This article assesses the validity of responses to sensitive questions using four different methods. In an experimental setting, the authors compared a computer-assisted self-interview (CASI), face-to-face direct questioning, and two different varieties of randomized response. All respondents interviewed had been identified as having committed welfare and unemployment benefit fraud. The interviewers did not know that respondents had been caught for fraud, and the respondents did not know that the researchers had this information. The results are evaluated by comparing the percentage of false negatives. The authors also looked for variables that might explain why some respondents admit fraud and others do not. The proportions of respondents admitting fraud are relatively low, between 19 percent and 49 percent. The two randomized response conditions were superior in eliciting admissions of fraud. A number of background variables, notably gender, age, still receiving benefit, and duration and perception of fraud, are related to admitting fraud. Although the randomized response conditions performed much better than face-to-face direct questioning and CASI, the percentage of respondents admitting fraud is only around 50 percent. Some possible reasons for this are discussed.

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

PETER G. M. VAN DER HEIJDEN
Utrecht University

GER VAN GILS
BeleidsOnderzoek en Advies

JAN BOUTS
Nederlands Instituut voor de Publieke Opinie en het Marktonderzoek

JOOP J. HOX
Utrecht University

1. INTRODUCTION

In both the United States and Europe, the issue of the welfare system has received growing attention in recent years, and efficient fraud

SOCIOLOGICAL METHODS & RESEARCH, Vol. 28 No. 4, May 2000 505-537

control is generally seen as an integral component of the welfare system. Therefore, accurate information on the types and extent of welfare and unemployment benefit fraud is extremely relevant for policy decisions. However, it is difficult to obtain valid and reliable information in this area. If standard survey research is used to assess fraud, respondents will often refuse to take part, or, if they do take part, they will often not answer truthfully, especially when they have committed fraud.

The problem of obtaining valid and reliable information is not unique to fraud. Posing a direct question in an interview tends to give poor results (Sudman and Bradburn 1979, 1982; Lee 1993). A few examples from the Netherlands illustrate the point. Junger (1989, 1990) compared police records with the answers to a direct question in a face-to-face interview. She showed that Dutch adolescents report only 60 percent to 70 percent of all offenses for which they have ever been caught by the police. When attention was restricted to the most recent year, only 23 percent of all adolescents in her research admitted their offenses, with important differences according to ethnic background. For instance, respondents with a Turkish background reported only 9 percent of their offenses. In an investigation of unemployment benefit fraud using face-to-face direct questioning, Elffers, Robben, and Verlind (1989) found that 43 percent of those respondents who had been caught for fraud did not admit this. In telephone interviews with people already caught for vehicle tax fraud, Berghuis and Kommer (1982) found that only approximately 10 percent admitted evading this tax. Hessing, Elffers, and Wiegel (1988; see also Elffers, Robben, and Hessing 1992), using face-to-face direct questioning, report that 70 percent of their respondents denied ever having evaded income tax, while in fact all their respondents had been found guilty of this offense.

Survey methodologists have generally hypothesized that a major source of error in reports of sensitive behavior is deliberate misreporting (Jobe et al. 1997), and validity studies have indicated a pattern of underreporting for socially disapproved items (for a concise summary,

AUTHORS' NOTE: *We thank Dr. Edith de Leeuw and three anonymous reviewers for their comments. We gratefully acknowledge the support of the Department of Social Affairs.*

see Droitcour et al. 1991). Likewise, Dutch surveys on moonlighting are plagued by false answers. Moonlighters receiving unemployment benefit may be less inclined to cooperate with survey research on moonlighting than those who have a job, since the former group is more at risk. Apart from juridical punishment, moonlighters also risk losing part of their income (Van Eck and Kazimier 1990; see also Koopmans 1988). It is clear that posing a direct question about welfare and unemployment benefit fraud in an interview is of limited value.

To combat respondents' underreporting of socially disapproved behavior, survey methodologists have developed a number of different measurement procedures designed to ensure the confidentiality of the answers and reduce potential respondents' concerns on self-presentation. When surveying sensitive topics, researchers often used self-administered questionnaires, either in the form of a postal survey or as a questionnaire handed out by the interviewer. Compared with face-to-face interviews, self-administered questionnaires evoke a greater sense of privacy and lead to greater self-disclosure (Sudman and Bradburn 1979; Tourangeau and Smith 1996). Survey research has shown that compared with interviews, self-administered questionnaires produce more valid reports of sensitive behavior and fewer socially desirable answers in general (e.g., Aquilino 1994; Hochstim 1967; Siemiatycki 1979; Turner, Lessler, and Devore 1992; for a meta-analysis, see De Leeuw 1992).

However, self-administered questionnaires have a serious drawback: Only relatively simple questionnaires can be used (Dillman 1978). Computer-assisted self-interviewing (CASI) overcomes this problem and makes it possible to use very complex questionnaires without the aid of an interviewer.

In CASI, the interview program takes over the questionnaire logic and question flow. Respondents simply read each question from the screen and type in an answer, so they are no longer burdened with complex routing instructions. An interviewer may bring the computer to the respondent's home, or the respondent may be invited to a data collection site equipped with computers. Generally, an interviewer or fieldworker is present to assist at the start, but the respondents operate the computer on their own, which gives them the privacy of a self-administered questionnaire (De Leeuw, Hox, and Snijkers 1995; Nicholls, Baker, and Martin 1997).

The visible presence of a computer in CASI may have a special effect on the respondent's sense of privacy. De Leeuw et al. (1995) distinguish two different potential effects of the presence of a computer in the interviewing situation. The first effect is a feeling of less privacy. If one is totally unfamiliar with computers, there could be a Big Brother effect, leading to more refusals and socially desirable answers to sensitive questions. When researchers first began to use computer-assisted interviewing, there was considerable apprehension about this possible effect. Using a computer could also lead to the expectancy of more privacy by the respondents, since responses are typed directly into the computer and cannot be read by anyone who happens to find the questionnaire. In the Western world, where computers are widespread and familiar, this reaction is more likely than the Big Brother reaction, although much depends on the setting and the specific interview situation.

In general, empirical research indicates that using a computer appears to enhance the feeling of privacy. Once an answer is given, it disappears from the screen, whereas an answer written down remains on the paper for everyone to see. Beckenbach (1995) showed in an experimental field study that respondents evaluated sensitive questions as somewhat less threatening when CASI was used compared with traditional paper-and-pencil methods. Respondents were also more positive about data privacy and found that answering sensitive questions was less unpleasant when computer-assisted methods were used. Beckenbach (1992) also cites two well-controlled laboratory experiments that compare paper-and-pencil self-administered questionnaires with CASI. In the first, Martin and Nagao (1989) compared CASI with face-to-face interviewing and with a paper-and-pencil self-administered questionnaire. Using the Crowne-Marlowe social desirability scale, they found less social desirability bias in the paper-and-pencil self-administered questionnaire (and even less in CASI) than in the interview. In the second experiment, Evan and Miller (1969) compared CASI with a paper-and-pencil self-administered questionnaire. They found that CASI leads to more openness with questions that are perceived as threatening, whereas no differences were found for nonthreatening questions. In a comparison of a postal (paper) questionnaire and an e-mail health questionnaire, Kiessler and

Sproull (1986) found fewer socially desirable answers in the electronic version; similar results were found by Johnston and Walton (1995). Finally, an extensive meta-analysis of 39 studies by Weisband and Kiessler (1996) reports that in general, computer administration increases self-disclosure when compared with paper-and-pencil (self-administered) forms.

Researchers have recently started to use interview programs that ask the questions through a headphone set. This provides greater privacy protection and makes it possible to use CASI for samples in which illiteracy may be a problem (Tourangeau and Smith 1998; Turner et al. 1998).

The randomized response (RR) technique is an entirely different approach to obtaining valid answers to sensitive questions. In RR, a randomizing technique is used to hide the answer given by the respondent from the interviewer and the researcher. RR was first proposed by Warner (1965), and different versions have been developed. We will describe RR techniques in more detail in the next section. Comparative research shows that RR techniques generally lead to more valid answers (Umesh and Peterson 1991; Hosseini and Armacost 1993; Scheers 1992). Locander, Sudman, and Bradburn (1976) compared face-to-face interviews, telephone interviews, and mail surveys with RR for a variety of sensitive questions. These authors conclude that RR is a promising technique, but none of the methods they studied was clearly superior to the others for all sensitive questions used.

