A comparison of randomized response, CASAQ, and direct questioning; eliciting sensitive information in the context of social security fraud

Peter G.M. van der Heijden*, Ger van Gils**, Jan Bouts*** and Joop Hox*

Abstract

The aim of the study is to assess the validity of the responses to sensitive questions on social security fraud obtained using four different methods. We compare two different varieties of randomized response with CASAQ and direct questioning in an experimental setting. Validity could be assessed because all respondents interviewed had already been identified as committing social security fraud. The experiment was set up in such a way that the interviewers did not know that respondents had been caught for fraud, and the respondents did not know that the researchers had this information. Since the actual status of the respondents is known, it is possible to compare the results of the four approaches by comparing the percentage of false negatives.

Two additional questions are these: which respondents are willing to admit to having practised fraud, and are the respondent characteristics that predict positive responses to the sensitive questions the same for all methods? For the direct question method and CASAQ these questions are answered by straightforward logistic regression. For the randomized response answers, logistic regression models are adjusted.

Although the RR conditions perform much better than more traditional approaches, the percentage of respondents admitting to fraud is far less than 100 %. Some reasons for this are discussed.

Keywords: sensitive questions, CASAQ, randomized response, logistic regression

---

*Vakgroep Methodenleer en Statistiek, FSW, Universiteit Utrecht, Postbus 80.125, 3508 TC Utrecht. Tel.nr. 0302534688, p.vanderheijden@fss.uu.nl, j.hox@fss.uu.nl. **BOA, Nieuwe Gracht 98, 3512 LX Utrecht. Tel.nr. 0302328449, boavgils@cuci.nl. ***NIPO, Postbus 247, 1000 AE Amsterdam. Tel.nr. 0205225444, jan.bouts@nipo.nl.

1. Introduction

Information on the types and extent of social security fraud are relevant for any assessment of its importance as a policy issue. Both in the US and in Europe the issue of social security has commended growing attention in recent years, both in politics and society as a whole. Because efficient fraud control is generally seen as an integral component of the social security system, organisations responsible for carrying out social security have increased their efforts to deal with the problem. An important aspect of the problem is that it is difficult to obtain valid and reliable information in this area. If 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 themselves engaged in fraud.

The problem of obtaining valid and reliable information is not unique to fraud, but also plays a role in other sensitive topics. Standard approaches to obtaining sensitive information, such as self-reports, tend to give poor results in general. A few examples from the Netherlands illustrate the point. By comparing police records with self-reports, Junger (1989, 1990) showed that Dutch adolescents report only 60 to 70 % of all offences for which they have ever been caught by the police. When attention is restricted to the most recent year, only 23 % of all adolescents in her research admitted their offences, with important differences according to ethnic background. For instance, this percentage for Dutch youth is 38 % but for Turkish youth only 9 %. In an investigation of social security fraud, Elffers, Robben and Verlind (1989) found that 43 % of those respondents who had been caught for fraud, did not admit this. It should be noted that this research was plagued by a non-response of 87 %, which highlights another problem of research on sensitive topics. Other examples deal with tax fraud. In a telephone interview among people already caught for vehicle tax fraud, Berghuis and Kommer (1982) found that only approximately 10 % admitted evading this tax. Hessing, Elffers and Weigel (1988; see also Elffers, Robben and Hessing 1992) report that, in a face-to-face interview, 70 % of their respondents denied ever having evaded income tax, whereas in fact all their respondents had been found guilty of this offence.

All surveys and also Dutch surveys are plagued by a high non-response rate (de Heer 1996). Surveys on moonlighting are also plagued by item non-response and false answers; this can be deduced from comparisons of surveys with macro-economic estimates. Van Eck and Kazimier (1990; see also Koopmans 1988) argue that moonlighters receiving unemployment benefit are even less inclined to cooperate with survey research into moonlighting than those who have a job, since the former group is more at risk. In addition to an administrative sanction, they also risk loosing part of their income. It is clear that the standard approaches of surveys on social security fraud are of strictly limited value.

When surveying sensitive topics, researchers often use self-administered questionnaires, either in the form of a postal survey or as a paper questionnaire handed over by the interviewer. Compared to face-to-face interviews, self-administered questionnaires evoke a greater sense of

privacy and lead to greater self-disclosure (Sudman & Bradburn 1974; Tourangeau & Smith 1996). Empirical research has shown that compared to 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 & Devore 1992; for a review see De Leeuw 1992; Lee 1993).

Computer Administered Self Administered Questionnaires (CASAQ) make it possible to put very complex questionnaires without the intermediary of an interviewer. In these applications the interview program takes over the questionnaire logic and question flow. Respondents simply read each question from the screen, type in an answer, and are no longer burdened with complex routing instructions.

