A Recommended Practice Manual for the Standardized Randomized Response Strategy

Andreas Quatember

January 2008
Abstract

The subject of this paper is the performance of the estimator of the standardized randomized response strategy (see: Quatember 2007). The comparison of the estimator's efficiency under simple random sampling without replacement for different designs has to take into account the level of privacy protection provided by the designs. This is done by the Leysieffer-Warner measures. A “recommended practice manual” is added, which helps the user to choose the optimal values for the design parameters of the different designs. The recommendations depend on whether the subject of interest is sensitive as a whole or only the possession but not the nonpossession of a certain attribute is awkward.

KEY WORDS: Data quality; Sampling theory; Survey methodology; Nonresponse; Randomized response technique; Privacy protection; Efficiency comparison

1 Introduction

The standardization of randomized response strategies for the estimation of proportions was introduced by Quatember (2007): Let $U$ be the universe of $N$ population units and $U_A$ be a subset of $N_A$ elements, that belong to a class $A$ of a categorial variable under study. Moreover let $U_{A^c}$ be the group of $N_{A^c}$ elements, that do not belong to this class ($U = U_A \cup U_{A^c}, U_A \cap U_{A^c} = \emptyset, N = N_A + N_{A^c}$). Let the parameter of interest be the relative size $\pi_A$ of the subpopulation $U_A$:

$$\pi_A = \frac{N_A}{N} \quad (1)$$

($\sum_U x_k$ is abbreviated notation for $\sum_{k \in U} x_k$).

Now each respondent of a simple random sample $s$ consisting of $n$ elements drawn without replacement has either to answer randomly

- with probability $p_1$ the question “Are you a member of group $U_A$?”,
- with probability $p_2$ the question “Are you a member of group $U_{A^c}$?” or
- with probability $p_3$ the question “Are you a member of group $U_B$?”

or is instructed just to say

- “yes” with probability $p_4$ or
- “no” with probability $p_5$

($\sum_{i=1}^{5} p_i = 1, 0 \leq p_i \leq 1$ for $i = 1, 2, ..., 5$). The $N_B$ Elements of group $U_B$ should be characterized by the possession of a completely innocuous attribute $B$ (for instance a season $B$ of birth), that should not be related to the possession or nonpossession of attribute $A$ (see: Horvitz, Shah and Sommons 1967). $\pi_B = N_B/N$ (with $0 < \pi_B < 1$) is the relative size of group $U_B$ in the population (Notice: Choosing an attribute
with relative size $\pi_B = 1$ or $0$ would mean nothing else than the instruction to answer "yes" or "no").

$\pi_B$ and the probabilities $p_1, p_2, ..., p_5$ are the design parameters of the standardized randomized response technique. For this strategy the probability $\pi_y$ of a "yes"-answer is

$$\pi_y = p_1 \cdot \pi_A + p_2 \cdot (1 - \pi_A) + p_3 \cdot \pi_B + p_4.$$  \hspace{1cm} (2)

Let

$$y_i = \begin{cases} 
1 & \text{if unit } i \text{ answers "yes"}, \\
0 & \text{otherwise}
\end{cases} (i = 1, 2, ..., n).$$

For $p_1 \neq p_2$ an unbiased estimator of $\pi_A$ is then given by

$$\hat{\pi}_A = \frac{\hat{\pi}_y - p_2 - p_3 \cdot \pi_B - p_4}{p_1 - p_2}$$  \hspace{1cm} (3)

with $\hat{\pi}_y = \sum_s y_k / n$, the proportion of "yes"-answers in the sample.

For simple random sampling without replacement (wor) the variance of the standardized estimator $\hat{\pi}_A$ (3) is given by

$$V_{wor}(\hat{\pi}_A) = \frac{\pi_y \cdot (1 - \pi_y)}{n \cdot (p_1 - p_2)^2} - \frac{\pi_A \cdot (1 - \pi_A)}{n} \cdot \frac{n - 1}{N - 1}$$  \hspace{1cm} (4)

(for a proof see: Quatember 2007).

Before we are able to look for the "variance-optimum" values of the design parameters of the standardized randomized response strategy we have to think about the level of privacy protection, which is offered by different choices of these parameters. The efficiency of different questioning designs can just be compared at the same level of this protection. The variance of the estimator gets smaller when the level of the individual’s privacy protection decreases, but if the variable under study is sensitive, at the same time the nonresponse rate increases. Therefore it would be desirable to find the optimum design parameters in such a way, that all respondents’ willingness to cooperate will just be guaranteed. Choosing a lower privacy protection would then automatically produce nonresponse and therefore set us back to the starting point of the problem.