We compare two different varieties of RR, CASI, and face-to-face direct questioning in an experimental setting. A review of such research is given by Umesh and Peterson (1991), who remark that most often RR reveals a higher proportion of respondents who report the sensitive characteristic under study. However, Umesh and Peterson correctly argue that finding a higher proportion is insufficient proof of the validity of the RR technique. It is possible that although there is a higher proportion, there still are many false negatives, that is, people who do not admit the sensitive characteristic. Umesh and Peterson therefore propose individual validation studies to investigate the validity of the RR technique. In an individual validation study, respondents are interviewed about a sensitive characteristic for which their status is known to the researcher to check whether those respondents answer truthfully.

All respondents used in our study had in fact been caught for welfare or unemployment benefit fraud. The experiment was set up in such a way that the interviewers did not know that respondents had been caught for fraud, and respondents did not know that the researchers had this information. Because the status of the respondents is known, it is possible to compare the results of the four experimental conditions by comparing the percentage of false negatives.

Based on the comparative studies cited in the introduction, we decided to compare two different RR procedures: a self-administered question method employing CASI and a simple direct question using face-to-face interviewing. We hypothesized that the RR versions would register a higher proportion of respondents admitting fraud than CASI and that CASI would outperform face-to-face direct questioning.

Many earlier studies on RR made use of students or relatively highly educated respondents. To show how RR behaves in a more realistic setting, we selected respondents whose level of education was below average. Our sample also contained relatively many people who were not born in the Netherlands and did not possess native mastery of the Dutch language.

Two additional questions are as follows: (1) Are there explanatory variables that explain why some respondents admit fraud and others do not? and (2) Are the effects of these variables the same over all experimental conditions? To examine these questions, we must relate the available explanatory variables to the response to the sensitive question about fraud. For face-to-face direct questioning and CASI, this can be done by straightforward logistic regression. For the RR conditions, we used adjusted logistic regression techniques that take into account that in RR, the dependent variable is measured with a known form of error (see Maddala 1983; Scheers and Dayton 1988; van der Heijden and van Gils 1996). Answers to questions (1) and (2) will help us understand the process of answering honestly in surveys. Once a good understanding of this process exists, we can tailor the interviews for specific types of respondents using specific types of interview methods.

We distinguish three groups of explanatory variables. First, from earlier research on sensitive questions, it is known that opinions about sensitivity of questions differ for different (cultural) groups (Dohren-

wend 1966; Johnson et al. 1997; Junger 1989; Lyberg and Dean 1992). In addition, it is useful to include other general variables (i.e., not specific to the topic of welfare and unemployment benefit fraud) that are easy to determine (e.g., based on sampling frame). This leads to the first group of explanatory variables: individual attributes that are easily determined, such as gender, age, education, and being born in the Netherlands. The second group consists of general psychological variables that cannot be determined without the collaboration of respondents, such as control of the Dutch language and whether respondents found the sensitive questions annoying. The third group consists of variables that are specific to the topic of this study and would be different in research on other sensitive topics, such as perception of fraud, amount of fraud, and whether respondents still receive welfare or unemployment benefit.

In section 2, we give more technical details about the respondents, the four experimental conditions used, the operationalization of the questions, and the method of analysis. The results are described in section 3. The discussion is in section 4.

2. METHOD

2.1. RESPONDENTS

This experiment was part of a larger interview that investigated how clients of local welfare departments make ends meet and how they evaluate their local welfare department. We made use of databases in three Dutch cities with addresses of people who had been found guilty of welfare and unemployment benefit fraud between 1991 and 1994. We concentrate on fraud due to the nonreporting of additional income, since for other types of fraud the number of respondents was too small.

A few remarks are in order to describe the Dutch welfare system, which is rather different from the welfare systems in the United States, Australia, and Canada (for an elaborate comparison of European and non-European welfare systems, see Engbersen et al. 1993). The Netherlands has a complicated welfare system consisting of compulsory insurance and voluntary insurance, welfare, and pension. To understand

some of the details of this study, it is important to know that there is an institution called the local welfare department that is responsible for the distribution and control of welfare and unemployment benefit in the Netherlands. Unemployment benefit are based on the so-called Unemployment Insurance Act and equal 70 percent of the last earned wage for a period of several months up to several years, depending on the work career of the beneficiary. Welfare is based on the National Assistance Act and provides an income for those who are no longer entitled to any other unemployment benefit. This so-called social minimum is a percentage of the legal minimum wage and amounts to a maximum of 70 percent for singles, 90 percent for single parents, and 100 percent for couples. Thus, the local welfare department is responsible for only part of the welfare system. It is not responsible for other welfare acts (e.g., Sickness Benefits Act, Disability Insurance Act, General Widows and Orphans Act).

The respondents were approached in the second half of 1995. All had been found guilty of welfare fraud and received either an administrative sanction or a cut in their welfare or benefit. This makes it highly unlikely that the detection had passed unnoticed by a respondent. It is therefore assumed that respondents still remember this administrative sanction or income cut. However, cases of fraud are not always clear cut. For example, local welfare workers consider getting a job and not reporting it on time as fraud, although respondents who had done this may not necessarily intend to commit fraud, and might therefore consider themselves as "clean" cases. Another problem is cases in which more than one type of fraud has been assessed. The local welfare department usually registers the most important type, but the type it considers the most important is not necessarily the type considered most important by the respondent. We assume that the number of cases for whom this holds is small.

The local welfare department combined respondents identified as having committed fraud with respondents not so identified (it is assumed that it is likely that the latter respondents did not commit fraud). This sample then received a letter from the local welfare department stating that they might be approached by the independent commercial interview bureau Nederlands Instituut voor de Publieke Opinie en het Marktonderzoek (NIPO) (which is well known in the

Netherlands for its election polls and public opinion research) with the request to participate in a study about making ends meet while receiving welfare or unemployment benefit. The respondents were given the possibility of returning a postcard indicating that they did not want to participate (the so-called passive consent procedure). The respondents were informed that the information they provided in the interviews would not be reported to the local welfare department.

To guarantee the privacy of the respondents, those who did not refuse to participate were allocated numbers. Numbered names and addresses were provided to NIPO, but NIPO did not know which numbers referred to respondents caught for fraud. The interviewers were not aware that fraud was an issue in the study. Once the interviews had been carried out by NIPO, the researchers received the data files without the names and addresses; only the local welfare department knew which respondent numbers were related to being caught for fraud. Thus, none of the three parties involved had access to all three sources of information: addresses, answers, and respondents caught for fraud. The analyses reported in section 3 were performed on the sample of those respondents who have actually committed fraud.

A description of the response can be found in Table 1. The total sample approached by the local welfare departments announcing the interview was 1,774. Twenty percent of those refused to participate by sending in the reply card (passive consent). Thus, the interview bureau was provided with 1,418 addresses (c) by the local welfare department. For various reasons, including incorrect addresses, a lack of known telephone numbers, and a hot summer that led to many respondents' being away from home, the fieldwork took much longer than planned. To reduce costs, the fieldwork was terminated after five months, at which time 221 addresses (d) either were not used or had not yet been approached five times. Therefore, it is impossible to determine a clear response rate in terms of the sample approached (phase [a]). The remaining 1,197 addresses were approached up to five times until contact was made. For 406 addresses, we were unable to contact the respondent; in 35 percent of these 406 cases, this was due to incorrect addresses. For 196 addresses, contact was made but the respondent declined to participate (g). The most important reasons given were lack of interest (41 percent) and refusal after an appoint-

ment had been made (13 percent). Furthermore, 35 interviews were only partially successful and 26 interviews had incoherent answers or behavior and needed to be "cleaned." This left 534 interviews. In terms of the total sample approached initially by letter, the response rate was 30 percent; in terms of the number of individuals who did not use the reply card for refusal, the response rate was 38 percent; and in terms of the number of individual addresses actually approached, the response was 45 percent. Given the number of addresses not fully pursued in phase (d) because of time constraints, and the incorrect addresses in phase (f), these response rates are reasonable for Dutch standards (de Heer 1996). In total, 426 of the 534 interviews were realized for respondents identified as having committed fraud. A comparison of respondents and nonrespondents on background variables that were known for both groups revealed that the nonresponse was not significantly related to known background variables such as having been found guilty of fraud and level of education (van Gils, van der Heijden, and Landsheer 1996).

The educational level of the 534 respondents is low for Dutch standards. Thirty-five percent reported having only primary education, 34 percent only lower vocational training, 16 percent middle vocational training, 4 percent higher general education or secondary education preliminary to attending a university, and 5 percent college or university education.

2.2. THE FOUR EXPERIMENTAL CONDITIONS

At the start of the interview, the respondents were randomly distributed over the four experimental conditions, which we will discuss below in some detail.

