Sociological Methods & Research
http://smr.sagepub.com/

Sensitive Questions in Online Surveys: Experimental Results for the Randomized Response Technique (RRT) and the Unmatched Count Technique (UCT)

Elisabeth Coutts and Ben Jann
Sociological Methods & Research 2011 40: 169
DOI: 10.1177/0049124110390768

Sociological Methods & Research
40(1) 169–193

DOI: 10.1177/0049124110390768
http://smr.sagepub.com

Sensitive Questions in Online Surveys: Experimental Results for the Randomized Response Technique (RRT) and the Unmatched Count Technique (UCT)

Elisabeth Coutts[1] and Ben Jann[2]

Abstract

Gaining valid answers to so-called sensitive questions is an age-old problem in survey research. Various techniques have been developed to guarantee anonymity and minimize the respondent's feelings of jeopardy. Two such techniques are the randomized response technique (RRT) and the unmatched count technique (UCT). In this study the authors evaluate the effectiveness of different implementations of the RRT (using a forced-response design) in a computer-assisted setting and also compare the use of the RRT to that of the UCT. The techniques are evaluated according to various quality criteria, such as the prevalence estimates they provide, the ease of their use, and respondent trust in the techniques. The results indicate that the RRTs are problematic with respect to several domains, such as the limited trust they inspire and nonresponse, and that the RRT

[1]ETH, Zurich, Switzerland
[2]University of Bern, Switzerland

Editor's Note: After a long and severe illness, Elisabeth Coutts died of cancer on August 5, 2009.

Corresponding Author:
Ben Jann, University of Bern, Institute of Sociology, Lerchenweg 36, CH-3000 Bern 9, Switzerland
Email: jann@soz.unibe.ch

estimates are unreliable due to a strong false no bias, especially for the more sensitive questions. The UCT, however, performed well compared to the RRTs on all the evaluated measures. The authors conclude that the UCT is a promising alternative to RRT in self-administered surveys and that future research should be directed toward evaluating and improving the technique.

Keywords

sensitive questions, online survey, randomized response technique, unmatched count technique, item count technique, methodological experiment

Introduction

Gaining valid answers to so-called sensitive questions, that is, questions pertaining to private, socially frowned upon, or illegal behavior, is an age-old problem in survey research. Some people might underreport such behavior while overreporting socially desirable behaviors (Barnett 1998; Lee 1993; Rasinski et al. 1999; Singer, von Thurn, and Miller 1995; Tourangeau, Rips, and Rasinski 2000; Tourangeau and Yan 2007). There is evidence that such bias stems from several sources, including the sensitivity of the topic being asked about, question format, data collection mode, respondent characteristics, and interviewer characteristics and behavior.

Researchers have tried to combat this presumed response bias, or systematic over- or underreporting depending on the desirability of the behavior in question, in a variety of ways. Several of these methods are geared toward providing the respondent greater perceived confidentiality. Various dejeopardizing techniques have been developed toward that end. Lee (1993) describes these as a variety of statistically based techniques designed to guarantee anonymity and minimize the respondent's feelings of jeopardy when asked to admit to a behavior that is stigmatized or incriminating. Two such techniques are the randomized response technique (RRT, introduced by Warner in 1965) and the unmatched count technique (UCT, also called the item count technique, the unmatched block design, or block total response, see Raghavarao and Federer 1979).

The Randomized Response Technique

The RRT has been implemented in various forms (Folsom et al. 1973; Fox and Tracy 1986; Greenberg et al. 1969; Kuk 1990; Scheers and Dayton

1987). These forms vary in their ease of implementation and the efficiency of the estimates they provide. However, all of them rely on the pairing of an unthreatening question with the sensitive question of interest. A randomizing device is used to determine whether the respondent will answer the sensitive question, an outcome known only to the respondent. For example, in a variant of Boruch's (1971) forced-response method, a respondent may be asked to flip a coin to determine whether to automatically answer yes (heads) or instead answer a sensitive question (tails). Since only the respondent knows whether he or she has flipped heads or tails, a yes answer cannot be interpreted as an admission of guilt. However, the proportion of the sample that has engaged in the behavior of interest can be calculated with knowledge of the properties of the randomizing device.

There are many indications in the literature that the RRT leads to more accurate estimates of the prevalence of socially undesirable behavior than asking the sensitive question directly. Mostly, the use of the RRT has resulted in increased reporting of sensitive or frowned-upon behaviors as varied as child abuse, drug use, abortion, employee theft, welfare fraud, and premature sign-offs on audits in comparison to the reporting of the same deeds in response to a direct question (Goodstadt and Gruson 1975; Lara et al. 2004; Reckers, Wheeler, and Wong-On-Wing 1997; Stem and Steinhorst 1984; Tracy and Fox 1981; van der Heijden et al. 2000; Zdep and Rhodes 1976). Of course, a higher reported prevalence does not necessarily imply more accurate measurement. However, the comparison of answers gained with the RRT with objective outside information on behavior indicates that while the RRT estimates of socially undesirable behavior are still too low relative to the actual prevalence, the RRT provides more accurate estimates than data gained through direct questioning (Tracy and Fox 1981; van der Heijden et al. 2000).

Although the RRT has been tested in various implementations and settings, most of our knowledge about the method comes from its use in face-to-face interviews. In general, the largest differences between estimates based on direct questioning and those based on the RRT have been obtained for very sensitive behaviors (Himmelfarb and Lickteig 1982; Lensvelt-Mulders et al. 2005) under nonanonymous survey conditions and in situations in which self-incrimination has potentially high costs. There are fewer differences between the two estimates for less sensitive topics and in situations in which respondents have reason to believe that their anonymity was guaranteed.

The Unmatched Count Technique

The UCT represents a similar approach, in that it does not allow the researcher to draw conclusions about respondent behavior on the basis of survey answers. With the UCT, the respondents are asked directly about their own sensitive behavior at the same time as they are asked about a number of neutral or socially desirable behaviors. Estimation of the prevalence of the sensitive behavior requires an estimate of the aggregate prevalence of the other behaviors. This method therefore usually requires two samples: a reference sample that answers questions only about unthreatening behaviors and a sample that answers a sensitive question in addition. For example, two lists of activities may be constructed. These lists are identical except for the fact that one list is longer by one behavior, namely, the sensitive behavior of interest. Respondents are asked to report only the number of activities in which they have participated, but not which ones. Subtracting the average number of behaviors in the reference group from the average number of behaviors in the sensitive-question group provides an estimate of the frequency of the sensitive behavior while preserving the anonymity of those in the sensitive-question group. For a more detailed description, see Droitcour et al. (1991) or Dalton, Wimbush, and Daily (1994).

