Department for Applied Statistics
Johannes Kepler University Linz

IFAS Research Paper Series
2010-50

A randomized response strategy to be used in opinion polls

Andreas Quatember

October 2010

A randomized response strategy to be used in opinion polls

1. INTRODUCTION

Nonresponse and untruthful answering are always present in the practice of sample and population surveys. Such a respondent's behavior causes serious problems in the analysis of sample as well as population data, if the underlying mechanisms are not "completely at random" (cf. Little and Rubin 2002, p.12). Subsequently we will call variables, for which this is true, "sensitive". Examples are domestic violence, illegal employment, drug use, tax evasion or voting behavior.

Let U be an universe of N population units. Furthermore, let s be a probability sample of size n with element inclusion probabilities $\rho_k > 0$ ($\forall\, k = 1, 2, ..., N$; $s \subseteq U$; $n \leq N$). In the presence of nonresponse and untruthful answering s is divided into three disjoint sets of sampled elements: A set t of truthfully answering units, a set u of elements, who answer untruthfully, and a "missing set" m ($s = t + u + m$). When we wish to estimate for instance the total τ of a variable y, in the Horvitz-Thompson estimator $\widehat{\tau}$ the sample sum of the product $y_k \cdot d_k$ with the design weights $d_k = \frac{1}{\rho_k}$ is decomposed into

$$\widehat{\tau} = \sum\nolimits_s y_k \cdot d_k = \sum\nolimits_t y_k \cdot d_k + \sum\nolimits_u y_k \cdot d_k + \sum\nolimits_m y_k \cdot d_k \tag{1}$$

($\sum_s$ is abbreviated notation for $\sum_{k \in s}$ or $\sum_{k=1}^{n}$). The y_k's in the last two sums of (1) cannot be observed. Furthermore, the estimation of τ on the basis of the observed data in the sets t and u can be biased seriously. This will happen, if the nonresponse and the untruthful answering mechanism are far from "completely at random" and the u and/or m are of non-negligible size. Methods like data imputation and weighting adjustment only account for nonresponse by estimating the missing y_k's in m or increasing the design weights of the observed elements according to some models.

But there also exists a family of statistical methods, which considers both nonresponse and untruthful answering. These procedures of "randomized response" try to minimize both rates by an increase of the respondents' privacy protection. For this purpose a questioning design is used, which – in contrast to the direct questioning on the sensitive subject – does not allow the data collector to assign the given information to the sensitive variable.

The pioneering work in this field was published by Warner (1965). Since then researchers have tried in different ways to increase the efficiency of the methods (cf. for instance: Yu et al. 2008, or Christofides 2003), have presented generalizations of such questioning designs (cf. for instance: Quatember 2009, or Chaudhuri 2001) and have applied the methods in various studies (cf. for instance: Miner and Center 2008, or Lensvelt-Mulders et al. 2005).

In the subsequent section the randomized response strategy for the estimation of the relative category sizes of a categorical variable presented by Liu and Chow (1976) for simple random sampling with replacement will be extended to all probability sampling designs. In section 3 the case of respondents choosing the direct response option over the randomized response option is considered and its effect on the estimation process is presented. In the fourth section the privacy protection offered by the questioning design is discussed. The concluding section considers the application of the proposed strategy to opinion polls along with an illustrative example.

2. THE PROPOSED STRATEGY

Let population U be divided by a categorical variable y with categories $1, 2, ..., m$ into m non-overlapping subgroups $U_1, U_2, ..., U_m$ of sizes $N_1, N_2, ..., N_m$ ($U = \bigcup U_i$, $N = \sum N_i$, $U_i \bigcap U_j = \emptyset \ \forall \ i \neq j, \ i, j = 1, 2, ..., m$). Furthermore the parameters of interest be the relative sizes $\pi_1, \pi_2, ..., \pi_m$ of the subgroups with

