameters (p_1^C, p_2^C) have better performances than a given Simmons' model with parameters (p_1^S, p_2^S) if $(p_1^C, p_2^C) \in \Theta^C$, where

$$\Theta^C = \left\{ (p_1, p_2) | 0 < p_2 < \frac{B}{A} \cap \frac{1 - p_2}{2} - \frac{p_1^S(1 - p_2)}{2C} < p_1 < \frac{1 - p_2}{2} - \frac{1}{2} p_1^S \sqrt{\frac{1 - p_2}{A}} \right\},$$

$A = B + C^2$, $B = 4(1 - p_1^S)(1 - 2p_2^S)(p_2^S - p_1^S p_2^S - \pi p_1^S)$ and $C = p_1^S + 2p_2^S(1 - p_1^S)$.

This result excludes the existence of a $\mathscr{PP}$-equivalence between Simmons' models and Christofides' models with $K \geqslant 3$. Moreover, it proves the superiority of the latter on the former when using the L-measure of Lanke (1976). $\quad\square$

4.6. The Mangat (1994) model

In the procedure proposed by Mangat (1994), each respondent is instructed to say 'Yes' if he/she has the attribute A or to use a Warner RD with probability p^M if he/she does not have attribute A.

The L- and LW-levels of PP are, respectively, given by

$$f_M(A) = \frac{\pi}{\lambda_M}, \quad g_M(Y, A) = \frac{1}{1 - p^M}, \quad g_M(N, A^c) = \frac{p^M}{0},$$

where $\lambda_M = \pi + (1 - \pi)(1 - p^M)$. The efficiency of the Mangat's estimator $\hat{\pi}_M = (\hat{\lambda} - 1 + p^M)/p^M$ is $V_M = \lambda_M(1 - \lambda_M)/[n(p^M)^2]$.

Proposition 8. *For every model of the Mangat (1994) family, for every $\pi \in (0, 1)$ and every PP-measure $\in \mathscr{PP}$, there exists one PP-equivalent Simmons' model with parameters $p_1^S = p^M$ and $p_2^S \to 1$. The converse is not true.*

5. Discussion

The propositions reported in this note prove some fundamental facts:

(1) About 25–30 years after the publication of the seminal works of Warner (1965) and Greenberg et al. (1969), some RR models have been proposed in the literature with the claim that they are superior to these earlier models. The equivalences stated in this paper show that many new procedures actually do not offer any improvement in terms of efficiency and protection (for any adopted measure of protection). Therefore, their usefulness must be determined on the basis of other characteristics: the degree of complexity, the level of understanding of the respondents, the total costs and others.

(2) The two-stage RR models of Mangat and Singh (1990) and Chang and Liang (1996) have more complex designs and higher costs of implementation than Warner's or Simmons' procedures without offering better performances. The method of Mangat (1994), equivalent to a Simmons' procedure, does not allow to set an upper bound on the level of jeopardizing of the 'No' answer with regard to A^c. In some cases, this could be a problem. Moreover, a major criticism of the model proposed by Chang et al. (2004) when seeking Institutional Review Board approvals is that it is somewhat unethical to ask a respondent to give untruthful response.

(3) The family of models proposed by Kuk (1990) is better than the Simmons' family in terms of efficiency and privacy protection. From an empirical point of view, the study of van der Heijden et al. (2000) shows that Kuk's procedure seems to perform slightly better than the forced-response procedure and markedly better than face-to-face direct questioning and computer assisted self-interviewing. Recommendations and successful applications of the Kuk's procedure have been also reported in van den Hout and van der Heijden (2002). These results should be even more marked for the model proposed by Christofides (2003).

(4) An adequate analysis of the efficiency and of the respondent's protection is always necessary when proposing new RR models.

References

Bhargava, M., Singh, R., 2000. A modified randomization device for Warner's model. Statistica Anno LX(2), 315–321.

Bhargava, M., Singh, R., 2002. On the efficiency comparison of certain randomized response strategies. Metrika 55, 191–197.

Chang, H.-J., Liang, D.-H., 1996. A two-stage unrelated randomized response procedure. Austral. J. Statist. 38, 43–51.

Chang, H.-J., Wang, C.-L., Huang, K.-C., 2004. On estimating the proportion of a qualitative sensitive character using randomized response sampling. Quality & Quantity 38, 675–680.

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

Greenberg, B.G., Abul-Ela, A.A., Simmons, W.R., Horvitz, D.G., 1969. The unrelated question randomized response model: theoretical frameworks. J. Amer. Statist. Assoc. 64, 520–539.

van der Heijden, P.G.M., van Gils, G., Bouts, J., Hox, J.J., 2000. A comparison of randomized response, computer-assisted self-interview, and face-to-face direct questioning. Eliciting sensitive information in the context of welfare and unemployment benefit. Sociological Methods & Res. 28, 505–537.

van den Hout, A., van der Heijden, P.G.M., 2002. Randomized response, statistical disclosure control and misclassification: a review. Internat. Statist. Rev. 70, 269–288.

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

Lanke, J., 1976. On the degree of protection in randomized interviews. Internat. Statist. Rev. 44, 197–203.

Leysieffer, F.W., Warner, S.L., 1976. Respondent jeopardy and optimal designs in randomized response models. J. Amer. Statist. Assoc. 71, 649–656.

Ljungqvist, L., 1993. A unified approach to measures of privacy in randomized response models: a utilitarian perspective. J. Amer. Statist. Assoc. 88, 97–103.

Mangat, N.S., 1994. An improved randomized response strategy. J. Roy. Statist. Soc. B 56, 93–95.

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

Mangat, N.S., Singh, S., Singh, R., 1990. On use of a modified randomization device in Warner's model. J. Indian Soc. Statist. Oper. Res. 16, 627–629.

Nayak, T.K., 1994. On randomized response surveys for estimating a proportion. Comm. Statist. Theory Methods 23, 3303–3321.

Warner, S.L., 1965. Randomized response: a survey technique for eliminating evasive answer bias. J. Amer. Statist. Assoc. 60, 63–69.

Zou, G., 1997. Two-stage randomized response procedures as single stage procedures. Austral. J. Statist. 39, 235–236.