Various studies point to the effectiveness of the UCT in providing higher estimates than direct questioning of such sensitive behaviors as employee misconduct, shoplifting, hate crime victimization, and risky sexual behaviors (Anderson et al. 2007; Dalton et al. 1994; LaBrie and Earleywine 2000; Recker Rayburn, Earleywine, and Davison 2003; Tsuchiya, Hirai, and Ono 2007; Wimbush and Dalton 1997; but also see Ahart and Sackett 2004, who found no effect). In general, the UCT seems to provide a benefit under similar conditions as the RRT: Less anonymous survey settings and more sensitive topics were associated with larger differences between prevalence estimates based on the UCT versus direct questioning. However, the UCT differs from the RRT in that no randomizing device is required. This presumably both increases respondent trust in the technique and makes it less effortful to use.

The Current Study

According to Lensvelt-Mulders et al. (2005:323), a ‘‘thorough look at the literature on RRTs reveals that 35 years of research have not led to a consensus or a description of best practices.’’ While many open issues remain about the use of RRTs in interviewer-administered modes, far less is known about the challenges peculiar to implementing the techniques in self-administered online surveys. For example, selecting a randomizing device that is likely to be handy to a respondent, trusted by the respondent, and used correctly in the absence of interviewer supervision appears critical to the RRT's success in a self-administered mode. Employing a randomizing device such as a coin may seem cumbersome to a respondent, who may forgo actually tossing a coin if no survey administrator is present. An electronic coin-toss simulator may be more convenient, but may also not be trusted by the respondent to deliver an outcome unknown to the survey administrator. This study therefore varies both the nature of the randomizing device employed and the amount of respondent control over that device in an effort to determine which device would work best in the context of a computer-administered survey. We decided to use only the forced-choice method described previously (Boruch 1971) in this exploratory study because of its high statistical efficiency. We are aware of only three studies comparing computer-administered direct questioning to a computer-administered RRT, and those studies also employed a forced-response method (one study additionally used the unrelated question method). In the first study, online respondents were more likely to admit to tax evasion using the RRT than direct questioning (Musch, Bröder, and Klauer 2001). The authors of the second study found some evidence for the benefit of using the RRT only for the behaviors that occurred with the highest frequency (Snijders and Weesie 2008). Furthermore, Holbrook and Krosnick (2010a) found that RRT performed *worse* than direct questioning for a socially desirable behavior (voting) for both the forced-response method and the unrelated question method.

We also compare the use of the RRT and the UCT. The instructions for the UCT are simpler, and all other things being equal, the UCT would therefore appear preferable to the RRT. However, we are aware of only two other studies directly comparing the use of the two techniques in a self-administered mode. The first study found that the two techniques produced similar estimates in a group administration context (in which a researcher was presumably present to answer questions about their rationale and use) (Wimbush and Dalton 1997). The second study found that UCT produced better estimates of socially desirable behavior than RRT (Holbrook and Krosnick, 2010a, 2010b). These comparisons were, however, limited to the size of the estimates. We extend the comparison to several other dimensions in an attempt to determine how difficult the techniques are to use in a context in which no survey practitioner is present to answer questions. More specifically, we examine respondents' assessments of ease of use and the degree to which they trust and believe they understand the various techniques. We do so in the nonanonymous self-administered context of an online access panel survey.[1]

Finally, the different techniques are applied to questions of varying sensitivity. Given results reported in other studies, we hypothesized that the greatest differences among the techniques would be seen when asking about the most sensitive topics.

Method

Measurement Techniques

In this study, various measurement techniques were employed to estimate the prevalence rates of six behaviors that varied in their degree of sensitivity. In addition to the baseline method of direct questioning (DQ), we implemented the unmatched count technique and five variants of a forced-choice randomized response technique. Given the number of techniques being tested on a sample of limited size, all the RRTs employed a forced-response design because of the efficiency of this technique. In all of these implementations, respondents were instructed to use a randomization device and then, depending on its outcome, either to answer the sensitive question truthfully or automatically provide a yes answer. The probabilities of being directed to answer the sensitive question or provide the yes answer were both one-half.

All RRT respondents were instructed to generate six randomizing-device outcomes before they viewed a screen with the sensitive questions (except in the case of RRT Variant 2, see the following). This procedure was intended to maximize compliance with the RRT instructions. All six sensitive questions were displayed on the same screen. Efforts were made to keep all instructions as simple and clear as possible. The following randomizing devices were used:

1. Manual coin toss: Respondents were instructed to get a coin, toss the coin six times, and note the results of those tosses (heads or tails) one after the other on a sheet of paper. After they had done so, a new screen appeared with the rules for answering the six sensitive questions (answer the first question honestly if the first result is heads; simply answer with yes if the result is tails; etc.) and with a detailed example. The sensitive questions followed on the same page. The basic rule (''Depending of the result of your *n*th coin toss, please answer the question either . . . ,'' with *n* as the question number) was again displayed below each question.

2. Electronic coin toss: Respondents were instructed to press a ''Toss Coin'' button that appeared next to each question and answer accordingly (answer honestly if the toss results in heads, simply answer yes if tails are tossed). Clicking on the button next to a question displayed the result of the toss (heads or tails) and the relevant instruction (''Answer the question honestly'' or ''Simply answer with 'yes'''). The buttons were programmed in such a way that the respondents could press them as many times as they liked to convince themselves that random results were being generated.[2] Results of a previous study employing an electronic coin toss indicate that many respondents prefer it to a manual one (Lensvelt-Mulders et al. 2006).
3. Banknote serial numbers: Respondents were instructed to get two Euro bills and write the last three digits of their serial numbers one after the other on a sheet of paper. Afterward, a new screen appeared with the rules for answering the six questions as a function of the parity of the numbers (answer the first question honestly if the first number in the list is even; simply answer with yes if the number is odd; etc.) and a detailed example. As in (1), the sensitive questions followed on the same page and the rules were repeated for each of these questions.
4. Telephone numbers: The same as (3), except that the respondents were instructed to use the last three digits of two telephone numbers of their choice.
5. Banknote serial numbers, with the option to use telephone numbers instead: Similar to (3), but with the instruction to use telephone numbers if no banknotes were available.