$$\pi_i = \frac{N_i}{N} \tag{2}$$

($i = 1, 2..., m$). For instance someone might be interested in the proportions of some categories itself or – in the case of an ordinal or a quantitative discrete variable y – the proportions could be needed to be able to calculate measures of position or dispersion. Liu and Chow (1976) explicated a randomized response questioning design for the estimation of such proportions for simple random sampling with replacement. Its theory can be extended to any probability sampling design in the following way: A data collector asks a survey unit k of a probabaility sample s with probability p_0 for his or her value y_k of variable y. Furthermore with probabilities $p_1, p_2, ..., p_m$ the respondent is instructed just to answer "category 1", "category 2", ..., "category m" ($\sum_{i=1}^{m} p_i = 1 - p_0$).

Let for unit k variable value $y_{ki} = \begin{cases} 1 & \text{if } y_k = i, \\ 0 & \text{otherwise} \end{cases}$ ($i = 1, 2, ..., m$) indicate the membership of group U_i. Moreover, let z_k be the response category of survey unit k and

$$z_{ki} = \begin{cases} 1 & \text{if } z_k = i, \\ 0 & \text{otherwise.} \end{cases}$$

Assuming cooperation, the probability of $z_{ki} = 1$ with respect to the randomized response questioning design R is for given y_{ki}:

$$P_R(z_{ki} = 1) = p_0 \cdot y_{ki} + p_i. \tag{3}$$

Then the term

$$\widehat{y}_{ki} = \frac{z_{ki} - p_i}{p_0} \tag{4}$$

($p_0 \neq 0$) is unbiased for the true value y_{ki} . Using these "estimates" the following theorems apply:

Theorem 1: For a probability sampling design P with design weights d_k

$$\widehat{\pi}_i = \frac{1}{N} \cdot \sum\nolimits_s \widehat{y}_{ki} \cdot d_k \tag{5}$$

is an unbiased estimator of the relative size π_i of group U_i ($i = 1, 2, ..., m$).

Theorem 2: The variance of $\widehat{\pi}_i$ is given by

$$\begin{aligned} V(\widehat{\pi}_i) &= \frac{1}{N^2} \cdot \Big[V_P \Big(\sum\nolimits_s y_{ki} \cdot d_k \Big) + \\ &+ \frac{1}{p_0^2} \cdot \Big(p_i \cdot (1-p_i) \cdot \sum\nolimits_U d_k + p_0 \cdot (1 - p_0 - 2 \cdot p_i) \cdot \sum\nolimits_U y_{ki} \cdot d_k \Big) \Big] \end{aligned} \quad (6)$$

$(i = 1, 2, ..., m)$. The first term within the squared bracket of (6) corresponds to the variance formula for the direct questioning under the assumption of full cooperation. Thereforer the second one can be seen as the price that has to be paid by the data analyst for the increased privacy protection of the respondents.

Theorem 3: Variance (6) can be estimated unbiasedly by

$$\begin{aligned} \widehat{V}(\widehat{\pi}_i) &= \frac{1}{N^2} \cdot \Big[\widehat{V}_P \Big(\sum\nolimits_s y_{ki} \cdot d_k \Big) + \\ &+ \frac{1}{p_0^2} \cdot \Big(p_i \cdot (1-p_i) \cdot \sum\nolimits_U d_k + p_0 \cdot (1 - p_0 - 2 \cdot p_i) \cdot \sum\nolimits_s \widehat{y}_{ki} \cdot d_k^2 \Big] \Big) \end{aligned} \quad (7)$$

$(i = 1, 2, ..., m)$. For the proofs of all three theorems see the Appendix.

Notice, that for simple random sampling with replacement, formula (7) yields Liu and Chow's (1976) variance estimator (ibd., p.31).