2.2.1. Face-to-Face Direct Questioning

The questionnaire began with questions about country of birth, length of stay in the Netherlands, knowledge of the Dutch language, and educational level attained; job history and orientation; income and whether the respondent was able to make ends meet; and respondent's relation with the local welfare department and attitude toward the Dutch system of welfare. We started the interview with these

TABLE 1: Description of Response

	n	% Total[a]	% Willing[b]	% Participated[c]
a. Sample approached	1,774	100		
b. Refusal by reply card	356	20		
c. No refusal	1,418	80	100	
d. Not used, or in process when stopped	221	12	16	
e. Approached (five times maximum)	1,197	67	84	100
f. Approached, no contact after five times	406	23	29	34
g. Approached, contact but no interview	196	11	14	16
h. Response resulting in interview	595	34	42	50
i. Interview only partially successful	35	2	2	3
k. Removed in data cleaning	26	1	2	2
l. Used for analyses	534	30	38	45

a. Percentage of respondents with respect to the total sample approached.
b. Percentage of respondents with respect to the individuals willing to participate.
c. Percentage of respondents with respect to the individuals who actually participated.

questions to foster an atmosphere that would encourage respondents to answer the sensitive questions honestly and to justify any possible fraud if necessary. See Appendix B1 for the face-to-face direct questioning introduction.

The sensitive question was phrased as follows: "Have you ever failed to declare part of your income to the local welfare department as you are required to do by law?" Income included job income, earnings on the side, gifts, maintenance, and so on. Possible answers were "yes" and "no."

2.2.2. CASI

In CASI, the respondents were offered all questions by computer and asked to type in their answers. After the introduction (see Appendix B2), the interviewer was seated in a location where he or she could not see the computer screen. In the CASI condition, the sensitive question was the same as in the face-to-face condition.

2.2.3. Forced-Response Procedure

RR was originally proposed by Warner (1965) as a method to obtain more valid answers to sensitive questions. Since then, many

new RR procedures have been worked out that are statistically more efficient than Warner's proposal and less threatening to respondents (see Fox and Tracy 1986; Chaudhuri and Mukerjee 1988). Warner's procedure is threatening in that it uses two statements, and both statements deal with the sensitive topic, so that respondents may think there is a mathematical trick to sort out their real status. We use two of the improved procedures as experimental RR conditions in this study (see Appendix B3).

The first procedure is the so-called forced-response procedure (see Fox and Tracy 1986:24; Chaudhuri and Mukerjee 1988:16-17). We asked respondents to roll two dice and add the results. They were instructed to answer "yes" to the sensitive question if the outcome of the sum was 2, 3, or 4; "no" if the outcome was 11 or 12; and "yes" or "no" if the sum was between 5 and 10.

In the forced-response condition, the interviewer put the dice on the table at the start of the interview. In addition to explaining the topic of the interview, the interviewer stated: "We are also going to make use of a way of questioning that has not been used before in the Netherlands. But we will see this later in the interview."

The probability of having to answer the sensitive question was determined by following Soeken and Macready (1982), whose experiments show that the probability of answering the sensitive question should be between .7 and .85 (here, it is .75). The proportion of respondents answering "yes" to the sensitive question can be estimated as follows. Let P_1 be the probability of forced "yes," P_2 be the probability of forced "no," P_3 be the probability of answering the sensitive question, and π be the probability of admitting fraud on the sensitive question. So $P_1 = 1/6$, $P_2 = 1/12$, and $P_3 = 3/4$. Then, $P(yes) = P_1 + P_3\pi$, so that $\pi = \{P(yes) - P_1\}/P_3$. The sampling variance of π is $\text{Var}\{\pi\} = P(yes)(1 - P(yes))/nP_3^2$.

Fox and Tracy (1986:38) point out the following advantages of the forced-response procedure over other RR procedures. First, it is efficient. Second, it is comparatively easy for respondents to comprehend. Third, use can be made of the observation of Moriarty and Wiseman (1976) that respondents tend to overestimate P_1 and P_2. This means that for the example above in which $P_1 = 1/6$ and $P_2 = 1/12$, respondents perceive this probability to be larger than 1/4, and

therefore think that they are "safer" than they actually are. However, Fox and Tracy (1986:39) also mention the problem that sometimes respondents refuse to say "yes" when the randomizing device directs them to do so (Edgell, Himmelfarb, and Duchan 1982). For this reason, we also employ an RR procedure proposed by Kuk (1990) in which respondents only have to answer the color of a card.

2.2.4. Kuk's Procedure

In Kuk's procedure, there are two stacks of cards. In the left stack, the proportion of red cards is set by us at $P_1 = .8$; in the right stack, the proportion is set at $P_2 = .2$. The respondent is asked to draw one card from each stack. Then the sensitive question is asked. When the answer is "yes," the respondent should name the color of the left stack; when it is "no," the respondent should name the color of the right stack. Kuk's procedure was announced at the start of the interview in a similar way as in the forced-response procedure. For the introduction of Kuk's procedure to the respondents, see Appendix B4.

In Kuk's procedure, the proportion of respondents saying "yes" to the sensitive question can be estimated as follows. Let π be the probability of choosing the left stack. It follows that $P(red) = P_1\pi + P_2[1-\pi]$. Thus, $\pi = \{P(red) - P_2\}/\{P_1 - P_2\}$. The sampling variance of π is $\mathrm{Var}\{\pi\} = P(red)(1-P(red))/n(P_1-P_2)^2$ (for details, see Kuk 1990).

2.2.5. Referrals

For cases in which a respondent had extreme difficulty in using CASI or the RR procedures, he or she was referred to the face-to-face direct questioning.

For RR, the referral procedure is described in phases B3.3 and B3.4 (see Appendixes B3 and B4). The interviewer was instructed to check whether the respondent understood the RR procedure. In a pilot study, this worked out well. (In the pilot, it turned out that doubts and misunderstanding of respondents always were related to the points mentioned in B3.4.) Note that a respondent is not *asked* to work with the RR procedure. The interviewers were instructed that if a respondent was doubtful or had objections, one attempt had to be made to identify

the doubts or objections and to remove them. If this attempt was not successful, the respondent was referred to the face-to-face interview. The effect of these referrals on our results is discussed in section 3.

2.3. ANALYSIS

We restricted the analyses to those respondents known to have committed fraud, since we could verify the answers of those respondents. For respondents not known to have committed fraud, verification is more ambiguous, since not having been caught does not necessarily mean that respondents have never practiced fraud. Thus, the interpretation of the dependent variable on income fraud has a clear interpretation as "admitting fraud" with answer yes/no. For the RR procedures, the way to estimate the proportion of respondents admitting fraud was described in sections 2.2.3 and 2.2.4.

The logistic regression model is a natural candidate to relate the responses on the dichotomous dependent variable on income fraud to explanatory variables. Let $\mathbf{x}_i$ be a vector of explanatory variables of respondent i and let x_{ik} be the kth element of this vector. Let $\pi(\mathbf{x}_i)$ be the probability of answering positively to the sensitive question, given $\mathbf{x}_i$. Then, the log odds of answering positively are predicted by a regression model:

$$\log \frac{\pi(\mathbf{x}_i)}{(1-\pi(\mathbf{x}_i))} = b_0 + \sum_k x_{ik} b_k, \qquad (1)$$

where b_0 is a constant and b_k is the unstandardized regression coefficient for explanatory variable k. These parameters are asymptotically distributed as normal variates, and z values are obtained by dividing by the standard error; for $\alpha = .05$, a one-sided test is significant when the value obtained is larger than 1.65 or smaller than –1.65. The log odds are most easily interpreted by taking the exponential transformation, since $\exp(b_k)$ is the ratio of the odds of admitting fraud comparing respondents having one unit difference on x_{ik}, conditional on the values of the other variables.

Another way to interpret logistic regression results is to use the parameter estimates for b_0 and b_k to derives estimates of $\pi(\mathbf{x}_i)$. This can

be done using another (equivalent) way to denote the logistic regression model:

$$\pi(\mathbf{x}_i) = \frac{\exp\left(b_0 + \sum_k x_{ik} b_k\right)}{1 + \exp\left(b_0 + \sum_k x_{ik} b_k\right)}. \quad (2)$$

If $\mathbf{x}_i$ consists of categorical variables only, then $\pi(\mathbf{x}_i)$ can be derived for each combination of the categories. If $\mathbf{x}_i$ also contains quantitative variables, meaningful values of these variables should be chosen, for example, the mean and plus and minus one standard deviation from the mean.