The unmatched count technique was implemented using six sets of statements, one set for each sensitive behavior. Each set contained five neutral statements and possibly also contained a statement on the sensitive behavior. In experimental Group 1, the sensitive behaviors were omitted from statement sets 1, 2, and 4; in Group 2, they were omitted from statement sets 3, 5, and 6. In other words, every respondent provided answers both for sets that contained sensitive statements and for sets that did not. This setup was intended to make the logic of the method as clear as possible to the survey participants. For every set, the respondents were instructed to provide the number of statements they would agree with and an example was given. Efforts were made to ensure that the list contained both relatively concrete and anonymous behaviors (see the online appendix at http://smr.sagepub.com/supplemental for a list of the UCT sets).[3]

All the RRT variants and the UCT were introduced with the following text: ''In order to ensure that your answers remain absolutely anonymous, we ask you to carry out the following procedure. In doing so, please adhere strictly to the instructions, otherwise the explanatory power of all the data collected will be compromised.'' Located between the instructions at the top of the screen and the sensitive questions, the RRTs contained an additional statement explaining that, ''Since we do not know the results of your coin tosses [or, e.g., depending on condition: ''the serial numbers of your bank notes''], we cannot know which kind of answer you provided. We can, however, calculate a frequency for the entire group containing all our respondents with the aid of probability calculus.'' For the UCT this statement was: ''We cannot know which of the individual statements apply to you. We can, however, calculate a frequency for the entire group containing all of our respondents with the aid of probability calculus.''[4]

Data Collection

The survey was implemented using the Unipark online research platform by Globalpark GmbH (see www.unipark.de). The respondents were recruited from the German ''Sozioland'' access panel by Respondi AG (see www.sozioland.de) between August 1 and September 30, 2007, with an e-mail invitation. Of the 10,000 invited panel members, 2,075 participated in the survey, yielding a response rate of 21%. The sample is by no means representative of either the general population or the Internet-using population. This point is not, however, critical as we are primarily interested in differences among experimental groups (Gosling et al. 2004; Reips 2002).[5] Female respondents were overrepresented (65 percent female, 35 percent male). Furthermore, the respondents were relatively young (about 60 percent were younger than age 30) and well educated (60 percent have some higher education) compared to the general population.

Participation in the study was nominally anonymous because we were not able to track response to the questionnaire. However, although we do not know the identities of the survey participants, respondents may have had reason to believe that such knowledge was possible. Their identities are known to the panel organizers and they may have feared that their answers would be connected to those identities by ''Sozioland.''

Respondents were randomly assigned to one of eight experimental groups once they had activated the questionnaire: a group for direct questioning, one group for each of the five RRT variants tested, and two groups for the unmatched count technique. The probability to be assigned to the first group was 30 percent. The other groups each had a probability of 10 percent. Table 1 provides an overview of the number of observations in each of the experimental groups. Since respondents were assigned to the experimental conditions on the fly at the time they started filling out the questionnaire, group sizes are subject to some random variation.

Respondents had been invited by e-mail to participate in a survey on ‘‘Security and Everyday Offenses.’’ The questionnaire began with a set of basic demographic questions, followed by some questions on the respondent’s living conditions and neighborhoods and an item battery measuring personality traits. Respondents then saw a text explaining that they were about to answer questions that some might consider personal and assuring that their responses would be treated confidentially.

The sensitive questions of interest, which addressed various illegal or frowned upon behaviors, were then posed using one of the seven techniques outlined previously. The questionnaire continued with some additional questions on the respondents’ attitudes toward the sensitive behaviors, for example whether committing the behavior would be all right and whether it would be uncomfortable for most people to admit to having done so. Respondents who had answered the sensitive questions using either an RRT or the UCT were also asked to evaluate how well they had understood the instructions for using the technique and whether the technique guaranteed the anonymity of their answers (see the Results section for the question wording). The median time required to complete the questionnaire was 6.9 minutes.

Sensitive Questions

Respondents were asked about six behaviors of varying sensitivity, each of which was likely to have been carried out in this sample with reasonable frequency. The questions on these behaviors were:

1. ‘‘Have you ever received too much change and knowingly kept it?’’ (Keeping too much change)
2. ‘‘Have you ever knowingly used public transportation without buying a ticket?’’ (Freeriding)
3. ‘‘Have you ever deliberately taken an article from a store without paying for it?’’ (Shoplifting)
4. ‘‘Have you used marijuana in the past month?’’ (Marijuana use)

Table 1. Sizes of the Experimental Groups

Group	*N*	%
Direct questioning	643	30.99
RRT: Manual coin toss	185	8.92
RRT: Electronic coin toss	201	9.69
RRT: Banknotes	194	9.35
RRT: Phone numbers	218	10.51
RRT: Banknotes or phone numbers	236	11.37
Unmatched count 1	210	10.12
Unmatched count 2	188	9.06
Total	2,075	100.00

Note: RRT = randomized response technique.

5. "Have you ever driven a car although your blood alcohol was almost certainly over the legal limit?" (Driving under influence, DUI)
6. "Have you ever cheated on your partner?" (Infidelity)

Results

We evaluated the techniques according to various quality criteria, namely, the estimates they provided, the ease of their use, and respondent trust they inspired. Before looking at the prevalence estimates obtained with the different measurement methods, we analyze the other indicators of how well the techniques worked in our survey. The indicators are (1) whether respondents felt they had understood the procedures, (2) whether respondents were convinced that the techniques guaranteed the anonymity of their answers, (3) the time required to read the instructions and answer the sensitive questions, and (4) the amount of nonresponse induced by the techniques. The results for these indicators are summarized in Table 2.

The instructions provided in the survey seem to have been clear to most respondents (first indicator). The proportion of respondents who believed that they had completely understood the instructions lies between 80 percent and 93 percent depending on method.[6] However, even a single respondent who does not understand the instructions is a potential source of bias. This is especially true for the RRTs, which require all respondents to exactly follow a relatively complex procedure. The understanding rates of around 80 percent to 85 percent for the manual coin toss, banknote, and telephone number RRTs are therefore clearly unsatisfactory. Significantly better rates

Table 2. Quality Measures for the Different Techniques

Experimental condition	N	Understood (%)	Trust (%)	Time (in seconds)	Nonresponse (%)
Direct questioning	643	n.a.	n.a.	28	0.0
RRT: Manual coin toss	185	85.7	21.1	175	4.9
RRT: Electronic coin toss	201	92.9	14.7	97	0.5
RRT: Banknotes	194	82.3	20.6	169	8.8
RRT: Phone numbers	218	84.5	18.4	159	6.4
RRT: Banknotes or phone numbers	236	79.5	22.3	166	5.5
Unmatched count	398	91.8	28.6	116	0.3

Note: Understood: percentage of respondents who felt they had completely understood the instructions; Trust: percentage of respondents who believed that the technique guaranteed the anonymity of their answers; Time: time spent reading the instructions and answering the six sensitive questions (median in seconds); Nonresponse: percentage of respondents who did not answer any of the six questions; RRT = randomized response technique.

are achieved for the electronic coin toss RRT (93 percent) and the unmatched count method (92 percent) (Fisher's exact $p < .001$ for the test against the combined manual RRTs in both cases).