3. COMBINING DIRECT AND RANDOMIZED RESPONSES

In practice it may often happen, that for instance a telephone interviewer recognizes during the necessary explanation of the randomized response questioning design, that the interviewee is willing to deliver the sensitive information without the randomization device. Such a respondent's behavior can be incorporated in the estimation procedure in the following way: Let population U be divided into a subpopulation U_D of elements not needing any privacy protection to provide the information requested and a disjoint group U_R, whose members will use the offered randomization $(U = U_D + U_R)$. Units $k \in U_D$ will deliver their true value y_k, the others the response category z_k $(k = 1, 2, ..., N)$. Therefore the probability sample s will also be divided into a group s_D with true and a group s_R with randomized responses on the sensitive variable $(s = s_D + s_R)$. This modification in the respondents' behavior affects the estimation procedure described in the previous section in the following way:

Theorem 4: For a mixture (M) of direct answers y_k and randomized responses z_k

$$\widehat{\pi}_i^M = \frac{1}{N} \cdot \Big(\sum\nolimits_{s_D} y_{ki} \cdot d_k + \sum\nolimits_{s_R} \widehat{y}_{ki} \cdot d_k \Big) \quad (8)$$

is unbiased for π_i $(i = 1, 2, ..., m)$ for any probability sampling design P.

Theorem 5: The variance of $\widehat{\pi}_i^M$ is given by

$$\begin{aligned} V(\widehat{\pi}_i^M) &= \frac{1}{N^2} \cdot \Big[V_P \Big(\sum\nolimits_s y_{ki} \cdot d_k \Big) + \\ &+ \frac{1}{p_0^2} \cdot \Big(p_i \cdot (1-p_i) \cdot \sum\nolimits_{U_R} d_k + p_0 \cdot (1 - p_0 - 2 \cdot p_i) \cdot \sum\nolimits_{U_R} y_{ki} \cdot d_k \Big) \Big] \end{aligned} \quad (9)$$

$(i = 1, 2, ..., m)$. The second term within the squared bracket of (9) is zero, when $U_R = \varnothing$ (i.e. y is a non-sensitive variable) and it corresponds to the second term of (6), when $U_D = \varnothing$. It follows, that $V(\widehat{\pi}_i^M) \leq V(\widehat{\pi}_i)$.

Theorem 6: Variance (9) can be estimated unbiasedly by

$$\widehat{V}(\widehat{\pi}_i^M) = \frac{1}{N^2} \cdot \left[\widehat{V}_P \left(\sum\nolimits_s y_{ki} \cdot d_k \right) + \right.$$
$$\left. + \frac{1}{p_0^2} \cdot \left(p_i \cdot (1 - p_i) \cdot \sum\nolimits_{s_R} d_k^2 + p_0 \cdot (1 - p_0 - 2 \cdot p_i) \cdot \sum\nolimits_{s_R} \widehat{y}_{ki} \cdot d_k^2 \right) \right] \quad (10)$$

$(i = 1, 2, ..., m)$. For the proofs of theorems 4 to 6 see the Appendix.

4. MEASURING PRIVACY PROTECTION

Let us define the following measures λ_i of loss of privacy according to the different categories i $(i = 1, 2, ..., m)$:

$$\lambda_i = \frac{P(z_k = i | y_k = i)}{P(z_k = i | y_k \neq i)} \quad (11)$$

with $1 \leq \lambda_i \leq \infty$ (cf. Leysieffer and Warner, 1976). λ_i refers to the privacy protection with respect to a response $z_k = i$. For the proposed technique of randomized response these measures are given by

$$\lambda_i = \frac{p_0 + p_i}{p_i}. \quad (12)$$

$\lambda_i = 1$ corresponds to a totally protected privacy. This can only happen for $p_0 = 0$, so that the answers are randomly drawn with design probabilities p_1 to p_m. The more λ_i differs from one, the more information on the membership of group U_i is contained in the given answer z_k and the less is the survey unit protected against disclosure. A probability $p_i = 0$ $(i = 1, 2, ..., m)$ yields $z_{ki} = y_{ki}$. The loss of privacy with respect to an answer given achieves its maximum $\lambda_i = \infty$.