For the RR approaches, we have adjusted the logistic regression model to incorporate the fact that the responses on the dependent variable are related in a known way to the sensitive question. This is worked out in Appendix A (Maddala 1983; Scheers and Dayton 1988; van der Heijden and van Gils 1996). Estimates for the regression coefficients from this adapted logistic regression can be interpreted as above. It should be remembered that for the RR approaches, the standard errors of the regression coefficients are relatively large, since RR generates considerable error in the dependent variable.

We also report combined b_k estimates plus standard error for the effect of the explanatory variables in all four experimental conditions, and a chi-square test for the significance of the systematic variation of the regression coefficients across the four experimental conditions. Regression coefficients were combined and their variance tested using standard random coefficient meta-analytic techniques (Raudenbush 1994).

3. RESULTS

3.1. PROCESS

In general, the interviewers felt the interviews went smoothly. In 80 percent of the interviews, language problems played no role, but 5

percent of the respondents did have serious problems with the Dutch language. In 90 percent of the cases, cooperation in the interviews was evaluated by the interviewers as good or very good, with no significant differences between the four experimental conditions. The attitude of more than 75 percent of the respondents was evaluated as unsuspicious, with an overrepresentation of suspicion for the forced-response procedure.

The interviewers indicated that 17 percent of the respondents found it difficult to understand the forced-response procedure; for Kuk's procedure, this figure was 21 percent. After the introduction and the first explanation of the RR procedures, 70 percent of the respondents were willing to cooperate. The other respondents made critical remarks, but after a second explanation they could be persuaded to cooperate. Eighteen percent of all respondents wanted to answer the sensitive questions straight away, and 9 percent declared they found the RR procedure nonsensical. In the RR procedures, 72 percent of the respondents were rated by the interviewers as not distrustful, 22 percent as a little distrustful, and 4 percent as distrustful of the procedure. In face-to-face direct questioning, these percentages were 77, 17, and 4, respectively (the difference between RR and face-to-face direct questioning was not significant). For CASI, cooperation and trustfulness could not be rated because the entire interview was filled in on the computer. However, 30 percent of the respondents required help or interviewer assistance while answering the questionnaire.

In comparing respondents known to have committed fraud with those not caught for fraud, there were almost no significant differences except that the cooperation of the former group was less forthcoming.

If respondents had extreme difficulty in mastering the RR procedure or CASI, interviewers were in principle allowed to switch respondents to the face-to-face interview. In CASI, the number of changes was largest (24 of 103 switched to face-to-face), whereas in the forced-response procedure and Kuk's procedure, it was much smaller (in CASI, 18 of 163 switched to face-to-face; in Kuk's procedure, 2 of 103 switched to CASI and 7 switched to face-to-face). In addition, 7 of 125 respondents switched from face-to-face direct questioning to CASI. There is a small overrepresentation of lower educational levels among the switchers.

It is clear that the switching in principle undermined the randomization, but that did not influence the marginal estimates of fraud. We also analyzed the data after removing the respondents who switched and obtained essentially the same results. Because the logistic analyses of the RR data require large samples, we present the results including respondents who switched.

3.2. PROPORTIONS OF RESPONDENTS ADMITTING FRAUD

The proportions of respondents admitting fraud are relatively low, ranging between 19 percent and 49 percent (see Table 2). The RR procedures performed significantly better than face-to-face direct questioning and CASI. Kuk's procedure performed the best, with 49 percent of the respondents admitting fraud, followed by the forced-response procedure with 43 percent. The difference between the two RR procedures was not significant. Contrary to our expectations, CASI (19 percent) produced even lower figures than face-to-face direct questioning (25 percent), although the difference was not significant.

People who switched from one condition to another were slightly less educated than people who did not. To check whether the switching between methods led to the lower estimates in face-to-face direct questioning, we checked the estimates of fraud for the face-to-face interview when the switchers are excluded. The figure decreases only marginally (23 percent). We conclude that face-to-face direct questioning is producing a higher figure than RR because "bad" respondents are switching from RR and CASI to face-to-face direct questioning. We proceed in the next sections with analyses of all respondents.

3.3. RELATIONSHIPS WITH EXPLANATORY VARIABLES

In this section, we report on the relationships between explanatory variables and admitting fraud. The results of logistic regression analyses are presented in Table 3, and estimates of admitting fraud for specific levels of the explanatory variables are given in Table 4. In Tables 3 and 4, three blocks of variables are distinguished. In block 1, we find individual attributes that are easily determined; in block 2, general psychological variables that cannot be determined without the

TABLE 2: Proportion of Respondents Admitting Income Fraud

Method	n	P *(fraud)*	SE	z
Forced response	96	.43	.068	2.22
Kuk	105	.49	.082	2.59
CASI	47	.19	.058	–0.83
Face-to-face	99	.25	.044	

NOTE: One-sided z tests were carried out against face-to-face direct questioning ($\alpha = .05$ corresponds to $z = 1.65$). CASI = computer-assisted self-interview. P = probability.

collaboration of respondents; and in block 3, variables that are specific to the topic of this study. All variables in a block are entered simultaneously in the logistic regression. A model entering all variables simultaneously results in estimation problems (the Hessian matrix that is part of maximum likelihood estimation cannot be inverted in this case). Significance of relations for a specific experimental condition can be assessed by the Wald test, which is a z test that divides the estimate for b_k by its asymptotic standard error (a value larger than 1.65 or smaller than –1.65 implies significance at $p = .05$ for a one-sided test). Table 3 also gives a precision-weighted estimate of b_k (column 13) that shows the overall effect of an explanatory variable across all four experimental methods. Column 15 gives the systematic variance over the four experimental conditions. When the systematic variance is low, which is usually the case, the four b_k estimates of the distinct conditions are not very different. This can be tested with a Pearson chi-square test with three degrees of freedom (the test value is given in column 16). Two final remarks must be made. First, because the number of tests that can be performed is large and not based on clearly predefined theoretical ideas, such tests have to be interpreted with care. Second, the sample sizes were determined to have enough power to perform the tests reported in Table 2, but here they often had a lower power. For this reason, we performed univariate logistic regression analyses, using one explanatory variable at a time.

The mean b_k and corresponding standard errors in columns 13 and 14 in Table 3 show that there are significant effects for gender, age, still receiving benefit, and duration and perception of fraud. Column 16 (chi-square with $df = 3$) shows that the results seem to be mostly

homogeneous over the experimental conditions, except for differential effects of being born in the Netherlands and the opinion about the local welfare department.

For explanatory variables in block 1 (individual attributes that are easily determined), Tables 3 and 4 show that in all experimental conditions, males and younger respondents admit more fraud, whereas females and older respondents more often give the socially desirable answer of no admission. Interestingly, education does not seem to play a role in admitting fraud. Also, being born in the Netherlands leads to more admission of fraud in face-to-face direct questioning but not in the other experimental conditions (RR and CASI) that protect privacy. This was tested in a post hoc analysis by adding a dummy predictor to the regression equation with a value of 1 for the face-to-face condition and 0 otherwise. The regression coefficient for this dummy is 1.78 (SE = 0.70), and the residual between-techniques variance vanishes completely (residual systematic variance is 0.2 percent).

For explanatory variables in block 2 (general psychological variables that cannot be determined without the collaboration of respondents), Table 3 shows no effect. In addition, for the randomized response procedures, *understanding* the procedure led to a higher admission of fraud (for more details, see Landsheer, van der Heijden, and van Gils 1999).

For block 3 (variables that are specific to the sensitive topic in this study), the first variable is "still receiving welfare or benefit." Some time had passed since the respondents had been found guilty of committing fraud, and a reasonable number of them no longer received benefit or welfare. We find that respondents still receiving are more reluctant to admit fraud, especially in the RR conditions. This suggests that two processes play a role in admitting fraud: fear that local welfare workers will learn of their answers and/or socially desirable behavior toward the interviewer. Respondents who received benefit or welfare for a shorter period of time before they were caught for fraud were more willing to admit fraud. Generally, respondents who state that fraud happens (rather) often are also more willing to admit their own fraud. Interestingly, in the RR conditions, the perceived fairness of procedures of the local welfare department leads to a slightly higher admission of fraud, whereas in CASI and face-to-face direct question-