For this reason it is necessary to measure the respondents’ privacy protection. The following ratios $\lambda_1$ and $\lambda_0$ of conditional probabilities give “a natural measure for the different levels of jeopardy” (Leysieffer and Warner 1976, p.650):

$$\lambda_1 = \frac{P(y_i = 1 | i \in U_A)}{P(y_i = 1 | i \in U_{A^c})}$$  \hspace{1cm} (5)

and

$$\lambda_0 = \frac{P(y_i = 0 | i \in U_{A^c})}{P(y_i = 0 | i \in U_A)}.$$  \hspace{1cm} (6)

The first term refers to the privacy protection with respect to a “yes”-answer, whereas the second refers to it with respect to a “no”-answer. For a totally protected privacy these measures are $\lambda_1 = \lambda_0 = 1$. The probability of responding “yes” (or “no”) on the selected question, for the case that the individual does possess
the attribute $A$, will then be the same as if he or she does not. This means that
the answer of the responding person would contain absolutely no information on the
subject under study. The more the Leysieffer-Warn er measures of privacy protection
differ from unity (in either direction), the more information about the characteristic
under study is contained in the answer on the selected question and the lower is
the individual’s protection against the interviewer. For the direct questioning de-
sign, offering absolutely no such protection to the respondent, these measures are
$$\lambda_1 = \lambda_0 = \infty \text{ (or zero)}.$$

Certainly the variance-optimum values $\lambda_{1,\text{opt}}$ and $\lambda_{0,\text{opt}}$ of these measures that we
assume to only just guarantee full response, will depend on the subject of interest,
meaning that the more sensitive a subject is, the closer to unity these values have
to be determined. So the practical experience of the user in this field together with
empirical studies will surely help to determine $\lambda_{1,\text{opt}}$ and $\lambda_{0,\text{opt}}$. But in contrast to
almost all presentations of randomized response techniques in the past as far as
the author knows them (see for example: Greenberg et al. 1969, p.526f, Mangat
and Singh 1990, p.440, Singh et al. 2003, 518f) we want to state that only in
combination with such measures it is possible to seriously compare the efficiency of
different choices of the design parameters of the standardized randomized response

2 Efficiency Comparisons

Without loss of generality let us assume subsequently, that we will choose the two
categories of the variable under study in such way, that the membership of $U_A$ is
at least as sensitive as the membership of $U_{Ac}$, which means that $\lambda_{1,\text{opt}}$ has to be
smaller than or at most equal to $\lambda_{0,\text{opt}}$. For the standardized questioning design the
Leysieffer-Warn er measures of privacy protection $\lambda_1$ and $\lambda_0$ are given by

$$\lambda_1 = \frac{p_1 + p_3 \cdot \pi_B + p_4}{p_2 + p_3 \cdot \pi_B + p_4}$$  \hspace{1cm} (7)

and

$$\lambda_0 = \frac{1 - (p_2 + p_3 \cdot \pi_B + p_4)}{1 - (p_1 + p_3 \cdot \pi_B + p_4)}.$$  \hspace{1cm} (8)

2.1 Case I: The Membership of Both $U_A$ and $U_{Ac}$ is Sensitive

Let $\lambda_{1,\text{opt}} < \infty$ and $\lambda_{0,\text{opt}} < \infty$, meaning that the membership of both $U_A$ and $U_{Ac}$
is sensitive, although not necessarily equally sensitive (for instance: $U_A \ldots$ set of
married people, who had at least one sexual intercourse with their partners last
week; $U_{Ac} = U - U_A$). From (7) and (8) this results in the equations

$$p_1 + p_3 \cdot \pi_B + p_4 = \frac{\lambda_{1,\text{opt}} \cdot \lambda_{0,\text{opt}} - \lambda_{1,\text{opt}}}{\lambda_{1,\text{opt}} \cdot \lambda_{0,\text{opt}} - 1}$$  \hspace{1cm} (9)

and

$$p_2 + p_3 \cdot \pi_B + p_4 = \frac{\lambda_{0,\text{opt}} - 1}{\lambda_{1,\text{opt}} \cdot \lambda_{0,\text{opt}} - 1}.$$  \hspace{1cm} (10)
Subtracting equation (10) from (9) gives the following condition, which has also to be fulfilled if the limits of privacy protection are to be applied:

\[ p_1 - p_2 = \frac{(\lambda_{1,\text{opt}} - 1) \cdot (\lambda_{0,\text{opt}} - 1)}{\lambda_{1,\text{opt}} \cdot \lambda_{0,\text{opt}} - 1}. \]  

(11)

Inserting (11) and (10) into (2) results in the following expression of the “optimum” probability of a “yes”-answer:

\[ \pi_{y,\text{opt}} = \frac{(\lambda_{1,\text{opt}} - 1) \cdot (\lambda_{0,\text{opt}} - 1)}{\lambda_{1,\text{opt}} \cdot \lambda_{0,\text{opt}} - 1} \cdot \pi_A + \frac{\lambda_{0,\text{opt}} - 1}{\lambda_{1,\text{opt}} \cdot \lambda_{0,\text{opt}} - 1} \cdot \pi_A. \]  

(12)