A second important aspect determining the success of the different methods is how many respondents believed in the protection offered by the procedures. If respondents remain suspicious in using the methods, they might behave self-protectively and either provide biased answers or refuse to respond to the questions (Tourangeau and Yan 2007). Table 2 displays the proportion of respondents who believed that the technique they used guaranteed the anonymity of their answers.[7] The results are disillusioning, with trust rates ranging from 15 percent for the electronic coin toss RRT to 29 percent for the unmatched count technique. However, also note that an additional approximately 20 percent of the respondents indicated that they "did not think about" whether their anonymity would be protected (not shown).

The lower level of trust in the electronic coin toss implementation of the RRT compared to the other RRTs makes sense because, technically, the outcomes of the electronic randomization device could have been tracked and stored on the project computer (the difference compared to the manual RRTs is significant on the 10 percent level; Fisher's exact $p = .07$). Furthermore, it is interesting to see that the trust rate was higher for the UCT than for the RRTs (statistically significant with $p < .01$ even if the electronic coin toss RRT is omitted). Our interpretation of this result is that the UCT instructions are easier to understand than the RRT instructions and that more complicated instructions make respondents more skeptical. This assertion is supported by the positive association between the ''understood'' and ''trust'' variables (phi $= 0.13$, $p < .001$).

Table 2 also contains information on the time required to read the instructions and answer the sensitive questions as a function of the measurement technique used (our third indicator). Median times are reported (the median is preferred here over the arithmetic mean because there are large outliers, probably due to interruptions while completing the questionnaire; however, using the mean does not alter our conclusions). Clearly, direct questioning is the fastest method, with a median response time of 28 seconds for the six questions (about 5 seconds per question). Answer times increase by a factor of 5 to 6 in the case of the manual RRTs because respondents have to get paper and pen and possibly a coin or banknotes. As expected, the electronic coin toss RRT is faster (97 seconds) than the manual RRTs (167 seconds), as is the unmatched count technique (116 seconds; $p < .001$ for both differences using a Mann-Whitney U-test), but the answer times are still inflated by a factor of 3.5 to 4 compared to the direct questions.

Finally, Table 2 reports nonresponse rates for the different techniques. Nonresponse is measured here as the proportion of respondents who did not answer any of the six questions after having read the instructions (including respondents who answered some, but not all six questions somewhat increases the nonresponse rates but does not alter the pattern observed). The results are very clear: The methods that require respondents to engage in a mode shift (i.e., take a pen and paper, toss a coin, etc.) induce more nonresponse. The amount of nonresponse for these RRTs was between 5 percent and 9 percent (6.5 percent to 12 percent if respondents with at least one unanswered question are counted), whereas it was virtually nonexistent for the other methods (direct questions, electronic coin toss RRT, and UCT).

To summarize, the manual RRTs (manual coin toss, banknotes, and telephone numbers) were problematic with respect to several aspects of their use. Many respondents did not understand the procedures, and both answer times and levels of nonresponse were considerable. The electronic coin toss RRT, although easier to use and better understood by the respondents, induced less trust because, in principle, it would have been possible for the researchers to find out whether the innocuous or the sensitive question had been answered. However, the unmatched count technique performed well compared to the RRTs on all of these measures.

We now turn to an analysis of the prevalence estimates for the sensitive behaviors. The estimates based on the different techniques are listed in Table 3. The true rates of the sensitive behaviors are unknown for this sample, so we cannot say which method provided the most valid results. There are nevertheless several interesting conclusions that can be drawn from the results. For example, it is immediately evident from Table 3 that the RRTs must have been used incorrectly by at least some respondents, since strongly negative estimates are observed (in brackets).[8] These negative estimates indicate that survey participants were responding with a no when proper use of the technique would have meant answering with an automatic yes. Apparently, some respondents were reluctant to give an automatic yes answer, possibly because they feared that it could falsely be construed as an admission of guilt (Edgell, Himmelfarb, and Duchan 1982; Krumpal 2008; Lensvelt-Mulders and Boeije 2007; Nathan and Sirken 1988). The RRTs therefore strongly underestimate the rates of the behaviors in question. The combined RRT estimates are significantly lower than the estimates obtained by direct questioning with $p < .001$ (two-sided z tests[9]) for all items except the first two (keeping too much change and freeriding), which have low question sensitivity.

Assuming that the rate of affirmative answers in the ‘‘answer the sensitive question’’ condition of the RRT is at least as high as the number of yes answers to the direct questions (which seems reasonable for these items), we can compute an estimate for the lower bound of the proportion of respondents who answered no although they were instructed to give an automatic yes answer. In our RRT designs, the expected value for the proportion of observed yes answers can be written as

$$\lambda = q \cdot \pi_x + (1 - q) \cdot \pi_y, \tag{1}$$

where q is the probability of being directed to the sensitive question ($q = 0.5$ in our design), π_x is the (unknown) probability of answering yes to the sensitive question, and π_y is the probability of answering yes to the innocuous question. In our case, π_y equals 1 because the ‘‘innocuous question’’ is a direct yes response. Solving the equation and substituting in the observed yes proportion $\hat{\lambda}$ (and setting $q = 0.5$ and $\pi_y = 1$), we obtain an estimate for π_x, namely:[10]

$$\hat{\pi}_x = \frac{1}{q}\left(\hat{\lambda} - (1 - q) \cdot \pi_y\right) = 2\left(\hat{\lambda} - 0.5\right). \tag{2}$$

These estimates are reported in Table 3. We can slightly modify equation (1) and rewrite it as

Table 3. Prevalence Estimates Based on the Different Techniques (%)

Experimental condition	Keeping too much change	Freeriding	Shoplifting	Marijuana use	DUI	Infidelity
Direct questioning	56.1	61.8	23.4	4.7	29.0	26.2
SE	2.0	1.9	1.7	0.8	1.8	1.7
RRT: Manual coin toss	61.4	45.1	4.0	[–31.4]	5.7	9.1
SE	6.0	6.7	7.6	7.2	7.5	7.5
RRT: Electronic coin toss	59.0	67.8	22.0	[–7.0]	8.0	20.0
SE	5.7	5.2	6.9	7.1	7.0	6.9
RRT: Banknotes	58.2	54.3	[–4.0]	[–44.5]	1.1	[–4.5]
SE	6.1	6.3	7.6	6.8	7.5	7.5
RRT: Phone numbers	59.6	59.6	15.8	[–38.7]	2.5	[–3.0]
SE	5.6	5.6	6.9	6.5	7.1	7.1
RRT: Banknotes or phone numbers	54.1	55.2	6.3	[–35.1]	[–6.3]	0.5
SE	5.6	5.6	6.7	6.3	6.7	6.7
Unmatched count	43.5	76.5	17.5	32.5	19.0	35.9
SE	11.1	10.2	10.3	11.3	9.2	9.0
RRTs combined	58.3	56.7	9.2	[–31.1]	1.9	4.4
SE	2.6	2.6	3.2	3.1	3.2	3.2
Lower bound for proportion of false no answers in RRTs	0.0	5.1	14.2	35.7	27.0	21.8