If the membership of a certain category i of variable y is more sensitive than of another category, its measure λ_i of loss of privacy should be less than that of the other categories with regard to the respondents' willingness to cooperate. In other words the design probabilities p_0 and p_1 to p_m should be chosen in accordance to the sensitivity levels of the different categories.

5. AN APPLICATION TO OPINION POLLS

During the last decades results from opinion polls conducted shortly before a forthcoming election have increasingly differed from the actual outcome of the election. Particularly the proportions of parties from the two margins of the political spectrum have constantly been underestimated. In Austria, for instance, representatives of such parties occused opinion researchers of data manipulation. As a consequence, these politicians claimed a prohibition of opinion polls short before elections. Actually besides the increasingly late moment of voting decision mainly the growing sensitivity of the variable "voting behavior" and particularly of certain categories seem to be the main reason for this phenomenon.

In a telephone or face-to-face survey the proposed questioning design is implemented in the following way (for postal or web surveys the procedure has to be adapted; see for instance: Lensvelt-Mulders et al. 2006, or Lensvelt-Mulders and Boeije 2007): The interviewer tells the survey unit, that due to the sensitivity of the subject a questioning design will be applied, which allows to protect the respondent's privacy. Of course, the effect of this questioning design on data protection has to be explained to the respondents in a vivid way to produce the desired willingness to cooperate (cf. Landsheer et al. 1999, p.6ff). Then the respondent is asked to think of a person, of whom he or she knows the mobile phone number or the dates of birth without delivering this information to the interviewer (for other random devices see for instance: Warner 1986). If – let's say – the last two digits of the phone number are within a certain interval like from 00 to 79, the respondent shall answer truthfully to a question like "Imagine it's election day. Which party gets your vote?". The $(m-1)$ parties in question and a non-voting category give altogether m possible answers. But if the two digits of the mobile phone number lie within an interval of the remaining combinations (for example from 80 to 84), the respondent shall just answer "party 1". If the digits lie within another non-overlapping interval, the survey unit shall just answer "party 2" and so on. After all, these m disjoint intervals have to cover all possible combinations. The chosen intervals determine the design probabilities p_1 to p_m ($\sum p_i = 1 - p_0$). Telephone companies or organizations like the national statistical agencies might be able to deliver the required information.

If a respondent reveals during the design explanations that he or she is willing to answer directly on the sensitive question (for instance by saying "It's no problem for me to answer this question honestly!"), this answer must be flagged. Then estimator (8) can be applied.

6. A NUMERICAL EXAMPLE

Consider the following example just to get an idea of the effects of the proposed strategy: A large population (of eligible voters for instance) is given. The variable y of interest has – say – $m = 4$ categories with relative sizes $\pi_1 = 0.4$, $\pi_2 = 0.3$, $\pi_3 = 0.2$ and $\pi_4 = 0.1$ to be estimated. For this purpose a simple random sample without replacement of size $n = 1,000$ is drawn from the population. As design probabilities at first we choose $p_0 = 0.6$ and p_1 to p_4 all equal to 0.1. This yields measures of loss of privacy (section 4) of $\lambda_i = 7 \; \forall \; i = 1, ..., 4$.

Using the proposed randomized response strategy and assuming full cooperation of the respondents, the estimators are unbiased and their theoretical variances given by

$V(\widehat{\pi}_1) = 0.62\dot{3} \cdot 10^{-3}$
$V(\widehat{\pi}_2) = 0.56 \cdot 10^{-3}$
$V(\widehat{\pi}_3) = 0.47\dot{6} \cdot 10^{-3}$
$V(\widehat{\pi}_4) = 0.37\dot{3} \cdot 10^{-3}$.