With this variance-optimum value of \( \pi_y \) we get with (4) the minimum variance of the estimation of \( \pi_A \), which can be achieved using the standardized strategy (\( p_1 \) and \( p_2 \) according to (11)):

\[ V_{\text{wor},\text{opt}}(\hat{\pi}_A) = \frac{\pi_{y,\text{opt}} \cdot (1 - \pi_{y,\text{opt}})}{n \cdot (p_1 - p_2)^2} - \frac{\pi_A \cdot (1 - \pi_A)}{n} \cdot \frac{n - 1}{N - 1}. \]  

(13)

### 2.2 Recommended Practice for Case I

Which of the special cases of Table 1 of Quatember (2007) can achieve the minimum variance and which values for the design parameters have to be chosen for this purpose? The direct questioning on the subject (we call this strategy \( \text{ST1} \)) and the other strategies in the following according to Quatember 2007) cannot be used if the subject of interest is sensitive, because it is assumed that this would lead to nonresponse. Furthermore the \( \text{ST4} \)-design cannot be used with a subject that is sensitive as a whole, because for this design \( \lambda_0 = \infty \). This means that a “no-answer” indicates with probability 1, that the respondent is a member of the subpopulation \( U_A \) and therefore on our assumption he or she will not respond on this sensitive question. The third case, that cannot be used, is \( \text{ST5} \). This questioning design consists of a question on membership of \( U_A \) and an instruction to reply “no”. In this case a “yes”-answer identifies the respondent with certainty as an owner of attribute \( A \) (\( \lambda_1 = \infty \)). Therefore this design like the direct questioning cannot be used at all, if the subject under study is sensitive.

The other designs, which are special cases of the standardized randomized response strategy, can be used for sensitive topics. It turns out that the \( \text{ST8} \)-design and – for \( \lambda_{1,\text{opt}} < \lambda_{0,\text{opt}} \) – also Warner’s design \( \text{ST2} \) are the only ones, that can not achieve the optimum efficiency. For Warner’s design this is caused by the fact, that it always protects the respondent’s privacy with respect to a “yes”-answer equally to the case of a “no”-answer. Both \( \lambda_1 \) and \( \lambda_0 \) are given by \( p_1/p_2 \). Therefore if \( \lambda_{1,\text{opt}} < \lambda_{0,\text{opt}} \) the optimum efficiency cannot be achieved by this strategy, because it protects a “no”-answer more than it would have to. On the other hand for the \( \text{ST8} \)-strategy the Leysieffer-Warner measure \( \lambda_1 \) is always greater than \( \lambda_0 \), because this design always protects a “no”-answer more than a “yes”-answer. Therefore for \( \lambda_{1,\text{opt}} \leq \lambda_{0,\text{opt}} < \infty \) we cannot choose the design parameters \( p_1, p_2 \) and \( p_5 \) of \( \text{ST8} \) in such a way, that it is possible to achieve the minimum variance given by (13).

But all others of the combinations can perform optimally, if the design parameters are chosen according to formulae (9) to (11). (For the optimum choices of
the design parameters the reader is referred to Table 1 here). This means, that if \( \lambda_{1,\text{opt}} = \lambda_{0,\text{opt}} \), there is not one randomized response technique that can perform better than Warner’s technique ST2 when we use the optimum design parameters \( p_1 \) and \( p_2 \) according to Table 1. This fact was not recognized in publications of “more efficient” strategies in the past. Greenberg et al.’s strategy (ST3) with known \( \pi_B \) has on the one hand the advantage over Warner’s design to be able to perform optimally also if \( \lambda_{1,\text{opt}} < \lambda_{0,\text{opt}} \). On the other hand, however, it has the disadvantage, that the size \( \pi_B \) of subpopulation \( U_B \) is completely predetermined, if we want to achieve the optimum efficiency: \( \pi_B = (\lambda_{0,\text{opt}} − 1)/(\lambda_{1,\text{opt}} + \lambda_{0,\text{opt}} − 2) \). This means in practice, that to achieve this goal we have to use an appropriate subpopulation, which is exactly of this size.

If \( \lambda_{1,\text{opt}} = \lambda_{0,\text{opt}} \) the design parameter \( \pi_B \) of ST6 is exactly 0.5 because of (9), (11) and \( p_3 = 1 − p_1 − p_2 \). If \( \lambda_{1,\text{opt}} < \lambda_{0,\text{opt}} \) we might start with any subpopulation \( U_B \), for which the relative size \((\lambda_{0,\text{opt}} − 1)/(\lambda_{1,\text{opt}} + \lambda_{0,\text{opt}} − 2) < \pi_B < 1 \) applies. This follows again from (9), (11) and \( p_3 = 1 − p_1 − p_2 \). The other design parameters of ST6 can then be derived (see: Table 1).