Note: Keeping too much change: whether respondent once received too much change and knowingly kept it; Freeriding: whether respondent once knowingly used public transportation without buying a ticket; Shoplifting: whether respondent once deliberately took an article from a store without paying for it; Marijuana use: whether respondent used marijuana within the past month; DUI: whether respondent once drove a car although the blood alcohol was almost certainly over the legal limit; Infidelity: whether respondent once cheated on a partner; RRT = randomized response technique.

$$\lambda = q \cdot \pi_x + (1 - q) \cdot (1 - \gamma) \cdot \pi_y, \tag{3}$$

where γ is the probability of the respondent disregarding the instructions and giving a no answer although an automatic yes answer would have been indicated according to the RRT instructions. Ideally, if all respondents follow the instructions, γ is zero. A proportion greater than zero is a real problem for RRT because it translates directly into the RRT prevalence estimate (the bias of the estimate is $-\gamma$ in our design).

If we substitute reasonable values for π_x, we can compute estimates for γ. In particular, if we assume that π_x is at least as high as the observed rate based on direct questions, denoted by $\hat{\pi}_x^{\mathrm{DQ}}$, the following relation holds for γ:

$$\hat{\gamma} \geq 1 - \frac{1}{(1-q) \cdot \pi_y} \cdot \left(\hat{\lambda} - q \cdot \hat{\pi}_x^{\mathrm{DQ}}\right) = 1 - 2\left(\hat{\lambda} - 0.5 \cdot \hat{\pi}_x^{\mathrm{DQ}}\right). \tag{4}$$

The right-hand side of equation (4) is simply the difference between the prevalence estimate based on direct questions and the RRT prevalence estimate in our design. The last row of Table 3 contains these lower bound estimates for the proportion of false no answers.[11] It is evident that for at least some of the sensitive questions, the proportion must have been high. Furthermore, the proportion of false no answers is lowest for the apparently least sensitive items (see Table 4). This heavily biases the RRT estimates, but the exact amount of bias remains unknown.[12]

While the RRT estimates seem to be unreliable due to strong false no biases,[13] the unmatched count technique provides more reasonable estimates (the UCT estimates are computed as the mean differences between the counts for the two experimental groups).[14] It seems noteworthy that in the case of marijuana consumption, the UCT yields a much higher estimate than the direct question ($p < .05$, two-sided z test; no significant differences exist between the UCT estimates and the estimates from direct questioning for the other items). This makes sense given that the marijuana question is the only question that refers to current behavior (''within the past month'' as opposed to ''ever''). However, this result should not be overinterpreted since standard errors are large for the UCT. For the marijuana question, the UCT standard error is 11.3, so the 95 percent confidence interval for the prevalence estimate ranges between 10 percent and 55 percent. The high sampling variance makes interpretation of the other prevalence estimates difficult as well, but overall the UCT estimates seem to have more validity than the RRT estimates: The UCT estimates are higher than the RRT estimates in most cases (significant for marijuana use and infidelity; $p < .01$, two-sided z test), and no significant negative deviations can be observed compared to the direct questions.

Conclusions

Our results indicate that the unmatched count technique is superior to any of the randomized response techniques implemented in the study along several dimensions. The procedure's instructions were generally better understood

Table 4. Perceived Question Sensitivity

	Keeping too much change	Freeriding	Shoplifting	Marijuana use	DUI	Infidelity
Percentage agreeing with "It is all right to . . ."	43.0	24.7	3.1	33.2	2.5	12.7
Percentage believing that it would be uncomfortable for most people to admit that they had done	27.8	25.7	81.2	56.3	53.8	82.8
Total sensitivity score	20.4	22.0	79.2	42.6	52.7	72.8

Note: Total sensitivity score is calculated as the proportion of respondents who think that the behavior is not all right and that admitting it would be uncomfortable for most.

and more respondents believed that the technique guaranteed the anonymity of their answers. Furthermore, response times were shorter than for most RRT variants and nonresponse was almost nonexistent, compared to rates of up to 9 percent for RRT. Also, the prevalence estimates obtained by the UCT did not suffer from the strong negative biases observed for the RRT estimates. Results indicating that the UCT may perform better than the RRT have also been provided very recently in the context of socially desirable behavior by Holbrook and Krosnick (2010a, 2010b). A drawback of the UCT is, however, the high sampling variance. UCT estimates are relatively inefficient compared to forced-choice RRT estimates, so that large samples are required to obtain precise estimates.

The often negative prevalence estimates obtained with the various versions of the RRT indicate that noncompliance with RRT instructions was frequent in our study. Similar results have been reported in other studies, especially those in which a forced-choice method directs respondents to provide an automatic yes answer (Lensvelt-Mulders and Boeije 2007; Musch et al. 2001; Snijders and Weesie 2008; van der Heijden et al. 2000). The amount of noncompliance with instructions for forced-choice techniques has been found to increase with the sensitivity of the question (Edgell et al. 1982). Respondents seem to feel as if they are being asked to answer the sensitive question with yes, versus simply being asked to answer in accordance with the outcome of a randomizing device. Edgell and coauthors (1982:97) report that, ''Despite [a] favorable endorsement of the randomizing device and the RRT procedure, some subjects indicated that they did not like to be directed to give embarrassing answers,'' results echoed in Krumpal (2008). There are two ways to address the reluctance to provide an RRT-directed answer that are at odds with the answer one would like to give. One is in using a (potentially less efficient or valid) RRT implementation, for instance one that poses an alternate question to which a yes response is innocuous or allowing respondents to choose direct questioning over RRT (Chaudhuri and Saha 2005). Another is to address the respondents' reluctance directly. As Lensvelt-Mulders and Boeije (2007:604) write, ''to avoid cheating in a forced response questionnaire, it is necessary to acknowledge the fact that being forced to answer contrary to one's own truth is difficult and sometimes even painful.'' Although both approaches are likely to reduce noncompliance, it remains to be seen whether sufficient compliance levels are reached for use in a self-administered setting.