The visible presence of a computer may have an effect on the respondents' sense of privacy. De Leeuw, Hox and Snijkers (1996) distinguish two different 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 expectation 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.

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. For instance, Beckenbach (1992) reports the two following controlled laboratory experiments which compare traditional self-administered questionnaires with CASAQ. In the first, Martin and Nagao (1989) compared CASAQ with face to face interviewing and a traditional self-administered questionnaire; using the Crowne-Marlowe social desirability scale they found less social desirability bias in the self administered questionnaire, and even less with CASAQ. In the second, in which Evan and Miller (1969) compared CASAQ with a traditional self administered questionnaire, the authors found that CASAQ leads to more openness with questions that are perceived as threatening, and no difference with non-threatening questions. Beckenbach (1992) also conducted a small scale and well controlled study comparing CAPI, CASAQ and a paper-and-pen, face-to-face interview. After the interview, both interviewers and respondents filled in a questionnaire containing questions about the interview itself. Neither interviewers nor respondents reported problems with eye contact or social interaction. In the computer-assisted methods (both CAPI as CASAQ), respondents were more positive about data privacy, and found answering sensitive questions less unpleasant. In a direct comparison of a postal questionnaire and an electronic mail health-questionnaire Kiessler and Sproull (1986) also 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 Kiesler (1996) found that computer administration increases self-disclosure compared with paper-and-pencil (self-administered) forms.

The Randomized Response (RR) technique is an entirely different approach toward 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 later. We will describe RR techniques in more detail in the next section. Comparative research shows that RR techniques lead to more valid answers (cf. Umesh & Peterson 1991; Hosseini & Armacost 1993; Scheers 1992). One study (Locander, Sudman & Bradburn 1976) compares face-to-face interviews, telephone interviews, and postal 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.

In the study reported here, we compare two different varieties of RR, CASAQ, and 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 with the sensitive characteristic under study. However, they rightly argue that finding a higher proportion is insufficient proof of the validity of the RR technique. It is possible that, even though there is a higher proportion, there still are many false negatives, i.e. people who do not admit the sensitive characteristic, and false positives, i.e. respondents who boast of a sensitive characteristic that they do not in fact possess. They therefore propose validation studies to investigate the validity of the RR technique. In a validation study, respondents are interviewed about a sensitive characteristic for which their status is known to the researcher, so that it can be checked whether individual respondents answer truthfully.

All respondents used in our study had in fact been caught for social security fraud. The experiment was set up in such a way that the interviewers did not know that respondents had been caught for fraud, and the respondents did not know that the researchers had this information. Since the actual status of the respondents is known, it is possible to compare the results of the four approaches by comparing the percentage of false negatives (respondents that are caught for fraud but do not admit this).

On the basis of the comparative studies cited above, we hypothesize that the RR versions will register a higher proportion of respondents admitting fraud than the self-completion version, and that the self-completion version will outperform the standard approach.

Many earlier studies made use of students, who are known to be experienced respondents. In this study, however, the respondents in our sample are of below average education. The sample also contains relatively many people who were not born in the Netherlands and do not possess native mastery of the Dutch language. This also makes this

study interesting, because it shows how RR behaves in a realistic setting. For this reason we will also pay particular attention to the interview process.

Two additional questions concern which respondents are willing to admit to having practised fraud, and whether the respondent characteristics that predict positive responses to the sensitive questions are the same for all methods. To examine these questions, we need to relate the available explanatory variables to the yes/no response. For the direct question method and for CASAQ this is straightforward logistic regression. For RR, this is more complicated, because we never directly observe an individual response. Many people erroneously think that RR can only produce estimates of the proportion of people having a sensitive behavior, but that one is unable to relate the answers on RR to covariates. This is incorrect: it is possible to develop logistic regression models when the dependent variable is the RR variable. Pioneering work has been done here by Maddela (1983) and by Scheers and Dayton (1988), but this work seems to have gone largely unnoticed (see, for a recent contribution, van der Heijden and van Gils, 1996). In this paper we will use a number of explanatory variables to predict whether correct answers are given to our sensitive questions, using logistic regression for the observed responses in the direct question and the CASAQ approaches, and the adapted logistic regression for the two RR procedures.

In section 2 will we give more technical details about the two RR approaches used. In section 3 we will describe the experimental procedure, such as the operationalizations of the questions and the methods of analysis. In section 4 we will describe the results. We end with a discussion.