The special cases ST7, ST11 and ST14 of our standardized randomized response strategy do not make use of the question on membership of \( U_B \) and achieve the minimum variance as well, if we choose the design parameters according to Table 1. But for ST7 this is only valid for \( \lambda_{1,\text{opt}} < \lambda_{0,\text{opt}} \), which means that the membership of \( U_A \) has to be more (and not equally) sensitive than that of \( U_A^* \). If \( \lambda_{1,\text{opt}} = \lambda_{0,\text{opt}} \) the variance of this technique only converges to the minimum variance when the design parameters approach the variance-optimum design parameters of ST2 \( (p_1 → \lambda_{1,\text{opt}}/(\lambda_{1,\text{opt}} + 1), p_2 → 1/(\lambda_{1,\text{opt}} + 1), p_4 → 0) \). Therefore ST7 is the perfect supplement of ST2, for which the very opposite is true. Without any exception ST11 and ST14 are variance-minimum designs in the case of a subject, which is sensitive as a whole, if we choose the design parameters according to Table 1.

The other six strategies ST9, ST10, ST12, ST13, ST15 and ST16 are more complicated in their practical use, because in the randomization devices the question on membership of \( U_B \) is included. For this reason the problem of finding a subpopulation not related to the possession and nonpossession of attribute \( A \) and of appropriate size occurs again. But these six designs can also achieve the minimum variance. For design ST9 it is recommended to start with the search for a subset \( U_B \), for which the relative size \( \pi_B \) meets the condition \( 0 < \pi_B < (\lambda_{0,\text{opt}} − 1)/(\lambda_{1,\text{opt}} + \lambda_{0,\text{opt}} − 2) \). Using questioning design ST10 the subpopulation has to be of relative size \((\lambda_{0,\text{opt}} − 1)/(\lambda_{1,\text{opt}} + \lambda_{0,\text{opt}} − 2) < \pi_B < 1 \). The other optimum design parameters for both strategies can be calculated on the basis of \( \pi_B \). Obviously these techniques perfectly complement each other. Depending on the relative size of the desired subpopulation we can use one of these two techniques to achieve the maximum performance.