Interestingly, the RRT method that provided the highest prevalence estimates was the electronic coin toss method, in which the outcome of the randomizing device was computer generated. Respondents reported less trust in the technique than in the other RRT techniques, but apparently adhered more closely to the RRT instructions when using it. The same thought that may have led to the lack of trust they expressed, that is, that the outcome of the electronic coin flip could be recorded and used to determine which question had been answered, also seemed to have disciplined respondents into giving an automatic yes answer when it was called for. That said, the electronic coin toss RRT method did not provide significantly higher prevalence estimates than direct questioning for any item and often provided lower estimates, so we may surmise that few participants used it to report true transgressions if they were directed to answer the sensitive question honestly. Despite the higher prevalence derived with the electronic coin-toss method than with the other RRT methods, it does not seem to be a useful method for this setting because it provides no benefit over direct questioning in our study.[15]

In contrast to the electronic RRT, many respondents seem not to have complied with the other RRT methods in our study, not providing an automatic yes answer when directed to do so. Furthermore, while the necessity of time-consuming random outcome generation in RRT imposes strong costs on the respondent (as evinced by the high reaction times), the low level of understanding of the technique might lead respondents to ask why they should go to the greater effort required to answer the RRT questions. The result is higher nonresponse for RRTs that make use of a manual randomization device. These results are a strong argument against using the RRT in situations in which its logic cannot be explained in detail or the actual use of the randomizing device monitored. We interpret the aforementioned results as indicating that while respondents prefer a more convenient and less time-consuming version of the RRT and seem more likely to provide an automatic yes answer when using it, work remains to be done on determining the kinds of RRT implementations and items for which respondents will be willing to provide a true yes answer in an online setting.

Based on our results, the UCT seems a more promising approach for a self-administered setting. However, the superiority of the UCT might not apply to all implementations of the RRT, since only the forced-choice one was tested in this study. Also, much work remains to be done on determining optimal implementations of the UCT. For example, there is some debate on the optimal prevalence of the nontarget items (Droitcour et al. 1991; Tsuchiya et al. 2007). It is clear that anonymity will be compromised if all of the nonsensitive items are either generally agreed with or generally not agreed with.[16] On the other hand, statistical efficiency increases if the nonsensitive items have prevalence rates close to zero or one. Another direction for further research is the determination of an optimal length of item lists. All other things being equal, longer lists offer more protection to the respondent, but this gain is potentially offset by the memory load longer lists impose and by the increased variance. It remains to be seen if the actual protection provided by longer lists translates into a sufficient increase in the perceived protection to justify the reduction in accuracy and efficiency (Tsuchiya et al. 2007). Furthermore, investigation of modifications of the UCT that would increase statistical efficiency without compromising the advantages of the technique would be desirable.[17]

It is also worth noting that models have been developed to allow analysis of both RRT and UCT data with covariates. The models for the RRT have been described elsewhere (e.g., Maddala 1983; Scheers and Dayton 1988; van den Hout and Kooiman 2006; an implementation is provided by Jann 2005). Analysis of the UCT data with covariates can be conducted by regressing the response variable on the covariates, a treatment variable (0 = the set does not include the sensitive item, 1 = the set includes the sensitive item), and the interactions between the covariates and the treatment variable. The interactions represent the effects of interest (for similar suggestions see Holbrook and Krosnick, 2010b). However, as noted earlier, rather large sample sizes will be required for these analyses. This remains the main impediment to effective use of the technique.

Acknowledgments

We thank Jürg Schellenberg and Fabian Schenkel, who helped develop the questionnaire and program the survey, and Stephanie Eckman, Frauke Kreuter, Werner Pitsch, and anonymous reviewers for their helpful comments. We also thank Globalpark GmbH and Respondi AG for providing free access to the ''Sozioland'' respondent pool.

Declaration of Conflicting Interests

The author(s) declared no conflicts of interest with respect to the authorship and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research and/or authorship of this article: This research was supported by the German Research Foundation (DFG) Priority Program Number 1292 on ''Survey Methodology.''

Notes

1. See Postoaca (2006) for a definition of and more information about an online access panel.
2. Of course, this may also have had the effect of tempting respondents to click on the coin toss button until they got a result they wanted.
3. The choice of unmatched count technique (UCT) statements can affect the usefulness of the technique. For example, making statements about traits or behaviors that can easily be verified for an individual respondent (e.g., birth month for participants in an online panel) reduces the anonymity provided by the technique. Also, asking about attitudes, traits, or behaviors that leave the respondent with a large amount of freedom to decide on a yes versus a no answer (e.g., ''Do you like . . . ?'') seems somewhat dangerous. Whether to answer yes to such a question may be decided as a function of how many other questions have been answered with yes. Finally, lists likely to elicit all or no yes answers to innocuous statements are also clearly undesirable because they reduce the anonymity that the UCT provides.
4. All instructions and questions texts have been translated from the original German.
5. That is, we are primarily interested in the internal validity of the measurement techniques. External validity, however, may be compromised given our selective sample. A limitation of our study therefore is that we cannot directly generalize our results to the general population or the Internet-using population without making an assumption about the techniques working equally well for different subpopulations that are under- or overrepresented in our sample.
6. The question (translated from German) was: ''In this survey, we used a special technique to ask you questions about some personal topics. Do you feel that you completely understood the instructions provided for the method?'' Possible

answers were yes, no, and don't know. Table 2 reports the proportion of yes answers among all answers.

7. The question (translated from German) was: ''Do you believe that, as we explained, we cannot derive the answers to these questions due to the use of this special method?'' Possible answers were yes, no, cannot say, and have not thought about it. Table 2 reports the proportion of yes answers among all answers.
8. A rate, of course, cannot be negative. However, if the true prevalence is close to zero, the randomized response technique (RRT) can occasionally result in negative estimates due to random variation, even if the procedure is correctly applied by all respondents.
9. The z values are computed as $(\hat{\pi}_1 - \hat{\pi}_2)/(SE_1^2 + SE_2^2)^{1/2}$ where $\hat{\pi}_1$ and $\hat{\pi}_2$ are the estimates of interest and SE_1 and SE_2 are their standard errors (as reported in Table 3). Because the groups are independent, there is no covariance term to be taken into account. The same formula is used for the comparison with the UCT estimates that follow.
10. With a standard error equal to $(q^{-2}\hat{\lambda}(1 - \hat{\lambda})/N)^{1/2}$, where N is the sample size.
11. The boundary estimates are equivalent to the maximum likelihood estimate for the proportion of cheaters proposed by Clark and Desharnais (1998).
12. Note that we can compute the absolute minimum for γ by setting π_x to zero, which yields $1 - 2\hat{\lambda}$ (and happens to be equal to the negative of the RRT prevalence estimate) in our design. This absolute minimum of false no answers is 31 percent for the marijuana question.
13. Interestingly, the electronic coin toss RRT seems to be the least biased. The estimates from the electronic coin toss RRT are significantly higher with $p < .05$ (two-sided) than the combined estimates from the other RRTs for all items but ''Keeping too much change'' and ''Driving under influence.''
14. That is, the UCT prevalence estimate is $\hat{\pi} = \bar{x}_1 - \bar{x}_0$ where $\bar{x}_1$ and $\bar{x}_0$ are the sample means of the counts for the sensitive question group and the reference group without the sensitive item, respectively. For the standard errors we employ the usual formula for a two-sample mean difference with unequal variances, that is, $\mathrm{SE} = (s_1^2/N_1 + s_0^2/N_0)^{1/2}$ where s_1 and s_0 denote the sample standard deviations and N_1 and N_0 are the sample sizes for the two groups.
15. We are aware of two other recent studies that looked at the use of a forced-choice RRT with an electronic randomizing device by members of an online access panel. Snijders and Weesie (2008) also found occasionally negative RRT estimates with this design, although there were indications of a benefit of using this version of the RRT versus direct questioning for relatively prevalent frowned upon behaviors. In Lensvelt-Mulders et al.'s (2006) study, the RRT yielded low yet nonnegative prevalence estimates for all items (these RRT estimates were not compared with estimates obtained with direct questioning).
16. In our study, the probability that all nonsensitive statements applied to a respondent was very low. Only 2 percent or fewer respondents in the reference group

reported a count of five. Furthermore, at least one statement applied to 90 percent of the respondents or more.