2. Two randomized response techniques

Because we expect many readers not to be familiar with the specifics of RR, we start with an exposition of the procedure originally proposed by Warner (1965). In its original form each respondent is given two questions or statements that are mutually exclusive. For example, in the context of social security fraud the two statements could be:

A. I reported all extra income last year

B. I did not report all extra income last year

A randomizing device, for instance a die, then determines which statement the respondent has to respond to. The outcome of a throw with the die is only seen by the respondent. The interviewer only receives the answer "yes, this is correct" or "no, this is incorrect". Assume that the rule concerning the randomizing device is that the respondent responds to statement A when the outcome of the die is 1 or 2, and B when the outcome of the die is 3, 4, 5 or 6. Let π be the probability of a positive answer to sensitive question B. Then the probability of obtaining the answer "yes", P(yes), is

$$P(\text{yes}) = (1 - \pi)\, P(A) + \pi\, P(B)$$

For this choice of the outcomes of the randomized device P(A) = 1/3 and P(B) = 2/3 . This shows that it is possible to estimate π from the sample estimate for P(yes).

The sampling variance associated with the estimate π is Var $(\pi) = \pi(1 - \pi)/n + \{P(A)[1 - P(A)]/n[2P(A) - 1]^2\}$. The first part of this formula, $\pi(1 - \pi)/n$, is the usual sampling variance for a binomially distributed outcome. The second part, $\{P(A)[1 - P(A)]/n[2P(A) - 1]^2\}$, can be interpreted as a penalty, imposed as a consequence of introducing a random event in the response process. Note that if only question B is asked of the respondent, P(A) = 0, and the second part of the formula becomes zero. Also note that the closer P(A) is to 0.5, the larger the extra term $\{P(A)[1 - P(A)]/n[2P(A) - 1]^2\}$ becomes. If P(A) is 0.5, the RR procedure is completely uninformative. Thus, the RR procedure is the least informative when P(A) is close to 0.5, and the most informative when P(A) is close to 0 or 1. It is clear that Warner's RR procedure is less efficient than the standard way of asking questions: larger sample sizes are needed to reach the same precision as in the standard situation. On the other hand, the idea of using the RR procedure is that the estimate for P(yes|B) is less biased than the estimate for P(yes|B) in the standard interview situation.

Since the original proposal by Warner (1965), many alternatives have been worked out that are, firstly, statistically more efficient than Warner's proposal, and, secondly, less threatening to respondents (see Fox and Tracy 1986; Chaudhuri and Mukerjee 1988). The threatening nature of Warner's procedure is that both statements deal with the sensitive topic and respondents could think that there is a mathematical trick to sort out their real status. We give two examples of such alternatives here, the RR procedures that we have used in the current study.

The first method is the so-called forced alternative method (see Fox and Tracy 1986, p.24). Here the randomized device determines whether the respondent is forced to say "yes", or "no", or answer the sensitive question. For example, in this study we asked the respondent to roll two dice, and answer "yes" if the outcome of the sum were 2, 3 or 4, answer "no" if the outcome were 11 or 12, and answer the sensitive question with "yes" or "no" if the sum were between 5 and 10. For the probability of answering the sensitive question we followed 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). 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 let π be the probability of answering this sensitive question positively. So $P_1 = 1/6$, $P_2 = 1/12$, and $P_3 = 3/4$. Then

$$P(\text{yes}) = P_1 + P_3\,\pi$$

and a sample estimate for π can be derived from the sample estimate for P(yes). Fox and Tracy (1986, p.38) mention the following advantages of the forced alternative approach over other RR approaches. First, it is quite efficient. Secondly, the method is comparatively easy for respondents to comprehend. And thirdly, use can be made of the observation of Moriarty and Wiseman (1976) that respondents tend to overestimate P_1 and P_2, i.e., for the example above where $P_1 = 1/6$ and $P_2 = 1/12$, they perceive this probability to be larger than 1/4, and therefore they think that they are "safer" than they actually are.

However, Fox and Tracy (1986; p.39) also mention the problem that sometimes respondents refuse to say "yes" when the randomizing device directs them to do so (cf. Edgell, Himmelfarb and Duchan 1982). For this reason we also employ a procedure proposed by Kuk (1990), where respondents only have to answer the color of a card. In Kuk's procedure there are two stacks of cards. In the left stack the proportion of red cards is $P_1 = .8$, and in the right stack it is $P_2 = .2$. The respondent is asked to draw one card from each stack. Then the sensitive question is asked, and when the answer is "yes", they should name the color of the left stack, and when it is "no", they should name the color of the right stack. Thus π = P(left) is the probability of fraud. It follows that

$$P(\text{red}) = P_1\pi + P_2(1 - \pi).$$

The sample estimate for π can be derived from the sample estimate of P(red). In the current research we used $P_1 = .8$ and $P_2 = .2$.