TABLE 3: Logistic Regression

	Forced Response			*Kuk*			*CASI*			*Face-to-Face*			*Mean*		*Difference*	
	b_0 *(1)*	b_k *(2)*	SE *(3)*	b_0 *(4)*	b_k *(5)*	SE *(6)*	b_0 *(7)*	b_k *(8)*	SE *(9)*	b_0 *(10)*	b_k *(11)*	SE *(12)*	b_k *(13)*	SE *(14)*	%variable *(15)*	χ^2 *(16)*
Block 1																
Gender	.91	−.81	.58	1.04	−.79	.71	.58	−1.46	.87	.06	−.81	.49	−.89	.31	.1	.49
Age	1.51	−.049	.028	.69	−.019	.027	.62	−.060	.044	−.41	−.018	.023	−.031	.014	.2	1.31
Education	.15	−.13	.23	.14	−.05	.25	−1.24	−.05	.25	−1.14	.02	.16	−.04	.11	.1	.30
Born in Netherlands	−.42	.10	.56	.25	−.19	.65	−1.58	.10	.75	1.26	−1.77	.59	−.45	.47	53.6	6.65
Block 2																
Dutch speaking	−.30	.01	.35	−1.98	.55	.41	−.66	−.22	.45	−1.95	.24	.27	.17	.17	.2	1.89
Annoying	2.50	−1.01	.80	−.47	.16	.76	−4.23	.95	2.11	2.21	−1.14	.91	−.52	.46	.4	2.13
Block 3																
Still receiving	.54	−1.11	.65	.78	−1.22	.75	−1.39	−.07	.90	−.97	−.17	.50	−.59	.33	1.0	2.39
Duration	.39	−.12	.07	1.14	−.31	.22	−.85	−.09	.09	−.95	−.02	.05	−.67	.39	2.5	2.75
Evaluation of welfare department	−2.39	.77	.67	−.30	.10	.97	2.08	−1.23	1.03	4.32	−1.93	.71	−.55	.67	60.0	6.54
Perception of fraud	1.21	−.62	.39	.94	−.42	.46	−.80	−.26	.43	−.83	−.11	.31	−.32	.19	.2	1.12

NOTE: Response variable is income fraud. The covariates are gender (1 = male, 2 = female); age (in years); education (1 = *low*, 8 = *high*); born in the Netherlands (1 = yes, 2 = no); Dutch speaking (scale of four items, high score = *adequate mastering*); are questions annoying (scale of four items, high score = *not annoying*); benefit (dummies for health, unemployment, else, versus no benefit anymore); still receiving benefit (0 = not anymore, 1 = still receiving); duration of benefit (in years); evaluation of welfare department (scale of seven items, high score = *very fair*); and perception of fraud (scale of four items, high score = *fraud is often seen*). For each experimental condition, there are three columns, representing estimates for b_0, b_k, and the asymptotic standard error (*SE*) of b_k. Columns 13 and 14 provide the precision-weighted estimate of b_k and its standard error. Columns 15 and 16 provide the systematic variance over the four experimental conditions and a Pearson chi-square that tests whether there is evidence for differences in *b* values over the four methods. CASI = computer-assisted self-interview.

TABLE 4: Probability of Admitting Fraud Given Explanatory Variables

	Forced Response	*Kuk*	*CASI*	*Face-to-Face*
Block 1				
Gender				
Male (1)	.52	.56	.29	.32
Female (2)	.33	.37	.09	.17
Age				
30	.51	.53	.24	.28
50	.28	.44	.08	.21
Education				
Primary school (2)	.47	.51	.21	.25
Lower general education (4)	.41	.49	.19	.26
Born in the Netherlands				
Yes (1)	.42	.51	.19	.38
No (2)	.45	.47	.20	.09
Block 2				
Dutch speaking (1-5)				
Rather bad (2)	.43	.29	.25	.19
Almost good (4)	.44	.55	.18	.27
Questions annoying (1-3)				
Very (1)	.82	.42	.04	.75
Not (3)	.37	.50	.20	.23
Block 3				
Still receiving benefit				
No benefit (0)	.63	.69	.20	.27
Still receiving (1)	.36	.39	.19	.24
Duration				
1 year	.57	.70	.28	.27
10 years	.31	.12	.15	.24
Opinion of welfare department (1-5)				
Rather unfair (2)	.30	.48	.41	.61
Rather fair (4)	.67	.52	.06	.03
Perception of fraud (1-5)				
Happens rather often (2)	.49	.52	.21	.26
Does not happen often (4)	.22	.32	.14	.22

NOTE: Univariate logistic regression estimates derived from the values of *b* in Table 3. The values of the categories are given after the category labels. CASI = computer-assisted self-interview.

ing, it leads to a much lower admission of fraud. This was tested by including a dummy predictor with a value of 1 for the two RR techniques and 0 for face-to-face direct questioning and CASI. The regression coefficient for this dummy is 2.26 ($SE = 0.80$), and the residual

between-techniques variance vanishes completely (0.3 percent systematic variance).

4. DISCUSSION

We compared four different data collection methods for sensitive questions: face-to-face direct questioning, CASI, and two versions of RR. Because we had validating information available, we were able to check the answers with regard to false negatives.

The results are not comforting. Although the RR procedures performed much better than the more traditional procedures, they still result in serious underreporting. For the forced-response procedure, 43 percent of respondents admitted fraud; for Kuk's procedure, 49 percent; for face-to-face direct questioning, 25 percent; and for CASI, 19 percent. Because all respondents are known to have committed fraud, all these figures, of course, should have been 100 percent. Why are the proportions of respondents admitting fraud so low? Why do respondents not answer truthfully? From a legal point of view, their cases are already closed, and admitting fraud would not lead to any repercussions. One possibility is that some of those respondents not admitting fraud believe that what they did does not actually constitute fraud. Some reasons for this were given in section 3. A comparison of respondents who still receive benefit or welfare with respondents who do not suggests that the RR techniques succeed largely in removing bias from social desirability effects but are less successful in removing bias due to fear of repercussions.

Overall, the two RR techniques performed much better than either CASI or face-to-face direct questioning. In fact, CASI performs very poorly. This is surprising given the generally positive results for CASI such as reported in the meta-analysis by Weisbrand and Kiessler (1996). Our different result may be due to questioning a difficult group of respondents. As mentioned previously, the respondents in this study had received unemployment benefit or welfare, had been found guilty of fraud, and had a much lower educational level than that of the general population. Most studies synthesized by Weisbrand and Kiessler investigated either the general population or a "well-behaved" special population such as students. It is suggestive that in

our population, 23 percent of the persons placed in the CASI condition had to be moved to face-to-face direct questioning due to computer-related problems or illiteracy. In contrast, in the two RR conditions, only 5 percent to 9 percent of the respondents had to be moved. We suggest that *standard* CASI for eliciting sensitive information works well only with populations that are at least moderately familiar with computers. When special populations are investigated, special CASI adaptations to the needs of that population should be made (De Leeuw et al. 1997). One possible adaptation is to use audio CASI (Tourangeau and Smith 1998; Turner et al. 1998). It should also be noted that the meta-analysis by Weisbrand and Kiessler finds differences between studies, which indicates that moderator variables may influence the effectiveness of CASI interviewing.

If we compare the two RR techniques, Kuk's procedure seems to perform slightly better than the forced-response procedure. Kuk's procedure results in slightly more positive responses (49 percent vs. 43 percent). Most of the other comparisons between the forced-response procedure and Kuk's procedure show no significant differences, but in all cases the direction of the difference favors Kuk's procedure. Finally, slightly more respondents were able to work with Kuk's procedure (95 percent vs. 91 percent), and respondents felt less suspicious of Kuk's procedure.

There are a number of background variables that predict admitting fraud. The respondent most likely to admit fraud is a relatively young male who has received welfare or benefit only briefly and is no longer dependent upon it, and who perceives fraud as frequently occurring. Two background variables differ in their relationship with admitting fraud: being born in the Netherlands and perceiving the welfare department as very fair. Persons born in the Netherlands are more likely to admit fraud in response to a direct question. Perception of the welfare department as fair predicts more fraud admission in the RR procedures and less fraud admission in CASI and face-to-face direct questioning. The direction of the difference suggests that both RR techniques are perceived as different from CASI and direct questioning. Given the poor performance of CASI in our experiment, we believe that despite the use of CASI for the sensitive question, our respondents are still intimidated by the presence of an interviewer.

The more elaborate RR procedures are more successful in counteracting respondents' suspicion.

This study has a number of strong points. First, it provides a realistic field test of methods for eliciting sensitive information. The population studied is poorly educated and considered difficult to interview. We have shown that randomized response procedures can be used with this group. Second, because we know who had committed fraud, we could check respondents' answers for false negatives and use this information to evaluate the data collection procedures.

A weak point of the study is the level of unit nonresponse. Compared with other countries, the nonresponse in the Netherlands is rather high. For instance, in the year we collected our data (1995), the total nonresponse on the Dutch Labor Force Survey was 40 percent (de Heer 1996). Nevertheless, even for Dutch standards, our nonresponse is above average. However, because we know a number of background variables for both respondents and nonrespondents (see section 2.1), we were able to check our sample for selective nonresponse on some important variables. No statistically significant relations of nonresponse with education and with having committed fraud were found, the latter being the key variable in this validation study.