Assuming next, that 70% of those elements of U belonging to group U_1 during the interview do signal their willingness to answer directly on the question and 50% of group U_2, 30% of group U_3 and 10% of group U_4 do so, the variances of the unbiased estimators $\widehat{\pi}_i^M$ decrease to

$V(\widehat{\pi}_1^M) = 0.405 \cdot 10^{-3}$
$V(\widehat{\pi}_2^M) = 0.385 \cdot 10^{-3}$

$V(\widehat{\pi}_3^M) = 0.332 \cdot 10^{-3}$
$V(\widehat{\pi}_4^M) = 0.245 \cdot 10^{-3}$.

The reduction of the variances compared to the pure randomized response strategy lies between 30 to 35%. Note, that if all sampled units would answer truthfully on the direct question, the four variances would be $0.24 \cdot 10^{-3}$, $0.21 \cdot 10^{-3}$, $0.16 \cdot 10^{-3}$ and $0.09 \cdot 10^{-3}$.

At last we increase the probability p_0 to 0.8 and fix the other four design probabilities at 0.05. This yields a higher measure of loss of privacy for a less sensitive variable as above: $\lambda_i = 17 \; \forall \; i = 1, ..., 4$. The variances of the estimators $\widehat{\pi}_i$ with the pure randomization and those using the mixture of direct and randomized responses ($\widehat{\pi}_1^M$) according to the mixing proportions given above would then decrease to
$V(\widehat{\pi}_1) = 0.364 \cdot 10^{-3}$, $V(\widehat{\pi}_1^M) = 0.292 \cdot 10^{-3}$,
$V(\widehat{\pi}_1) = 0.322 \cdot 10^{-3}$, $V(\widehat{\pi}_2^M) = 0.266 \cdot 10^{-3}$,
$V(\widehat{\pi}_1) = 0.259 \cdot 10^{-3}$, $V(\widehat{\pi}_3^M) = 0.215 \cdot 10^{-3}$,
$V(\widehat{\pi}_1) = 0.177 \cdot 10^{-3}$, $V(\widehat{\pi}_4^M) = 0.138 \cdot 10^{-3}$.

The interested readers may try to obtain further results by their own by inserting differing design probabilities and mixing proportions into the formulae.

7. SUMMARY

The randomized response questioning design of Liu and Chow (1976) for sensitive categorical variables was extended to all probability sampling designs. The statistical properties of estimators of the relative sizes of m disjoint subpopulations were presented. Furthermore the case of survey units, who disclaim the randomization procedure in favor of the direct answer on the sensitive question under study was implemented and developed. The privacy protection view of the questioning design was also considered and an application on opinion polls presented along with a numerical example. The results show exemplarily the effect of the proposed randomized response questioning design on the efficiency of estimators.

If the direct questioning on the sensitive subject leads to considerable biases of the estimators, the higher complexity of the randomized response questioning design will pay. The accuracy of the estimators increases then although their variances exceed the theoretical ones of the direct questioning.

APPENDIX:

PROOFS OF THEOREMS 1 TO 3:

$$E(\widehat{\pi}_i) = \frac{1}{N} \cdot E_P \left(E_R \left(\sum\nolimits_s \widehat{y}_{ki} \cdot d_k | s \right) \right) = \frac{1}{N} \cdot E_P \left(\sum\nolimits_s y_{ki} \cdot d_k \right) = \frac{1}{N} \cdot \sum\nolimits_U y_{ki} = \pi_i.$$

Furthermore the variance is given by

$$V(\widehat{\pi}_i) = V_P(E_R(\widehat{\pi}_i|s)) + E_P(V_R(\widehat{\pi}_i|s)),$$

where

$$V_P(E_R(\widehat{\pi}_i|s)) = \frac{1}{N^2} \cdot V_P \left(\sum\nolimits_s y_{ki} \cdot d_k \right).$$