3. Method

3.1 Respondents

The experiment reported here was part of a larger interview investigating how clients of the social security department make ends meet, and how they evaluate the social security department. We could make use of data bases in three Dutch cities with addresses of people found guilty of social security fraud between 1991 to 1994. We concentrate here on income fraud, because for other types of fraud the number of respondents was too small.

It is important to note that every respondent found guilty of fraud received at least either an administrative sanction or a cut in their income, making it highly unlikely that the discovery had passed unnoticed by a respondent. The respondents were approached in the second half of 1995. It is therefore assumed that respondents still remember having experienced this administrative sanction or cut. However, cases of fraud are not always clear-cut. For example, social security officers consider getting a job and not reporting this on time as fraud, although respondents who had done this did not always necessarily intend to commit fraud, and might therefore consider themselves as "clean" cases. Another problem is cases where more than one type of fraud has been assessed. The social security department usually registers the most

important type, but what they consider the most important one is not necessarily the one considered important by respondents. We assume that the number of cases for whom this holds is small.

As a first step, the social security department mixed the addresses of respondents identified as having committed fraud with addresses of respondents not so identified. This sample then received a letter from the local social security department stating that they might be approached with the request to take part in research about making ends meet while receiving social security money. They were also told that new methods of data collection would be used in this research. The respondents were given the possibility of sending a card indicating that they did not want to participate (the so-called 'passive consent' procedure; for 'active consent', not employed in this study, respondents have to send a card stating that they wish to participate). The respondents were informed that the information that they provided in the interviews would not be reported to the social security 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 a commercial interview bureau (NIPO), which did not know which numbers referred to respondents caught for fraud. Once the interviews had been carried out by this bureau, the researchers received the data files without the addresses, and the social security office provided the key showing which respondent numbers referred to respondents caught for what type of fraud. Thus none of the three parties involved (i.e. researchers, interview bureau and the social security department) had access to all three sources of information: addresses, answers and respondents caught for fraud.

The interviewers were not told that there were respondents known to have committed fraud. At the same time, respondents known to have committed fraud were not told that the researchers eventually would get this information.

The respondents were randomly distributed over the four conditions of the experiment.

Although there is quantitative information about social security fraud from respondents not caught for fraud, we do not report this here because it is not the scope of this study. We are reluctant in giving fraud estimates for this group at this stage but rather wait untill we have results on a nationwide survey that is currently running (see also the discussion section). Therefore, in section 4 we concentrate on the respondents who are caught for fraud and investigate the validity of their answers, except when we discuss the interview process.

A description of the response of the total group of respondents (i.e. those caught and not caught) can be found in table 1. The total sample approached by the social security department announcing the interview was 1774; 20 % of those indicated their refusal to participate by sending in the reply card (passive consent); thus the interview bureau was provided with 1418 addresses (c) by the social security department. For various reasons, such as incorrect addresses, a lack of known telephone numbers, and a hot summer that led to many respondents to absent themselves from home, the field work took much longer than planned. To reduce

Table 1: Description of response. Column 1 is sample size; column 2 gives percentages w.r.t. total sample approached; column 3 gives percentages w.r.t. individuals who did not refuse to participate; column 4 gives percentages w.r.t. individuals who were approached.

	n	%	%	%
a. sample approached	*1774*	*100*		
b. refusal by reply card	356	20		
c. no refusal	*1418*	*80*	*100*	
d. not used, or in process when stopped	221	12	16	
e. approached (maximally five times)	*1197*	*67*	*84*	*100*
f. approached, no contact after five times	406	23	29	34
g. contact, but no interview	196	11	14	16
h. response	*595*	*34*	*42*	*50*
i. partly successful	35	2	2	3
k. cleaned by researchers	26	1	2	2
l. used for analyses	*534*	*30*	*38*	*45*

costs, the field work was terminated after five months, at which time 221 addresses (d) were either not used or had not yet been approached five times. As a result, it is impossible to determine a clear response rate in terms of the sample approached (phase (a)). The remaining 1197 addresses were approached up to five times, until contact was made. For 406 addresses, we were unable to contact the respondent; in 35 % 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 % of 196) and refusal after an appointment had been made (13 %); refusal to participate was not related to level of education. This left 595 respondents, of which 35 interviews were only partly successful, and 26 interviews had to be deleted by us because interviewers' reported incoherent answers or behaviour. This left us with 534 interviews. In terms of the total sample approached by a letter the response is 30 %; in terms of the number of individuals who did not use the reply card for refusal, the response is 38 %; in terms of the number of individuals approached five times, the response is 45 %. 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 the 426 of the 534 interviews were realised for respondents identified as having committed fraud. Non-response in phase (g) was not significantly related to having been found guilty of fraud.