17. Double lists as used by Droitcour et al. (1991) are a step in this direction. Another suggestion for improvement of the UCT is made by Chaudhuri and Christofides (2007). Given that dejeopardizing techniques such as the UCT and the RRT must be associated with a loss in statistical efficiency by design, further research should also be directed toward gaining a better understanding about when such techniques are to be applied. There is a trade-off between bias and efficiency, and in many cases direct questioning may turn out to be the most appropriate approach. Factors influencing the trade-off are the degree of question sensitivity and the perceived anonymity of the setting, and the best choice of technique may depend on characteristics of the population under study. Adaptive approaches that pick the most appropriate technique depending on individual context might also be a fruitful approach.

References

Ahart, Alison M. and Paul R. Sackett. 2004. ''A New Method of Examining Relationships Between Individual Difference Measures and Sensitive Behavior Criteria: Evaluating the Unmatched Count Technique.'' *Organizational Research Methods* 7:101-14.

Anderson, Drew A., Angela M. Simmons, Suzanne M. Milnes, and Mitchell Earleywine. 2007. ''Effect of Response Format on Endorsement of Eating Disordered Attitudes and Behaviors.'' *International Journal of Eating Disorders* 40:90-3.

Barnett, Julie. 1998. ''Sensitive Questions and Response Effects: An Evaluation.'' *Journal of Managerial Psychology* 13:63-76.

Boruch, Robert F. 1971. ''Assuring Confidentiality of Responses in Social Research: A Note on Strategies.'' *The American Sociologist* 6:308-11.

Chaudhuri, A. and A. Saha. 2005. ''Optional Versus Compulsory Randomized Response Techniques in Complex Surveys.'' *Journal of Statistical Planning and Inference* 135:516-27.

Chaudhuri, Arijit and Tasos C. Christofides. 2007. ''Item Count Technique in Estimating the Proportion of People With a Sensitive Feature.'' *Journal of Statistical Planning and Inference* 137:589-93.

Clark, Stephen J. and Robert A. Desharnais. 1998. ''Honest Answers to Embarrassing Questions: Detecting Cheaters in the Randomized Response Model.'' *Psychological Methods* 3:160-68.

Dalton, Dan R., James C. Wimbush, and Catherine M. Daily. 1994. ''Using the Unmatched Count Technique (UCT) to Estimate Base Rates for Sensitive Behavior.'' *Personnel Psychology* 47:817-28.

Droitcour, Judith, Rachael A. Caspar, Michael L. Hubbard, Teresa L. Parsely, Wendy Vischer, and Trena M. Ezzati. 1991. ''The Item Count Technique as a Method of Indirect Questioning: A Review of Its Development and a Case Study

Application.'' Pp. 185-210 in *Measurement Errors in Surveys*, edited by P. Biemer, R. M. Groves, L. E. Lyberg, N. A. Mathiowetz, and S. Sudman. New York: John Wiley.

Edgell, Stephen E., Samuel Himmelfarb, and Karen L. Duchan. 1982. ''Validity of Forced Responses in a Randomized-response Model.'' *Sociological Methods & Research* 11:89-100.

Folsom, Ralph E., Bernard G. Greenberg, Daniel G. Horvitz, and James R. Abernathy. 1973. ''The Two Alternate Questions Randomized Response Model for Human Surveys.'' *Journal of the American Statistical Association* 68:525-30.

Fox, James Alan and Paul E. Tracy. 1986. *Randomized Response: A Method for Sensitive Surveys*. London: Sage.

Goodstadt, Michael S. and Valerie Gruson. 1975. ''The Randomized Response Technique: A Test of Drug Use.'' *Journal of the American Statistical Association* 70:814-18.

Gosling, Samuel D., Simine Vazire, Sanjay Srivastava, and Oliver P. John. 2004. ''Should We Trust Web-based Studies? A Comparative Analysis of Six Preconceptions About Internet Questionnaires.'' *American Psychologist* 59:93-104.

Greenberg, Bernard G., Abdel-Latif A. Abul-Ela, Walt R. Simmons, and Daniel G. Horvitz. 1969. ''The Unrelated Question Randomized Response Model: Theoretical Framework.'' *Journal of the American Statistical Association* 64:520-39.

Himmelfarb, Samuel and Carl Lickteig. 1982. ''Social Desirability and the Randomized-response Technique.'' *Journal of Personality and Social Psychology* 43: 710-17.

Holbrook, Allyson L. and Jon A. Krosnick. 2010a. ''Measuring Voter Turnout by Using the Randomized Response Technique: Evidence Calling Into Question the Method's Validity.'' *Public Opinion Quarterly* 74:328-43.

Holbrook, Allyson L. and Jon A. Krosnick. 2010b. ''Social Desirability Bias in Voter Turnout Reports: Tests Using the Item Count Technique.'' *Public Opinion Quarterly* 74:37-67.

Jann, Ben. 2005. ''rrlogit: Stata Module to Estimate Logistic Regression for Randomized Response Data.'' Statistical Software Components S456203. Boston College Department of Economics. (http://ideas.repec.org/c/boc/bocode/s456203.html).

Krumpal, Ivar. 2008. ''Evaluation of the Effectiveness of the Randomized Response Technique and the Item Count Method in the Telephone Survey Mode.'' In *Proceedings of the 7th International Conference on Social Science Methodology, RC33—Logic and Methodology in Sociology (ISA)* [CD-ROM], edited by S. Balbi, G. Scepi, G. Russolillo, and A. Stawinoga. Napoli: Jovene Editore.

Kuk, Anthony Y. C. 1990. ''Asking Sensitive Questions Indirectly.'' *Biometrika* 77:436-38.

LaBrie, Joseph W. and Mitchell Earleywine. 2000. ''Sexual Risk Behaviors and Alcohol: Higher Base Rates Revealed Using the Unmatched-count Technique.'' *Journal of Sex Research* 37:321-26.