For questioning design ST12 a user has to start with the choice of the design parameters \( p_1 \) and \( p_2 \) according to Table 1. It is recommended to continue with the search for an adequate group \( U_B \), for which the relative size \( \pi_B \) is less than \([\lambda_{0,\text{opt}} − 1 − p_2 \cdot (\lambda_{1,\text{opt}} \cdot \lambda_{0,\text{opt}} − 1)]/[\lambda_{1,\text{opt}} + \lambda_{0,\text{opt}} − 2 − 2p_2 \cdot (\lambda_{1,\text{opt}} \cdot \lambda_{0,\text{opt}} − 1)]\), followed by the determination of \( p_3 \) and \( p_4 \). For strategy ST13 the adequate subpopulation \( U_B \) has to have a relative size greater than the upper bound of \( \pi_B \) for ST12. Therefore ST13 fits perfectly to ST12, so any subset \( U_B \) of the population can be used, if we
<table>
<thead>
<tr>
<th>Design</th>
<th>Variance-optimum design parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>ST1</td>
<td>not applicable</td>
</tr>
<tr>
<td>ST2 ((\lambda_{1,\text{opt}} = \lambda_{0,\text{opt}}))</td>
<td>(p_1 = \frac{\lambda_{1,\text{opt}}}{\lambda_{1,\text{opt}}+1}, p_2 = 1-p_1)</td>
</tr>
<tr>
<td>ST2 ((\lambda_{1,\text{opt}} &lt; \lambda_{0,\text{opt}}))</td>
<td>impossible to achieve the minimum variances (13) and (??)</td>
</tr>
<tr>
<td>ST3</td>
<td>(\pi_B = \frac{\lambda_{0,\text{opt}}-1}{\lambda_{1,\text{opt}}+\lambda_{0,\text{opt}}-2}, p_1 = \frac{(\lambda_{1,\text{opt}}-1)(\lambda_{0,\text{opt}}-1)}{\lambda_{1,\text{opt}}\lambda_{0,\text{opt}}-1}, p_3 = 1-p_1)</td>
</tr>
<tr>
<td>ST4</td>
<td>not applicable</td>
</tr>
<tr>
<td>ST5</td>
<td>not applicable</td>
</tr>
<tr>
<td>ST6 ((\lambda_{1,\text{opt}} = \lambda_{0,\text{opt}}))</td>
<td>(\pi_B = 0.5, p_1: \frac{\lambda_{1,\text{opt}}-1}{\lambda_{1,\text{opt}}+1} &lt; p_1 &lt; \frac{\lambda_{1,\text{opt}}}{\lambda_{1,\text{opt}}+1}, p_2 = \frac{\lambda_{1,\text{opt}}}{\lambda_{1,\text{opt}}+1}, p_3 = 1-p_1-p_2)</td>
</tr>
<tr>
<td>ST6 ((\lambda_{1,\text{opt}} &lt; \lambda_{0,\text{opt}}))</td>
<td>(\pi_B: \frac{\lambda_{0,\text{opt}}-1}{\lambda_{1,\text{opt}}+\lambda_{0,\text{opt}}-2} &lt; \pi_B &lt; 1, p_1 = \frac{(\lambda_{1,\text{opt}}-1)(\lambda_{0,\text{opt}}-1)}{\lambda_{1,\text{opt}}\lambda_{0,\text{opt}}-1} + \frac{(\lambda_{1,\text{opt}}-1)(\lambda_{0,\text{opt}}-1)(1-\pi_B)}{(\lambda_{1,\text{opt}}\lambda_{0,\text{opt}}-1)(2\pi_B-1)}, p_2 = \frac{\lambda_{0,\text{opt}}-1}{\lambda_{1,\text{opt}}\lambda_{0,\text{opt}}-1}, p_3 = 1-p_1-p_2)</td>
</tr>
<tr>
<td>ST7 ((\lambda_{1,\text{opt}} = \lambda_{0,\text{opt}}))</td>
<td>impossible to achieve the minimum variances (13) and (??)</td>
</tr>
<tr>
<td>ST7 ((\lambda_{1,\text{opt}} &lt; \lambda_{0,\text{opt}}))</td>
<td>(p_1 = \frac{\lambda_{1,\text{opt}}\lambda_{0,\text{opt}}-\lambda_{0,\text{opt}}}{\lambda_{1,\text{opt}}\lambda_{0,\text{opt}}-1}, p_2 = \frac{\lambda_{1,\text{opt}}-1}{\lambda_{1,\text{opt}}\lambda_{0,\text{opt}}-1}, p_4 = 1-p_1-p_2)</td>
</tr>
<tr>
<td>ST8</td>
<td>impossible to achieve the minimum variances (13) and (??)</td>
</tr>
<tr>
<td>ST9</td>
<td>(\pi_B: 0 &lt; \pi_B &lt; \frac{\lambda_{0,\text{opt}}-1}{\lambda_{1,\text{opt}}+\lambda_{0,\text{opt}}-2}, p_1 = \frac{(\lambda_{1,\text{opt}}-1)(\lambda_{0,\text{opt}}-1)}{\lambda_{1,\text{opt}}\lambda_{0,\text{opt}}-1}, p_3 = \frac{\lambda_{1,\text{opt}}-1}{\lambda_{1,\text{opt}}\lambda_{0,\text{opt}}-1}(1-\pi_B), p_4 = 1-p_1-p_3)</td>
</tr>
<tr>
<td>ST10</td>
<td>(\pi_B: \frac{\lambda_{0,\text{opt}}-1}{\lambda_{1,\text{opt}}+\lambda_{0,\text{opt}}-2} &lt; \pi_B &lt; 1, p_1 = \frac{(\lambda_{1,\text{opt}}-1)(\lambda_{0,\text{opt}}-1)}{\lambda_{1,\text{opt}}\lambda_{0,\text{opt}}-1}, p_3 = \frac{(\lambda_{1,\text{opt}}\lambda_{0,\text{opt}}-1)(1-\pi_B)}{\lambda_{0,\text{opt}}-1}, p_5 = 1-p_1-p_3)</td>
</tr>
<tr>
<td>ST11</td>
<td>(p_1 = \frac{(\lambda_{1,\text{opt}}-1)(\lambda_{0,\text{opt}}-1)}{\lambda_{1,\text{opt}}\lambda_{0,\text{opt}}-1}, p_4 = \frac{\lambda_{0,\text{opt}}-1}{\lambda_{1,\text{opt}}\lambda_{0,\text{opt}}-1}, p_5 = 1-p_1-p_4)</td>
</tr>
<tr>
<td>ST12</td>
<td>(p_1: \frac{\lambda_{1,\text{opt}}-1}{\lambda_{1,\text{opt}}\lambda_{0,\text{opt}}-1} &lt; p_1 &lt; \frac{\lambda_{0,\text{opt}}-1}{\lambda_{1,\text{opt}}\lambda_{0,\text{opt}}-1}, p_2 = \frac{\lambda_{1,\text{opt}}-1}{\lambda_{0,\text{opt}}-1}(1-\pi_B), p_4 = 1-\sum_{i=1}^{3} p_i)</td>
</tr>
<tr>
<td>ST13</td>
<td>(\pi_B: 0 &lt; \pi_B &lt; \frac{\lambda_{0,\text{opt}}-1}{\lambda_{1,\text{opt}}+\lambda_{0,\text{opt}}-2-2\pi_B(\lambda_{1,\text{opt}}\lambda_{0,\text{opt}}-1)}, p_3 = \frac{\lambda_{0,\text{opt}}-1-p_2(\lambda_{1,\text{opt}}\lambda_{0,\text{opt}}-1)}{\lambda_{1,\text{opt}}\lambda_{0,\text{opt}}-1}(1-\pi_B), p_5 = 1-\sum_{i=1}^{3} p_i)</td>
</tr>
<tr>
<td>ST14</td>
<td>(p_1: \frac{\lambda_{1,\text{opt}}-1}{\lambda_{1,\text{opt}}\lambda_{0,\text{opt}}-1} &lt; p_1 &lt; \frac{(\lambda_{1,\text{opt}}-1)(\lambda_{0,\text{opt}}-1)}{\lambda_{1,\text{opt}}\lambda_{0,\text{opt}}-1}, p_2 = \frac{\lambda_{1,\text{opt}}-1}{\lambda_{1,\text{opt}}\lambda_{0,\text{opt}}-1}, p_4 = \frac{\lambda_{0,\text{opt}}-1}{\lambda_{1,\text{opt}}\lambda_{0,\text{opt}}-1} - p_2), (p_5 = 1-p_1-p_2-p_4)</td>
</tr>
<tr>
<td>ST15</td>
<td>(\pi_B: 0 &lt; \pi_B &lt; 1, p_1 = \frac{(\lambda_{1,\text{opt}}-1)(\lambda_{0,\text{opt}}-1)}{\lambda_{1,\text{opt}}\lambda_{0,\text{opt}}-1}, p_3: 0 &lt; p_3 &lt; \frac{\lambda_{1,\text{opt}}-1}{(\lambda_{1,\text{opt}}\lambda_{0,\text{opt}}-1)(1-\pi_B)}, p_4 = \frac{\lambda_{0,\text{opt}}-1}{\lambda_{1,\text{opt}}\lambda_{0,\text{opt}}-1} - p_3 \cdot \pi_B), (p_5 = 1-p_1-p_3-p_4)</td>
</tr>
<tr>
<td>ST16</td>
<td>(\pi_B: 0 &lt; \pi_B &lt; 1, p_1: \frac{(\lambda_{1,\text{opt}}-1)(\lambda_{0,\text{opt}}-1)}{\lambda_{1,\text{opt}}\lambda_{0,\text{opt}}-1} &lt; p_1 &lt; \frac{(\lambda_{1,\text{opt}}\lambda_{0,\text{opt}}-1)}{\lambda_{1,\text{opt}}\lambda_{0,\text{opt}}-1}, p_2 = \frac{\lambda_{1,\text{opt}}\lambda_{0,\text{opt}}-1}{\lambda_{1,\text{opt}}\lambda_{0,\text{opt}}-1}, p_3: 0 &lt; p_3 &lt; \frac{\lambda_{0,\text{opt}}-1}{\lambda_{1,\text{opt}}\lambda_{0,\text{opt}}-1} - p_1), (p_4 = \frac{\lambda_{0,\text{opt}}-1}{\lambda_{1,\text{opt}}\lambda_{0,\text{opt}}-1} - p_2 - p_3 \cdot \pi_B), (p_5 = 1-\sum_{i=1}^{3} p_i)</td>
</tr>
</tbody>
</table>