Furthermore, the precise formulation of our key question could be improved. We asked respondents whether they had ever failed to declare part of their income (see section 2.2.1 for precise formulation). We did not ask the respondents whether they were ever charged or perhaps convicted of not having declared income to the local welfare department. Because a record of conviction was available as validating information, the second formulation would have followed the validation standard much closer. It would also have made it possible to check for false positives. In interviews about sensitive topics, underreporting of socially disapproved items is feared most (Aquilino 1994; Jobe et al. 1997; Locander et al. 1976), but the opposite danger of overreporting or boasting is also possible. As stated, asking whether respondents were ever convicted makes it possible to check for false positives among those who were not convicted and for underreporters among those who were convicted. The original question about committing fraud is a question that is used in this form in Dutch surveys,

and we use it to ensure the external validity of the experiment. However, it would be interesting to use both questions in another experiment.

APPENDIX A
Logistic Regression Models for Randomized Response Data

The logistic regression model relates explanatory variables to a dichotomous response variable. In this appendix, we show how this model is adjusted to deal with the two types of randomized response (RR) data described in the text (for estimation procedures, see van der Heijden and van Gils 1996).

First, consider computer-assisted self-interviewing (CASI) and face-to-face direct questioning. Let $\pi(\mathbf{x}_i)$ be the probability of giving a positive answer to the sensitive question as a function of explanatory variable vector $\mathbf{x}_i$, and let $[1-\pi(\mathbf{x}_i)]$ be the corresponding probability of a negative response. Let the kth element of $\mathbf{x}_i$ be denoted by x_{ik}. Let b_k be the regression coefficient for explanatory variable k. The logistic regression model is then defined as

$$\pi(\mathbf{x}_i) = \frac{\exp\left(b_0 + \sum_k x_{ik} b_k\right)}{1 + \exp\left(b_0 + \sum_k x_{ik} b_k\right)}. \tag{A1}$$

This definition of logistic regression is equivalent to the one in section 2.3. Let n_{i1} be the number of respondents giving a positive answer with explanatory variable vector $\mathbf{x}_i$, and let n_{i0} be the number of respondents giving a negative answer. The log likelihood for the model is then

$$\log L = \sum_i n_{i1} \log \pi(\mathbf{x}_i) + \sum_i n_{i0} \log[1-\pi(\mathbf{x}_i)]. \tag{A2}$$

This log likelihood can be maximized over the parameters b_k using numerical methods.

We will now adjust the logistic regression model for the forced-response procedure (for similar proposals for different RR procedures, see Maddala 1983:54-56; Scheers and Dayton 1988). Let the probability of answering by forced "yes" be P_1, the probability of answering by forced "no" P_2, and the probability of answering honestly $P_3 = 1 - P_1 - P_2$. The probability of a "yes" answer in the forced-response procedure, given $\mathbf{x}_i$, is then $P(yes|\mathbf{x}_i) = P_1 + P_3\pi(\mathbf{x}_i)$ and $P(no|\mathbf{x}_i) = P_2 + P_3[1-\pi(\mathbf{x}_i)]$. Now, n_{i1} and n_{i0} are defined in terms of the number of respondents giving "yes" answers and "no" answers in the forced-response procedure. Thus, the log likelihood for the RR data becomes

$$\log L = \sum_i n_{i1} \log[P_2 + P_3\pi(\mathbf{x}_i)] + \sum_i n_{i0} \log(P_1 + P_3[1-\pi(\mathbf{x}_i)]), \tag{A3}$$

and this log likelihood is to be maximized over the parameters b_k defined in (A1).

In Kuk's procedure, the interviewer has two stacks of cards in front of him or her. In the left-hand stack, the proportion of red cards is P_1, and in the right-hand stack, the proportion is P_2. The respondent has to draw one card from each stack and name the color of the card from the left-hand stack if he should answer "yes" to the sensitive question and the color of the card from the right-hand stack if he should answer "no" to the sensitive question. For ease of notation, we define $P_3 = 1 - P_1$ and $P_4 = 1 - P_2$. Thus, $P(red|\mathbf{x}_{ij}) = P_1\pi(\mathbf{x}_i) + P_2[1 - \pi(\mathbf{x}_i)]$ and $P(black|\mathbf{x}_i) = P_3\pi(\mathbf{x}_i) + P_4[1 - \pi(\mathbf{x}_i)]$. Now, n_{i1} and n_{i0} are defined in terms of the number of respondents giving "red" answers and "black" answers in Kuk's procedure. Thus, the log likelihood for the RR data becomes

$$\log L = \sum_i n_{i1} \log(P_1\pi(\mathbf{x}_i) + P_2[1 - \pi(\mathbf{x}_i)]) + \sum_i n_{i2} \log(P_3\pi(\mathbf{x}_i) + P_4[1 - \pi(\mathbf{x}_i)]), \quad \text{(A4)}$$

and this log likelihood is to be maximized over the parameters b_k defined in (A1).

APPENDIX B

The interviewers are instructed to follow the questionnaire literally. If problems or refusals by the respondent arise during the interview, the interviewer is instructed to solve them to the best of his or her ability and report them at the end of the interview. In this way, we would know how many respondents changed from one experimental condition to another.

We present the text blocks introducing the sensitive questions for each of the four experimental conditions. Some of the text blocks are used in more than one experimental condition, and therefore we number them. Text spoken by the interviewer is in quotes. Written instructions to the interviewer are in brackets.

B1. FACE-TO-FACE DIRECT QUESTIONING

B1.1. "We now would like to ask a couple of questions about topics that we already touched upon, for example, your income and possessions, extra high expenses, looking for work, and providing information to the local welfare department. This can have to do with, for example, declaring part of your income from a side job, family reunion, or living together. In short, about information that for all sorts of reasons often is not, only partly, or not in time provided to the local welfare department."

B1.2. "We ask you to answer these questions with 'yes' or 'no.'"

B1.3. "We understand that this can sometimes be difficult because you will not always have a ready-made answer. That is why we ask you to answer 'yes' when the answer is 'mostly yes' and no when the answer is 'mostly no.'"

B1.4. "We will now ask you a few questions about your expenses and income and about providing information to the local welfare department."

B1.5. [Important. The questions have to be read word by word, including the explanation of the terms, so that the respondent does not need to ask for clarification.]

B1.6. Questions follow about (1) saving for a large expenditure; (2) providing address information to the local welfare department; (3) officially having a car worth more than approximately $15,000; (4) having a motor home; (5) going abroad for holiday longer than four weeks; (6) gambling a large amount (more than $25) at the horses, in casinos, in playing halls, or on bets; (7) having hobbies about which you or household members think cost too much, given the income you have; (8) having refused jobs, or taken care that employers did not want you for a job while you had a good chance to get the job; (9) working more than 20 hours as a volunteer without the local welfare department's knowledge; (10) not declaring part of your income to the local welfare department, whereas this is obligatory by law; (11) living now with a partner without the local welfare department's knowledge; (12) having lived with a partner without the local welfare department's knowledge; and (13) giving the local welfare department insufficient or incorrect information about having a fortune. Note that (10) is the dependent variable that is the key variable in this article (see section 2.2.1 for the exact formulation). Also note that questions (1) to (7) are not referring to fraud in any way. They are meant simply to pave the way for more sensitive questions.

B2. Computer-Assisted Self-Interview (CASI)

The text blocks used here are B1.1, B1.2, B1.3, B1.4, and B1.6.

At the start of the interview, the computer is introduced as follows: [Read aloud and give the computer to the respondent] "Could you please handle the PC yourself for reasons of confidentiality. By showing some examples of questions, I will explain how you have to handle the PC. In a multiple question, you have to type in all the answer category numbers sequentially. Once you have indicated all answers, you end the question with [Enter] and proceed to the next question. If you have typed in a wrong number, then you can remove it with the [Backspace] button." Five distinct examples of questions follow. "Now we start with the real questionnaire. If you have problems or difficulties with certain questions or other things, then you may of course always ask the interviewer for an explanation." Subsequently, the interview proceeds on the computer.

B3. RANDOMIZED RESPONSE: FORCED-RESPONSE PROCEDURE

The sensitive block starts with B1.1. Then,

B3.1. "Many people find it difficult to answer these types of questions straightaway because they find the topics too private. Yet, we do not want to embarrass anyone.

Therefore, we ask you these questions, experimentally, in a roundabout way. We let you answer in such a way that your privacy is guaranteed so that nobody can ever find out what you have done personally, including me."