Lara, Diana, Jennifer Strickler, Claudia D. Olavarrieta, and Charlotte Ellertson. 2004. ''Measuring Induced Abortion in Mexico: A Comparison of Four Methodologies.'' *Sociological Methods & Research* 32:529-58.

Lee, Raymond M. 1993. *Doing Research on Sensitive Issues*. London: Sage.

Lensvelt-Mulders, Gerty J. L. M. and Hennie R. Boeije. 2007. ''Evaluating Compliance With a Computer Assisted Randomized Response Technique: A Qualitative Study Into the Origins of Lying and Cheating.'' *Computers in Human Behavior* 23:591-608.

Lensvelt-Mulders, Gerty J. L. M., Joop J. Hox, Peter G. M. van der Heijden, and Cora J. M. Maas. 2005. ''Meta-analysis of Randomized Response Research.'' *Sociological Methods & Research* 33:319-48.

Lensvelt-Mulders, Gerty J. L. M., Peter G. M. van der Heijden, Olav Laudy, and Ger van Gils. 2006. ''A Validation of a Computer-assisted Randomized Response Survey to Estimate the Prevalence of Fraud in Social Security.'' *Journal of the Royal Statistical Society Series A* 169:305-18.

Maddala, G. S. 1983. *Limited Dependent and Qualitative Variables in Econometrics*. Cambridge, UK: Cambridge University Press.

Musch, Jochen, Arndt Bröder, and Karl Christoph Klauer. 2001. ''Improving Survey Research on the World Wide Web Using the Randomized Response Technique.'' Pp. 179-92 in *Dimensions of Internet Science*, edited by U.-D. Reips and M. Bosnjak. Lengerich, Germany: Pabst Science Publishers.

Nathan, Gad and Monroe G. Sirken. 1988. ''Cognitive Aspects of Randomized Response.'' Pp. 173-78 in *Proceedings of the Section on Survey Research Methods*. Alexandria, VA: American Statistical Association.

Postoaca, Andrei. 2006. *The Anonymous Elect: Market Research Through Online Access Panels*. Berlin: Springer.

Raghavarao, Damaraju and Walter T. Federer. 1979. ''Block Total Response as an Alternative to the Randomized Response Method in Surveys.'' *Journal of the Royal Statistical Society Series B (Statistical Methodology)* 41:40-5.

Rasinski, K. A., G. B. Willis, A. K. Baldwin, W. C. Yeh, and L. Lee. 1999. ''Methods of Data Collection, Perceptions of Risks and Losses, and Motivation to Give Truthful Answers to Sensitive Survey Questions.'' *Applied Cognitive Psychology* 13:465-84.

Recker Rayburn, Nadine, Mitchell Earleywine, and Gerald C. Davison. 2003. ''Base Rates of Hate Crime Victimization Among College Students.'' *Journal of Interpersonal Violence* 18:1209-221.

Reckers, Philip M. J., Stephen W. Wheeler, and Bernard Wong-On-Wing. 1997. ''A Comparative Examination of Auditor Premature Sign-offs Using the Direct and the Randomized Response Methods.'' *Auditing: A Journal of Practice & Theory* 16:69-78.

Reips, Ulf-Dietrich. 2002. ''Standards for Internet-based Experimenting.'' *Experimental Psychology* 49:243-56.

Scheers, N. J. and C. Mitchell Dayton. 1987. ‘‘Improved Estimation of Academic Cheating Behavior Using the Randomized-response Technique.’’ *Research in Higher Education* 26:61-9.

Scheers, N. J. and C. Mitchell Dayton. 1988. ‘‘Covariate Randomized Response Models.’’ *Journal of the American Statistical Association* 83:969-74.

Singer, Eleanor, Dawn R. von Thurn, and Esther Miller. 1995. ‘‘Confidentiality Assurances and Response: A Quantitative Review of the Experimental Literature.’’ *Public Opinion Quarterly* 59:66-77.

Snijders, Chris and Jeroen Weesie. 2008. ‘‘The Online Use of Randomized Response Measurement.’’ Paper presented at General Online Research 2008, Hamburg, Germany.

Stem, Donald E. and R. Kirk Steinhorst. 1984. ‘‘Telephone Interview and Mail Questionnaire Applications of the Randomized Response Model.’’ *Journal of the American Statistical Association* 79:555-64.

Tourangeau, R., Lance J. Rips, and Kenneth Rasinski. 2000. *The Psychology of Survey Response*. Cambridge, UK: Cambridge University Press.

Tourangeau, Roger and Ting Yan. 2007. ‘‘Sensitive Questions in Surveys.’’ *Psychological Bulletin* 133:859-83.

Tracy, Paul E. and James Allan Fox. 1981. ‘‘The Validity of Randomized Response for Sensitive Measurements.’’ *American Sociological Review* 46:187-200.

Tsuchiya, Takahiro, Yoko Hirai, and Shiguro Ono. 2007. ‘‘A Study of the Properties of the Item Count Technique.’’ *Public Opinion Quarterly* 71:253-72.

van den Hout, Ardo and Peter Kooiman. 2006. ‘‘Estimating the Linear Regression Model With Categorical Covariates Subject to Randomized Response.’’ *Computational Statistics & Data Analysis* 50:3311-323.

van der Heijden, Peter G. M., Ger van Gils, Jan Bouts, and Joop J. Hox. 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 & Research* 28:505-37.

Warner, Stanley L. 1965. ‘‘Randomized-response: A Survey Technique for Eliminating Evasive Answer Bias.’’ *Journal of the American Statistical Association* 60:63-9.

Wimbush, James C. and Dan R. Dalton. 1997. ‘‘Base Rate for Employee Theft: Convergence of Multiple Methods.’’ *Journal of Applied Psychology* 82:756-63.

Zdep, S. M. and Isabelle N. Rhodes. 1976. ‘‘Making the Randomized Response Technique Work.’’ *Public Opinion Quarterly* 40:531-37.

Bios

Elisabeth Coutts was a graduate student at the Chair of Sociology at the ETH Zurich, Switzerland. Her research interests included social-science methodology, judgment and decision making, and scientific misconduct. Her most recent publication was a review of issues in Web survey methodology published in *Kölner Zeitschrift für*

Soziologie und Sozialpsychologie (with Mick Couper). Elisabeth Coutts died of cancer on August 5, 2009.

Ben Jann is assistant professor of sociology at the University of Bern, Switzerland. His research interests include social-science methodology, statistics, social stratification, and labor market sociology. Recent publications include ‘‘The Blinder-Oaxaca Decomposition for Linear Regression Models’’ in the *Stata Journal* and the German book *Erwerbsarbeit, Einkommen und Geschlecht* (Wiesbaden, 2008).