Table 1: Design Parameters to Achieve the Minimum Variances for \(\lambda_{1,\text{opt}} < \infty\) and \(\lambda_{0,\text{opt}} < \infty\)
use either ST12 or ST13.

The special cases ST15 and ST16 of our standardized randomized response strategy can both be used with any subpopulation \( U_B \subset U \). Questioning design ST15 allows to start with the calculation of \( p_1 \). The probability \( p_3 \) has to be calculated according to the inequality given in Table 2. \( p_4 \) and \( p_5 \) can be calculated thereafter. For ST16 to be equally efficient one has just one more step to go through, because from \( p_1 \) according to \([\lambda_{1,\text{opt}} - 1] \cdot (\lambda_{0,\text{opt}} - 1]) / (\lambda_{1,\text{opt}} \cdot \lambda_{0,\text{opt}} - 1) < p_1 < (\lambda_{1,\text{opt}} \cdot \lambda_{0,\text{opt}} - \lambda_{0,\text{opt}}) / (\lambda_{1,\text{opt}} \cdot \lambda_{0,\text{opt}} - 1)\) the additional design parameter \( p_2 \) has to be calculated according to (11). The other steps are very similar to those of ST15.

**Example 1:** Let \( x \) be a binary variable, which is sensitive as a whole (like sexual behaviour). Let furthermore the membership of group \( U_A \) be equally sensitive to the one of group \( U_{A^c} \) and \( \lambda_{1,\text{opt}} = \lambda_{0,\text{opt}} = 4 \). This means that we allow the probability of a “yes”-answer (“no”-answer) to be at most four times higher given the membership of \( U_A \) \((U_{A^c}) \) than given the membership of \( U_{A^c} \) \((U_A) \). Let us further suppose that the subpopulation \( U_B \), we want to use, is of relative size \( \pi_B = 0.2 \) and let \( N = 1,000 \), \( n = 250 \) and \( \pi_A = 0.1 \).

Inserting \( \lambda_{1,\text{opt}} = \lambda_{0,\text{opt}} = 4 \) into (12) gives \( \pi_{y,\text{opt}} = 0.26 \). For sampling without replacement this means that the minimum achievable standard deviation for the estimation of \( \pi_A \) with the standardized randomized response design is \( 4.53 \cdot 10^{-2} \).