Let $I_k = \mathbf{1}(k \in s)$ $(k = 1, 2, ..., N)$ indicate the sample inclusion of a survey unit k with $E_P(I_k) = \frac{1}{d_k}$. Because covariance $C_R(\widehat{y}_{ki}, \widehat{y}_{li}|s) = 0 \; \forall \; k \neq l$, the following expectation is derived:

$$\begin{aligned} E_P(V_R(\widehat{\pi}_i|s)) &= \frac{1}{N^2} \cdot E_P\left(V_R\left(\sum\nolimits_U I_k \cdot \widehat{y}_{ki} \cdot d_k|s\right)\right) \\ &= \frac{1}{N^2} \cdot E_P\left(\sum\nolimits_U I_k^2 \cdot d_k^2 \cdot V_R(\widehat{y}_{ki})\right) \\ &= \frac{1}{N^2} \cdot \sum\nolimits_U V_R(\widehat{y}_{ki}) \cdot d_k. \end{aligned}$$

$V_R(\widehat{y}_{ki})$ results in

$$V_R(\widehat{y}_{ki}) = \frac{1}{p_0^2} \cdot V_R(z_{ki})$$

and

$$\begin{aligned} V_R(z_{ki}) &= E_R(z_{ki}^2) - E_R^2(z_{ki}) \\ &= p_0 \cdot y_{ki} + p_i - (p_0 \cdot y_{ki} + p_i)^2 \\ &= p_0 \cdot y_{ki} + p_i - p_0^2 \cdot y_{ki}^2 - 2 \cdot p_0 \cdot p_i \cdot y_{ki} - p_i^2 \\ &= p_i \cdot (1 - p_i) + p_0 \cdot y_{ki} \cdot (1 - p_0 - 2 \cdot p_i). \end{aligned}$$

It follows:

$$E_P(V_R(\widehat{\pi}_i|s)) = \frac{1}{N^2} \cdot \frac{1}{p_0^2} \cdot \left((p_i \cdot (1 - p_i) \cdot \sum\nolimits_U d_k + p_0 \cdot (1 - p_0 - 2 \cdot p_i) \cdot \sum\nolimits_U y_{ki} \cdot d_k\right)$$

This completes the proof of Theorem 2.

The theoretical variance $V(\widehat{\pi}_i)$ (6) can be estimated unbiasedly by inserting an unbiased estimator $\widehat{V}_P(\sum_s y_{ki} \cdot d_k)$ for $V_P(\sum_s y_{ki} \cdot d_k)$ and $\sum_s \widehat{y}_{ki} \cdot d_k^2$ for $\sum_U y_{ki} \cdot d_k$, because

$$\begin{aligned} E\left(\sum\nolimits_s \widehat{y}_{ki} \cdot d_k^2\right) &= E\left(\sum\nolimits_U \widehat{y}_{ki} \cdot d_k^2 \cdot I_k\right) \\ &= \sum\nolimits_U d_k^2 \cdot E(\widehat{y}_{ki} \cdot I_k) \\ &= \sum\nolimits_U y_{ki} \cdot d_k, \end{aligned}$$

which proofs Theorem 3.

PROOFS OF THEOREMS 4 TO 6

The unbiasedness of $\widehat{y}_{ki}$ for y_{ki} yields $E(\widehat{\pi}_i^M) = \pi_i$.

Again

$$V(\widehat{\pi}_i^M) = V_P(E_R(\widehat{\pi}_i^M|s)) + E_P(V_R(\widehat{\pi}_i^M|s)).$$

and

$$V_P(E_R(\widehat{\pi}_i^M|s)) = \frac{1}{N^2} \cdot V_P\left(\sum\nolimits_s y_{ki} \cdot d_k\right).$$