The educational level of the 534 is low for Dutch standards: 35 % reported having received only primary education, 34 % only lower vocational training, 16 % middle vocational training, 4 % higher general education or secondary education preliminary to university, and 5 % vocational high school or university.

3.2 Questionnaire

The questionnaire began with questions dealing with country of birth, length of stay in the Netherlands, knowledge of the Dutch language and educational level attained; questions about job history, and job orientation; questions about their income and about whether they were able to make ends meet; questions on their relation with the social security department and their attitude towards the Dutch system of social welfare. The interview began with these questions in order to foster an atmosphere that would encourage respondents to answer the sensitive questions honestly.

In the CASAQ the respondents were offered the questions from the screen and asked to type in their answers. The interviewer was unable to see the screen.

The exact text with which the RR procedures were explained can be found in van Gils, van der Heijden and Landsheer (1996). Before asking the sensitive question that concerns us here, the procedures were practised by offering some questions of a less sensitive nature. In those cases where respondents had extreme difficulties in using the CASAQ or the RR procedures, they were moved to the direct question condition of the experiment.

3.3 Operationalizations

In the questionnaire the term "fraud" was carefully circumvented. The sensitive question was phrased as follows: "have you ever failed to declare part of your income to the social security department as you are required by law? [This might be income from work, earnings on the side, gifts, maintenance and such like.]"

For other operationalizations we refer to van Gils et al. (1996).

3.4 Analysis

We restricted the analyses to those respondents known to have committed fraud, because we could verify the answers of these respondents. For respondents not known to have committed fraud this is more ambiguous, because not having been caught does not necessarily mean that they have never practised fraud. Thus the interpretation of the dependent variable on income

fraud has a clear interpretation as 'admitting fraud' with answers yes/no. The way to estimate the probability of respondents admitting their fraud has been described in section 2.

We test for differences associated with the four experimental conditions by using the approximation of the binomial distribution by the normal distribution, and by using the standard errors for the proportions of respondents admitting fraud.

The logistic regression model is a natural candidate to relate the responses on a dichotomous dependent variable to covariates. For the RR approaches we have adjusted the logistic regression model to take care of the fact that the responses on the dependent variable are related in a known way to the sensitive question. For the forced response model this can be worked out as is worked out in Maddela (1983; p.54-56). Van der Heijden and van Gils (1996) also work out an adjusted logistic regression model for Kuk's procedure. We give a short outline of these adjustments.

Let $\pi_{j|i}$ be the probability of giving answer j (j=1,2) as a function of covariate vector i. In this context assume j=1 means 'no' and j=2 means 'yes'. Let the k'th covariate value in covariate vector indexed by i be denoted by x_{ik}, where x_{ik} is either a continuous variable or a dummy-coded variable. Let β_k be the regression parameter determining the relation between covariate k and response j. Then the logistic regression model is defined as

$$\pi_{j|i} = \frac{1}{1 + \exp \sum_k x_{ik}\beta_k} \text{ if } j=1 \qquad (1)$$

$$\pi_{j|i} = \frac{\exp \sum_k x_{ik}\beta_k}{1 + \exp \sum_k x_{ik}\beta_k} \text{ if } j=2 \qquad (2)$$

Let n_{ij} be the number of responses j for the observations with covariate vector i. Then the loglikelihood for the model is

$$\log L = \sum_i \sum_j n_{ij} \log \pi_{j|i} = \sum_i n_{i1} \log \pi_{1|i} + \sum_i n_{i2} \log \pi_{2|i} \qquad (3)$$

This loglikelihood can be maximized over the parameters β_k.

The adjustment made for the forced response is as follows. The interpretation of $\pi_{j|i}$ remains the same, and is defined in (1) and (2). In section 2 we found equations for P(yes) and P(no) that we now adjust for covariates into $P_i(\text{yes}) = P_1 + P_3\pi_{1|i}$, and $P_i(\text{no}) = P_2 + P_3\pi_{1|i}$. Thus the loglikelihood for the RR data becomes

$$\log L = \sum_i n_{i1} \log (P_2 + P_3\pi_{1|i}) + \sum_i n_{i2} \log (P_1 + P_3\pi_{2|i}) \qquad (4)$$

and this loglikelihood is to be maximized over the parameters of the logit model for $\pi_{j|i}$. The adjustment made for Kuk's procedure is similar. Details for both adjustments can be found in van der Heijden and van Gils (1996).