In Table 2 optimum design parameters for the special cases of the standardized randomized response technique can be found. All of these designs perform best. In the case of an infinite number of possibilities for these values \((ST6, ST9, ST10\) and \( ST12 \) to \( ST16) \) only one example is given.

Like many others Warner’s strategy can perform optimally because in our example \( \lambda_{1,\text{opt}} = \lambda_{0,\text{opt}} \). To achieve this goal the question on membership of \( U_A \) has to be asked with probability 0.8 and the alternative question on membership of \( U_{A^c} \) with the remaining probability of 0.2. Two techniques, of which the question “Are you a member of group \( U_B \)?” is part of the randomization device, cannot be used with the supposed subpopulation of relative size 0.2 \((ST3 \) and \( ST6) \). Anyhow, the designs \( ST10 \) and \( ST13 \) could be used just as their “twins” \( ST9 \) and \( ST12 \), because we can change the notations of subsets \( U_B \) and \( U_{B^c} \), so that \( \pi_B \) results in 0.8.

### 2.3 Case II: Only the Membership of \( U_A \) is Sensitive

Let \( \lambda_{1,\text{opt}} < \infty \) and \( \lambda_{0,\text{opt}} = \infty \), meaning that only the membership of \( U_A \), but not of \( U_{A^c} \) is sensitive (for instance: \( U_A \) ... set of drug users within the last month; \( U_{A^c} = U - U_A \)). For \( \lambda_{0,\text{opt}} \) to be able to reach infinity \( 1 - (p_1 + p_3 \cdot \pi_B + p_4) \) has to be zero and therefore (7) and (8) lead to equations

\[
p_1 + p_3 \cdot \pi_B + p_4 = 1 \tag{14}
\]

and

\[
p_2 + p_3 \cdot \pi_B + p_4 = \frac{1}{\lambda_{1,\text{opt}}} \tag{15}
\]

Subtracting equation (15) from (14) gives the following condition, which has to be kept:
Table 2: Design Parameters to Achieve the Minimum Variances in Example 1.

<table>
<thead>
<tr>
<th>Design</th>
<th>p_1</th>
<th>p_2</th>
<th>p_3</th>
<th>p_4</th>
<th>p_5</th>
<th>(\pi_B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ST1</td>
<td>\text{not achievable}</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ST2</td>
<td>0.8</td>
<td>0.2</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>ST3</td>
<td>\text{no optimum efficiency if } \pi_B = 0.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ST4</td>
<td>\text{not achievable}</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ST5</td>
<td>\text{not achievable}</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ST6</td>
<td>\text{no optimum efficiency if } \pi_B = 0.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ST7</td>
<td>\text{no optimum efficiency for } \lambda_{1,\text{opt}} = \lambda_{0,\text{opt}}</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ST8</td>
<td>\text{no optimum efficiency}</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ST9</td>
<td>0.6</td>
<td>–</td>
<td>0.25</td>
<td>0.15</td>
<td>–</td>
<td>0.2</td>
</tr>
<tr>
<td>ST10</td>
<td>0.6</td>
<td>–</td>
<td>0.25</td>
<td>–</td>
<td>0.15</td>
<td>0.8</td>
</tr>
<tr>
<td>ST11</td>
<td>0.6</td>
<td>–</td>
<td>–</td>
<td>0.2</td>
<td>0.2</td>
<td>–</td>
</tr>
<tr>
<td>ST12</td>
<td>0.7</td>
<td>0.1</td>
<td>0.125</td>
<td>0.075</td>
<td>–</td>
<td>0.2</td>
</tr>
<tr>
<td>ST13</td>
<td>0.7</td>
<td>0.1</td>
<td>0.125</td>
<td>–</td>
<td>0.075</td>
<td>0.8</td>
</tr>
<tr>
<td>ST14</td>
<td>0.7</td>
<td>0.1</td>
<td>–</td>
<td>0.1</td>
<td>0.1</td>
<td>–</td>
</tr>
<tr>
<td>ST15</td>
<td>0.6</td>
<td>–</td>
<td>0.2</td>
<td>0.16</td>
<td>0.04</td>
<td>0.2</td>
</tr>
<tr>
<td>ST16</td>
<td>0.7</td>
<td>0.1</td>
<td>0.05</td>
<td>0.09</td>
<td>0.06</td>
<td>0.2</td>
</tr>
</tbody>
</table>

\[ p_1 - p_2 = \frac{\lambda_{1,\text{opt}} - 1}{\lambda_{1,\text{opt}}}. \]  
\[ \pi_{y,\text{opt}} = \frac{\lambda_{1,\text{opt}} - 1}{\lambda_{1,\text{opt}}} \cdot \pi_A + \frac{1}{\lambda_{1,\text{opt}}}. \]

Finally inserting (17) into (13) we get the minimum achievable variance for the case where only the membership of \(U_A\) but not of \(U_{Ac}\) is sensitive.