B3.2. "You may answer in a few moments using two dice. With those, you can throw 2 or 12 or something in between. You answer is dependent on what you throw with the dice." [Give the box to the respondent and look at it together.] "In the box you will find a card showing what you have to say when you have thrown the dice." [Let interviewee look and give directions with the next explanation.] "If you throw 5, 6, 7, 8, 9, or 10, you always answer 'yes' or 'no' honestly. If you throw 2, 3, or 4, you always answer 'yes.' If you throw 11 or 12, you always answer 'no.' So, if you throw 2, 3, or 4, or 11 or 12, then your answer is based on the outcome of the throw. Because I cannot see what you have thrown, your personal privacy is guaranteed; thus, your answer always remains a secret.

"This technique is a bit strange. But it is useful, since it allows people working for Utrecht University to estimate how many people of the group that we interviewed answered 'yes' because they threw 2, 3, or 4 and how many people answered 'yes' because they had to give an honest answer.

"Let us take an example. I ask you the question: 'Do you live in Utrecht?' and you throw a 3. You answer with 'yes.'

"We can imagine that you find this a bit awkward, but it does not mean that you are lying or that someone can think that the honest answer to the question is also 'yes.' It means only that you stick to the rules of the game by which your privacy and that of everybody else taking part in this investigation is fully guaranteed. I propose that we now try out a few questions to practice."

B3.3. [Turn around] and B1.5.

"I ask you the first six questions to practice."

Questions follow about whether the respondent (1) read a newspaper today, (2) ignored a red traffic light, (3) received a fine for driving under the influence of alcohol, (4) used public transportation last year without paying at least once, (5) paid the obligatory fee for television and radio, (6) ever bought a bicycle suspecting it was stolen.

The instruction goes on with the following.

"Is it clear now? Then we will now ask the questions we are really interested in. Please take your time to answer them."

[Do not start with the real questions before you are certain the next points are understood. Do not read the following points aloud. Read one of the points aloud only when that point is unclear to the respondent.]

B3.4. [We do this to guarantee your privacy. Nobody sees what you throw and nobody will know what your personal answer is. According to the rules of the game, answers are possible that are in conflict with your feelings: "yes" when it is "no" and "no"

when it is "yes." It is not lying; it simply guarantees your privacy. Based on all answers of the people that we interviewed, we can estimate afterward how many people have read a newspaper today or ignored a red traffic light, and so on.] Followed by B1.4, B1.5, and B1.6.

B4. RANDOMIZED RESPONSE: KUK'S PROCEDURE

The sensitive block starts with B1.1 and B3.1. Then,

B4.1. "Let us take an example to see how it works. I have two stacks of cards and a box behind in which I place the cards. [Give the box to the respondent and look at it together.] In the box, you find a card on which it is written what the stack means: the left-hand stack is the 'yes' stack, and the right-hand stack is the 'no' stack. [Let interviewee look and give directions with the next explanation.] In the 'yes' stack [Point to left-hand stack] there are more red cards than in the 'no' stack. [Point to right-hand stack. Respondent may check.] If you want, you may shuffle the two stacks [Separately]. Now, please take from each stack an arbitrary card. You may take the card on top or from within the stack. [Take a card from each stack.] Now I ask you the question: 'Have you have read a newspaper today?' If you have read a newspaper, your answer is 'yes,' but you are not allowed to say this. Instead, you mention the color of the card that you have taken from the 'yes' stack, that is, the left-hand stack. Do not answer too fast and take care that you mention the color of the correct card and not 'yes' or 'no.' Nobody but you can see the colors of your cards; when you mention a card color, we do not know the stack from which you took the card. Thus, your personal privacy is guaranteed: Your answer will always remain a secret.

"This technique is a bit strange. But it is useful, since it allows people working for Utrecht University to estimate how many people of the group that we interviewed answered 'red' or 'black' because the answer was 'yes' and how many people answered 'red' or 'black' because the answer was 'no.'

"We can imagine that you find this a bit awkward, but by the rules of the game your privacy and that of everybody else taking part in this investigating is fully guaranteed. I propose that we now try out a few questions to practice." Followed by B3.3.

B4.2. [By this way of questioning, we can guarantee your privacy completely. Based on all answers of the people that we interviewed, we can estimate afterward how many people have answered "yes" or "no." Red does not mean "yes" and black does not mean "no."] Followed by B1.4, B1.5, and B1.6.

REFERENCES

Aquilino, W. 1994. "Interview Mode Effects in Surveys of Drug and Alcohol Use." *Public Opinion Quarterly* 58:210-40.

Beckenbach, Andreas. 1992. "Befragung mit dem Computer, Methode der Zukunft? Anwendungsmöglichkeiten, Perspektiven und experimentelle Untersuchungen zum Einsatz des Computers bei Selbstbefragung und persönlich-mündlichen Interviews" (Computer Assisted Interviewing, the Method of the Future? An Experimental Study of the Use of a Computer in Self-Administered Questionnaires and Face-to-Face Interviews). Ph.D. thesis, Universität Mannheim.

———. 1995. "Computer Assisted Questioning: The New Survey Methods in the Perception of the Respondent." *Bulletin de Méthodologie Sociologique* 48:82-100.

Berghuis, A. C. and Max M. Kommer. 1982. *Effecten van voorlichting en controle: Een experiment bij de motorrijtuigenbelasting*. Den Haag, the Netherlands: Department of Justice, Research and Documentation Center (RDC).

Chaudhuri, Arijit and Rahul Mukerjee. 1988. *Randomized Response: Theory and Techniques*. New York: Dekker.

de Heer, Wim. 1996. "International Response Trends." Presented at the 4th International Social Methodology Conference, July, Colchester, UK.

De Leeuw, Edith D. 1992. *Data Quality in Mail, Telephone, and Face-to-Face Surveys*. Amsterdam: TT-Publicaties.

De Leeuw, Edith D., Joop J. Hox, Sabina Kef, and Marion van Hattum. 1997. "Overcoming the Problems of Special Interviews on Sensitive Topics: Computer-Assisted Self Interviewing Tailored for Young Children and Adolescents." Pp. 1-14 in *Proceedings of the Sawtooth Software Conference*. Sequim, WA: Sawtooth Software Inc.

De Leeuw, Edith D., Joop J. Hox, and Ger Snijkers. 1995. "The Effect of Computer Assisted Interviewing on Data Quality: A Review." *Journal of the Marketing Research Society* 37:325-44.

Dillman, Don A. 1978. *Mail and Telephone Surveys: The Total Design Method*. New York: Wiley.

Dohrenwend, B. 1966. "Social Status and Psychological Disorder: An Issue of Substance and an Issue of Method." *American Sociological Review* 31:14-34.

Droitcour, Judith, Rachel A. Caspar, Micheal L. Hubbard, Teresa L. Parsley, Wendy Visscher, and Trena M. Ezzati. 1991. "The Item Count Technique as a Method of Indirect Questioning: A Review of Its Development and a Case Study of Its Application." Pp. 185-210 in *Measurement Errors in Surveys*, edited by P. P. Biemer, R. M. Groves, L. E. Lyberg, N. A. Mathiowetz, and S. Sudman. New York: Wiley.

Edgell, Stephen E., Samuel Himmelfarb, and Karen L. Duchan. 1982. "Validity of Forced Response in a Randomized Response Model." *Sociological Methods & Research* 11:89-110.

Elffers, Henk, Henry S. J. Robben, and Dick J. Hessing. 1992. "On Measuring Tax Evasion." *Journal of Economic Psychology* 13:545-67.

Elffers, Henk, Henry S. J. Robben, and W.F.M. Verlind. 1989. "Determinanten van individueel misbruik van sociale verzekeringen: Het geval van WW-misbruik. Verslag van een onderzoek verricht in opdracht van de Federatie van Bedrijfsverenigingen door de Projectgroep Belastingweerstand" (Determinants of Individual Misuse of Social Security). Unpublished manuscript, Faculteit der Rechtsgeleerdheid, Erasmusuniversiteit, Rotterdam.

Engbersen, Godfried, Kees Schuyt, Jan Timmer, and Frans van Waarden. 1993. *Cultures of Unemployment: A Comparative Look at Long-Term Unemployment and Urban Poverty*. Oxford: Westview Press.

Evan, W. M. and J. R. Miller. 1969. "Differential Effect of Response Bias of Computer vs. Traditional Administration of a Social Science Questionnaire: An Exploratory Methodological Experiment." *Behavioral Science* 14:216-27.

Fox, James A. and Paul E. Tracy. 1986. *Randomized Response: A Method for Sensitive Surveys*. Beverly Hills, CA: Sage.