Estimates for the regression coefficients from adapted logistic regression can be interpreted in the usual way, as though no RR question were involved. It is possible to estimate the probability of fraud for specific values of the covariates using equations (1) and (2). It should be borne in mind that, for the RR approaches, the standard errors of the regression coefficients are rather large, due to the fact that the sample size is rather low, the number of variables included in the equations is comparatively large, and RR generates considerable error on the dependent variable. Another problem is that in the RR approaches the logistic regression model is numerically less stable, which means that the number of covariates that can simultaneously be included in the model is smaller than in the standard situation. Therefore the number of covariates included in the analyses is limited.

In order to compare logistic regression results for each of the four methods, we also report a combined estimate plus standard error for the effect of the covariates in all four experimental groups, and a chi-squared test for the significance of the variation of the regression coefficients across the four groups. Regression coefficients were combined and their variance tested using standard random coefficient meta analysis techniques (Raudenbush 1994).

4. Results

4.1 Process

In general, the interviewers evaluated the interviews as going smoothly. In 80 % of the interviews language problems played no role, but 5 % of the respondents did have serious problems with the Dutch language. In 90 % of the cases co-operation in the interviews was evaluated by the interviewers as good or very good, with no significant differences between the four methods. The attitude of more than 75 % of the respondents was evaluated as unsuspicious, with an overrepresentation of suspicion for the forced alternative procedure.

The interviewers indicated that 17 % of respondents found it difficult to understand the forced alternative procedure; for Kuk's procedure this was 21 %. After the introduction and the first explanation of the RR procedures 70 % of the respondents were willing to cooperate. The other respondents made critical remarks, but after a second explanation they could be persuaded to cooperate; 18 % of all respondents would rather answer the sensitive questions straight away, and 9 % declared they found the RR procedure nonsensical. In the RR procedures 72 % of the respondents were rated as not distrustful, 23 % as a little distrustful and 3 persons indicated

Table 2: Proportion of respondents admitting income fraud. One-sided z-tests are carried out against the standard method ($\alpha = .05$ corresponds to $z = 1.65$)

Method	N	P(fraud)	s.e.	z	p
Forced alternative	96	.43	.068	2.22	.01
Kuk	105	.49	.082	2.59	<.01
Self-completion	47	.19	.058	-.83	.20
Direct	99	.25	.044		

thorough distrust of the procedure. In the direct question approach 83 % of the respondents were not distrustful and 12 % were a little distrustful (difference between the RR procedures and the direct approach is significant). For the self-completion approach, the co-operation and trustfulness could not be rated because the whole interview was filled in on the computer. However, 30 % of the respondents required help with this.

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

If respondents had extreme difficulties mastering the RR or CASAQ procedure, they were allowed to switch to the direct question approach. The number of changes was largest for CASAQ (22 %), whereas for the forced alternative method and Kuk's procedure it was much smaller (9 % and 5 % respectively). CASAQ produces significantly more switches than both RR techniques ($p = .00$), which did not differ significantly from each other ($p = .13$). There is a small but not insignificant effect of overrepresentation of lower education levels among the switchers.

4.2 Proportions of respondents admitting fraud

The proportions of respondents admitting fraud are relatively low, ranging between 19 % and 49 % (see table 2). The RR procedures performed significantly better than both the direct question approach and the CASAQ approach. Kuk's procedure did best, with 49 % of the respondents admitting fraud, followed by the forced alternative approach with 43 % (the difference between the two RR procedures is not significant). Contrary to our expectations the CASAQ approach gave even lower figures than the direct question approach, with 19 % against 25 % (this difference is not significant either).

4.3 Logistic regression analyses

The multivariate logistic regression analyses are presented in two blocks in table 3. Analysis of all variables in one block was not possible because one of the RR models would not then be identified. In the first block of variables we have background variables like gender and age and variables that are possibly related to understanding, i.e. educational level, being born in the Netherlands, and control of Dutch language. In the second block we find social security benefit variables, i.e. the type of benefit and the duration, and psychological variables, i.e. the attitude towards the social security department, the perception of fraud, and whether the questions are considered an annoying intrusion. Positive b-estimates in table 3 indicate that higher values of this variable predict more admission of fraud, and negative estimates predict less admission. For example, for the variable Gender we find always negative estimates for b. It follows that higher values of Gender (in this case: females, since 1 = males and 2 = females) lead to fewer admissions. For reasons of space we discuss only the overall trends in the b-estimates.

Starting with the first block, the background variables Gender and Age show over all methods that women and older respondents are less willing to admit to fraud. The variables possibly related to understanding generally have no relation to admission of fraud, except being born in the Netherlands: individuals not born here admit fraud less often when asked directly. Interestingly, the other methods seem to protect, since there the relation has disappeared. There is also a significant relation that a better control of the Dutch language leads to more honest answers in Kuk's procedure.