### 2.4 Recommended Practice for Case II

Looking for those values of the design parameters, for which the standardized randomized response strategy can achieve the minimum variance and for which equations (14) to (16) hold, we do find that there is only one solution! The only questioning design, that is able to perform optimally, is ST4 – a strategy, which could not be used at all with a subject, that is sensitive as a whole (Sections 2.1 and 2.2). The design parameters of ST4, that result in the minimum variances are given by \(p_1 = (\lambda_{1,\text{opt}} - 1)/\lambda_{1,\text{opt}}\) and \(p_4 = 1 - p_1\). This means that with probability \((\lambda_{1,\text{opt}} - 1)/\lambda_{1,\text{opt}}\) a respondent is asked the question on membership of \(U_A\) and with the remaining probability he or she is instructed to say “yes”. In this way the interviewer is only able to conclude from a “no”-answer directly on the nonpossession of \(A\) but not from a “yes”-answer on the possession of this sensitive attribute. But as
the membership of $U_{A^c}$ is nonsensitive, this fact does not cause any additional item nonresponse problem into the survey.

Questioning designs $ST1$ and $ST5$ are not applicable for Case II, too.

Because in design $ST2$ $\lambda_0$ can only be as large as $\lambda_1$, in the case of a “no”-answer the privacy of the interviewee is protected more than necessary, if only the possession of attribute $A$ is sensitive. Therefore Warner’s procedure cannot be as efficient as $ST4$. In fact there is only one design, that performs even worse than $ST2$. This is $ST8$, because for this design $\lambda_1$ is always larger than $\lambda_0$ and therefore, in the case of a “no”-answer the individual’s privacy is protected even more than for Warner’s technique.

For all other procedures $\lambda_1 < \lambda_0 < \infty$ applies. This means, that these are able to protect a “no”-answer less than Warner’s design (and therefore are more efficient than $ST2$), but still more than necessary. For instance Greenberg et al.’s $ST3$-strategy performs the better, the closer the design parameters are to the design parameters of $ST4$ ($p_1 \rightarrow (\lambda_{1,\text{opt}} - 1)/\lambda_{1,\text{opt}}$, $p_3 \cdot \pi_B \rightarrow 1/\lambda_{1,\text{opt}}$). The minimum variance of (13) is the limit to which the variance of the estimation of $\pi_A$ converge in this case. This limit also applies to all other special cases of our standardized randomized response technique.

**Example 2:** Let $x$ be a variable, for which the membership of class $U_A$ and not of $U_{A^c}$ is sensitive. Let furthermore the limits of privacy protection, that just guarantee full response, be $\lambda_{1,\text{opt}} = 4$ and $\lambda_{0,\text{opt}} = \infty$. This means that we allow the probability of a “yes”-answer to be 1 given the membership of $U_A$ and the probability of such an answer to be 0.25 given the membership of $U_{A^c}$. Let us assume that the proposed subpopulation $U_B$ is of relative size $\pi_B = 0.2$ and let $N = 1,000$, $n = 250$ and $\pi_A = 0.1$.

Inserting $\lambda_{1,\text{opt}} = 4$ into (17) results in $\pi_{y,\text{opt}} = 0.325$ and the minimum standard deviation, that can be achieved regarding our assumptions, is for sampling without replacement given by $3.83 \cdot 10^{-2}$. As described above these minimum variance can only be achieved by using the questioning design $ST4$, in which either we ask the respondents the question “Are you a member of group $U_A$?” with probability 0.75 or we instruct the person just to say “yes” with the remaining probability of 0.25. The choice of the design parameters of Warner’s strategy, that deliver the best performance and keeps the condition $\lambda_1 \leq 4$ to guarantee full response at our assumptions leads us to $p_1 = 0.8$ and $p_2 = 0.2$. The estimator’s standard deviation for simple random sampling without replacement is $4.53 \cdot 10^{-2}$. The difference between this standard deviation and the minimum achievable one of strategy $ST4$ is a result of the unnecessary high level of privacy protection in the case of a “no”-answer, that is implicit in Warner’s questioning design.

For all other questioning designs not using the question on $U_B$ the following result holds: If we choose the design parameters close to those of $ST4$ the performance can converge to the best performance and therefore at least be more efficient than strategy $ST2$. For those containing the question on membership of $U_B$ the same applies as for Greenberg et al.’s strategy: For a wanted subpopulation with $\pi_B = 0.2$ (or 0.8, if we change the notations of $U_B$ and $U_{B^c}$) we can estimate $\pi_A$ more efficiently than $ST2$, but only less than $ST4$. For $ST3$ for instance choosing $p_1 = 0.706$, $p_3 = 0.294$ and $\pi_B = 0.8$ results in a standard deviation of $4.02 \cdot 10^{-2}$. With these
design parameters the Warner-Leysieffer measures of Greenberg et al.’s strategy are $\lambda_{1,\text{opt}} \approx 4$ and $\lambda_{0,\text{opt}} \approx 13$.

References