Furthermore

$$\begin{aligned} E_P(V_R(\widehat{\pi}_i^M|s)) &= \frac{1}{N^2} \cdot E_P\left(V_R\left(\sum\nolimits_{U_D} I_k \cdot y_{ki} \cdot d_k|s + \sum\nolimits_{U_R} I_k \cdot \widehat{y}_{ki} \cdot d_k|s\right)\right) \\ &= \frac{1}{N^2} \cdot E_P\left(\sum\nolimits_{U_R} I_k^2 \cdot d_k^2 \cdot V_R(\widehat{y}_{ki})\right) \\ &= \frac{1}{N^2} \cdot \sum\nolimits_{U_R} V_R(\widehat{y}_{ki}) \cdot d_k. \end{aligned}$$

Because of $V_R(\widehat{y}_{ki}) = \frac{1}{p_0^2} \cdot V_R(z_{ki})$ and $V_R(z_{ki})$ as above we have the result of Theorem 5.

The theoretical variance $V(\widehat{\pi}_i^M)$ (9) can be estimated unbiasedly by inserting again an unbiased estimator $\widehat{V}_P(\sum_s y_{ki} \cdot d_k)$ for $V_P(\sum_s y_{ki} \cdot d_k)$ and $\sum_{s_R} \widehat{y}_{ki} \cdot d_k^2$ for $\sum_{U_R} y_{ki} \cdot d_k$. Additionally the term $\sum_{s_R} d_k^2$ estimates $\sum_{U_R} d_k$ unbiasedly, which gives Theorem 6.

BIBLIOGRAPHY

Chaudhuri, A. (2001). Using randomized response from a complex survey to estimate a sensitive proportion in a dichotomous finite population. *Journal of Statistical Planning and Inference*, **57**, 37–42.

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

Landsheer, J. A., van der Heijden, P., and van Gils, G. (1999). Trust and Understanding, Two Psychological Aspects of Randomized Response. *Quality and Quantity*, **33**, 1–12.

Lensvelt-Mulders, G.J.L.M., and Boeije, H.R. (2007). Evaluating compliance with a computer assisted randomized response technique: a qualitative study into origins of lying and cheating. *Computers in Human Behavior*, **23**, 591–608.

Lensvelt-Mulders, G.J.L.M., van der Heijden, P.G.M. and Laudy, O. (2006). A validation of a computer-assisted randomized response survey to estimate the prevelance of fraud in social security. *Journal of the Royal Statistical Society, Series A*, **169**, Part 2, 305–318.

Lensvelt-Mulders, G.J.L.M., Hox, J. J., van der Heijden, P. G. M., and Maas, C. J. M. (2005). Meta-Analysis of Randomized Response Research. *Sociological Methods & Research*, **33** (3), 319–348.

Leysieffer, F. W. and Warner, S. L. (1976). Respondent Jeopardy and Optimal Designs in Randomized Response Models. *Journal of the American Statistical Association*, **71**, 649–656.

Little, R. J. A. and Rubin, D. B. (2002). *Statistical Analysis with Missing Data.* Hoboken: Wiley and Sons.

Liu, P.T., and Chow, L.P. (1976). A New Discrete Quantitative Randomized Response Model. *Journal of the American Statistical Association*, **71**, Number 353, 30–31.

Miner, M.H., and Center, B. (2008). Improving the Measurement of Criminal Sexual Behavior: The Application of Randomized Responding Technique. *Sexual Abuse: A Journal of research and Treatment*, **20**, Number 1, 88–101.

Quatember, A. (2009). A standardization of randomized response strategies. *Survey Methodology*, **32**, No.2, 143–152.

Warner, S. L. (1965). Randomized Response: A Survey Technique for Eliminating

Evasive Answer Bias. *Journal of the American Statistical Association*, **60**, 63–69.

Warner, S. L. (1986). The omitted digit randomized response model for telephone applications. *Proceedings of the Social Survey Research Methods, American Statistical Association*, 441–443.

Yu, J.-W., Tian, G.-L. and Tang, M.-L. (2008). Two new models for survey sampling with sensitive characteristic: design and analysis. *Metrika*, **67**, 251–263.