In the second block, the three benefit variables should be interpreted with respect to the reference category "no longer receiving benefit". There is a general effect; unemployed respondents are less willing to admit fraud. For the direct approach we find an additional significant effect of still receiving health benefit, in the sense that these individuals are less inclined to admit fraud. For the psychological variables we find no general effect. In the forced alternative approach people who claim that they find the questions annoying admit less fraud, and for the direct approach people who have a negative attitude towards the social security department are less inclined to admit to fraud.

There is no strong evidence that the covariates have different effects in the different procedures. The chi-squared tests are all non-significant. Since we are comparing only four conditions, the power of this test is very low. If we interpret a chi-square value larger than its degrees of freedom as an indication of possible variation across the procedures, then we have to conclude that not being born in the Netherlands, the duration of welfare benefits, attitude towards the social security office and perception of fraud may act differently in different response procedures. In general, the Kuk methods seems to be least affected by the available covariates.

5. Discussion

This study has a number of strong points. The population studied is generally seen as difficult to approach, and as such this study provides a test of the performance of the questioning techniques in a difficult group. Secondly, since we know that the respondents have been confirmed as engaging in fraud, the study is a validation study that allows the detection of false negatives. Thirdly, four different conditions for eliciting the sensitive information are compared. The level of non response is a weak point of the study, which may be partly due to the fact that respondents were approached by the social security department.

The results are not comforting. Although the RR approach performs much better than more traditional approaches, the percentage of respondents admitting to fraud is far less than 100 %. Why are the proportions of respondents admitting their fraud so low? These percentages are especially low when we keep in mind that, from a legal point of view, their cases are already closed. One possibility is that some of those respondents not admitting to fraud believe that they did not really commit fraud. Some reasons for this were given in section 3. A comparison of respondents who still receive social benefit to respondents who are no longer in receipt of these benefits, 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 CASAQ or the direct question method. CASAQ did not perform well at all. This is somewhat surprising, given the positive results for CASAQ in the meta-analysis by Weisbrand & Kiesler (1996). Our different result may be due to questioning a difficult group of respondents. As mentioned above, in this special group of persons who (1) had received social security benefits, and (2) had been found guilty of fraud, respondents had a much lower educational level than the general population. Most studies synthesized by Weisbrand and Kiesler looked at either the general population or a 'well-behaved' special population such as students. It is suggestive that in our population 22% of the persons placed in the CASAQ condition had to be moved to the direct question condition because they could not cope with the CASAQ procedure because of, e.g., computer-related problems or illiteracy. In contrast, in the RR conditions only five to nine percent of the respondents had to be moved. We suggest that the CASAQ method for eliciting sensitive information works well only with populations that are at least moderately familiar with computers.

If we compare the two RR techniques, Kuk's method seems to perform slightly better. Most comparisons between the forced choice technique and Kuk's method show no significant differences, but in all cases the direction of the difference favors Kuk's method. Slightly more respondents were able to work with Kuk's method (95% compared to 91%), and respondents felt less suspicious of Kuk's method. In addition, Kuk's method results in slightly more positive responses (49% versus 43% with the forced alternative).

Table 4. Results CAPIBUS research. Percentage of respondents admitting to questions 1 to 7. For questions, see text.

Question	Direct		Kuk	
	%	n	%	n
1. ABN AMRO most important bank?	60	68	60	90
2. Do you have a car?	67	233	68	293
3. Personal loan	16	230	30	292
4. Tax fraud	3	122	15	179
5. Alcohol	9	232	17	292
6. Extreme right winged party	0	180	0	291
7. Social security fraud	3	70	20	90

Since Kuk's method appears promising, we investigated the practicality of including it in a standard survey of the general population. Consequently, we included a number of RR questions following Kuk's method in CAPIBUS, a weekly sample of 2000 respondents on different topics carried out by the NIPO. After weighting this sample is known to be representative for the Dutch population. For data collection, many interviewers were used who had received no special training for working with Kuk's method. The number of problems encountered was low, and the interviewers judged the understanding of respondents in 66 % of the cases as complete, in 20 % as approximate, 6 % partial, 3 % low and 5 % "don't know".

Neutral and sensitive questions were used. Neutral questions were 1. Is ABN AMRO for you the most important bank? and 2. Do you have a car? Sensitive questions were 3. Do you have a personal loan of more than DFL 1000?, 4. Did you give the tax authorities purposely incorrect information in the last three years?, 5. Do you drink three or more glasses of alcohol daily, on average?, 6. Did you vote for the last national election for the CD or CP'86? (CD and CP'86 are extreme right winged political parties, often accused of propagating racist beliefs), and 7. Did you ever purposely provide incorrect information to social security offices, or commit benefit fraud? The results are shown in table 5.