Hessing, Dick J., Henk Elffers, and Russell H. Weigel. 1988. "Exploring the Limits of Self-Reports and Reasoned Action: An Investigation of the Psychology of Tax Evasion Behavior." *Journal of Personality and Social Psychology* 54:405-13.

Hochstim, J. R. 1967. "A Critical Comparison of Three Strategies of Collecting Data From Households." *Journal of the American Statistical Association* 62:976-89.

Hosseini, Jamskid C. and Robert L. Armacost. 1993. "Gathering Sensitive Information in Organizations." *American Behavioral Scientist* 36:443-71.

Jobe, Jared B., William F. Pratt, Roger Tourangeau, Alison K. Baldwin, and Kenneth A. Rasisnski. 1997. "Effects of Interview Mode on Sensitive Questions in a Fertility Survey." Pp. 311-29 in *Survey Measurement and Process Quality*, edited by L. Lyberg, P. Biemer, M. Collins, E. De Leeuw, C. Dippo, N. Schwarz, and D. Trewin. New York: Wiley.

Johnson, Timothy, Diane O'Rourke, Noel Chavez, Seymour Sudman, Richard Warnecke, Loretta Lacey, and John Horm. 1997. "Social Cognition and Responses to Survey Questions Among Culturally Diverse Populations." Pp. 87-113 in *Survey Measurement and Process Quality*, edited by L. Lyberg, P. Biemer, M. Collins, E. De Leeuw, C. Dippo, N. Schwarz, and D. Trewin. New York: Wiley.

Johnston, J. and C. Walton. 1995. "Reducing Response Effects for Sensitive Questions: A Computer-Assisted Self Interview With Audio." *Social Science Computer Review* 13:304-19.

Junger, Marianne. 1989. "Discrepancies Between Police and Self-Report Data for Dutch Racial Minorities." *British Journal of Criminology* 29:273-83.

———. 1990. *Delinquency and Ethnicity*. Deventer, the Netherlands: Kluwer.

Kiessler, S. and L. S. Sproull. 1986. "Response Effects in Electronic Surveys." *Public Opinion Quarterly* 50:402-13.

Koopmans, C. C. 1988. *De informele arbeidsmarkt, een onderzoek naar vraag, aanbod en prijzen. OSA voorstudie nr. V 27* (The Informal Labor Market). Den Haag, the Netherlands: Department of Social Affairs and Employment.

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

Landsheer, Johannes A., Peter G. M. van der Heijden, and Ger van Gils. 1999. "Trust and Understanding, Two Psychological Aspects of Randomized Response: A Study of a Method for Improving the Estimate of Social Security Fraud." *Quality and Quantity* 33 (1): 1-12.

Lee, Raymond M. 1993. *Doing Research on Sensitive Topics*. Newbury Park, CA: Sage.

Locander, William, Seymour Sudman, and Norman Bradburn. 1976. "An Investigation of Interview Method, Threat and Response Distortion." *Journal of the American Statistical Association* 71:269-75.

Lyberg, Lars E. and Pat Dean. 1992. "Methods for Reducing Nonresponse Rates." Presented at the annual conference of the American Association for Public Opinion Research, May 15-19, St. Petersburg, FL.

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

Martin, C. L. and D. H. Nagao. 1989. "Some Effects of Computerized Interviewing on Job Applicant Responses." *Journal of Applied Psychology* 74:72-80.

Moriarty, Mark and Frederick Wiseman. 1976. "On the Choice of a Randomization Technique With the Randomized Response Model." Pp. 624-26 in *Proceedings of the Social Statistics Section of the American Statistical Association*. Washington, DC: American Statistical Association.

Nicholls, William L. II, Reginald P. Baker, and Jean Martin. 1997. "The Effect of New Data Collection Technology on Survey Data Quality." Pp. 221-48 in *Survey Measurement and Process Quality*, edited by L. Lyberg, P. Biemer, M. Collins, E. De Leeuw, C. Dippo, N. Schwarz, and D. Trewin. New York: Wiley.

Raudenbush, Steven W. 1994. "Random Effects Models." Pp. 301-21 in *The Handbook of Research Synthesis*, edited by H. Cooper and L. V. Hedges. New York: Russell Sage Foundation.

Scheers, N. J. 1992. "A Review of Randomized Response Techniques." *Measurement and Evaluation in Counseling and Development* 25:27-41.

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

Siemiatycki, J. 1979. "A Comparison of Mail, Telephone, and Home Interview Strategies for Household Health Surveys." *American Journal of Public Health* 69:238-45.

Soeken, Karen L. and George B. Macready. 1982. "Respondents' Perceived Protection When Using Randomized Response." *Psychological Bulletin* 92:487-89.

Sudman, Seymour and Norman M. Bradburn. 1979. *Improving Interview Method and Questionnaire Design: Response Effects to Threatening Questions in Survey Research.* San Francisco: Jossey-Bass.

———. 1982. *Asking Questions: A Practical Guide to Questionnaire Design.* San Francisco: Jossey-Bass.

Tourangeau, Roger and Tom W. Smith. 1996. "Asking Sensitive Questions: The Impact of Data Collection, Question Format, and Question Context." *Public Opinion Quarterly* 60:275-304.

———. 1998. "Collecting Sensitive Information With Different Modes of Data Collection." Pp. 431-54 in *Computer Assisted Survey Information Collection*, edited by M. P. Couper, R. P. Baker, J. Bethlehem, C.Z.F. Clark, J. Martin, W. L. Nicholls II, and J. M. O'Reilly. New York: Wiley.

Turner, Charles F., Barbara H. Forsyth, James M. O'Reilly, Philip C. Cooley, Timothy K. Smith, Susan M. Rogers, and Heather G. Miller. 1998. "Automated Self-Interviewing and the Survey Measurement of Sensitive Behavior." Pp. 455-74 in *Computer Assisted Survey Information Collection*, edited by M. P. Couper, R. P. Baker, J. Bethlehem, C.Z.F. Clark, J. Martin, W. L. Nicholls II, and J. M. O'Reilly. New York: Wiley.

Turner, Charles F., J. T. Lessler, and J. W. Devore. 1992. "Effects of Mode of Administration and Wording on Reporting of Drug Use." Pp. 177-219 in *Survey Measurement of Drug Use: Methodological Studies*, edited by Charles F. Turner, J. T. Lessler, and J. C. Gfroerer. Washington, DC: Government Printing Office.

Umesh, U. N. and Robert A. Peterson. 1991. "A Critical Evaluation of the Randomized Response Method: Applications, Validation, and Research Agenda." *Sociological Methods & Research* 20:104-38.

van der Heijden, Peter G. M. and Ger van Gils. 1996. "Some Logistic Regression Models for Randomized Response Data." Pp. 341-48 in *Statistical Modelling: Proceedings of the 11th International Workshop on Statistical Modelling*, edited by Antonio Forcina, Gianfranco M. Marchetti, Reinhold Hatzinger, and Giovanni Galmatti. Perugia, Italy: University of Perugia, Department of Statistics.

Van Eck, R. and Bregt Kazimier. 1990. *Zwarte inkomsten uit arbeid: Een nadere analyse van de uitkomsten van enquete-onderzoek* (Unofficial Income From Labor). Manuscript 36. Voorburg: Statistics Netherlands, Statistische onderzoekingen.

van Gils, Ger, Peter G. M. van der Heijden, and Hans Landsheer. 1996. *Rapportage van uitkeringsfraude in surveys* (Reporting Welfare Fraud in Surveys). Den Haag, the Netherlands: Department of Social Affairs and Employment.

Warner, Stanley L. 1965. "Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias." *Journal of the American Statistical Association* 60:63-69.

Weisband, Suzanne and Sara Kiessler. 1996. "Self Disclosure on Computer Forms: Meta-Analysis and Implications." In *CHI'96 Electronic Proceedings*, edited by R. Bilger, S. Guest, and M. J. Tauber. ACM. Available: http://www1.acm.org:82/sigchi/chi96/proceedings

Peter G. M. van der Heijden is a professor of statistics in the Faculty of Social Sciences at Utrecht University. He is interested in the estimation of population sizes in cases of sensitive topics and categorical data analysis.

Ger van Gils is director and senior researcher at BeleidsOnderzoek en Advies, Utrecht. His interests include the estimation of population sizes in cases of sensitive topics.

Jan Bouts is senior consultant at Nederlands Instituut voor de Publieke Opinie en het Marktonderzoek.

Joop J. Hox is a professor of survey methodology in the Faculty of Social Sciences at Utrecht University. His main research interests are survey methodology, data quality, and the analysis of complex data with multilevel models and structural equation models.