Kuk's approach seems to do well here. On the two non-sensitive questions the proportion of respondents "admitting" is as high for the direct question approach as it is for Kuk's approach. For the sensitive questions Kuk's approach gives significantly higher proportions than the direct approach, except for question 6. For question 3 it is also known from the Dutch Bureau for Credit Registration that the proportion of people having a loan is

indeed 30 %, which provides a global validation for our result (in a global validation it is possible that this is due to under- and overreporting cancelling each other out). It is possible that this is due to the fact that having a personal loan is much less sensitive than fraud. For question 6 the proportion in the last national election was indeed around zero, which result therefore shows that respondents have followed the instructions correctly.

The CAPIBUS results are much more comforting than the results from our validation study. It turns out that Kuk's RR-method is relatively easily explained to samples of the general population, even if the interviewers did not receive special training. It seems that here the validity of results is much increased over direct questioning. It is true that this increase in validity has its price in terms of the larger sample size needed to get the same power (power calculations can be done using formulas for standard errors reported in Kuk, 1990), but it seems to us that direct questioning is no alternative when quantitative information on sensitive topics is needed.

The authors are still involved in the further development and use of RR procedures. For example, currently a nationwide study into fraud is conducted by the authors for the Ministry of Social Affairs and Employment where respondents are approached using a two-step procedure. Roughly, this two-step procedure is as follows: using CAPIBUS we collect a pool of respondents who receive unemployment benefit. Those respondents are asked if they can be approached for further research. If the answer is positive, they are then interviewed about fraud using Kuk's procedure. We have optimized Kuk's procedure in terms of understanding because it turned out that there is a strong relation between answering truthfully and understanding of Kuk's procedure (see Landsheer, van der Heijden and van Gils, 1997, for details). This two-step approach has the advantage that we do not approach respondents via the social security department. We also work on the development of correction factors to correct for the underreporting of fraud. We will report on this in later publications.

A referee asked us to comment on the ethical aspects of our validation study, especially on the fact that respondents were not told that they were taking part in a validation study. This is not unusual in validation studies, and as far as we know, this did also not happen in the validation studies discussed in the introduction. However, many precautions were taken to guarantee the individual privacy of respondents (see section 3.1), and in our view a respondent is not individually hurt by the fact that he or she is not told that the information he or she provided is validated against official sources after anonimization of his or her questionaire. In one of the three cities the client council of the social security department gave permission for this procedure (in the two other cities it was not necessary to obtain this permission). All these procedures are, in fact, in accordance with the section on "Privacy and confidentiality" on the American Psychological Association's "Ethical principals of psychologists and code of conduct" (APA, 1992). In all, our opinion is that, first, the existence of data collection methods that provide valid information about sensitive topics such as fraud is important for Dutch society;

second, in order to assess whether RR is indeed promising in this respect validation studies are necessary; and third, there is no other realistic way to do such a validation study than the way we chose, but we are open to suggestions for alternative ways to conduct validation studies.

References.

APA (1992). *Ethical principles of psychologists and code of conduct.* Available: http://www.apa.org/ethics/code.html.

Aquilino, W. (1994). Interview Mode Effects in Surveys of Drug and Alcohol Use. *Public Opinion Quarterly, 58,* 210-240.

Beckenbach, A (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, A method of the future? An experimental study of the use of a computer by self-administered questionnaires and face to face interviews]. Ph.D. thesis. Universität Mannheim.

Beckenbach, A. (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 M.M. Kommer (1982). *Effecten van voorlichting en controle: een experiment bij de motorrijtuigenbelasting.* Den Haag: Ministerie van Justitie, WODC.

Chaudhuri, A. and R. Mukerjee (1988). *Randomized response. Theory and Techniques.* New York: Dekker.

De Heer, W. (1996). International response trends. Paper, 4th International Social Methodology Conference, Colchester, UK, July 1996.

De Leeuw, E.D. (1992). *Data quality in mail, telephone, and face-to-face surveys.* PhD Thesis, Vrije Universiteit, Amsterdam.

De Leeuw, E.D., Hox, J.J. & Snijkers, G. (1995). The effect of computer assisted interviewing on data quality; A review. *Journal of the Marketing Research Society,* 37, 325-344.

Edgell, S.E., S. Himmelfarb and K.L. Duchan (1982). Validity of forced response in a randomized response model. *Sociological methods and research, 11,* 89-110.

Elffers, H., H.S.J. Robben and D.J. Hessing (1992). On measuring tax evasion. *Journal of economic psychology, 13,* 545-567.

Elffers, H., H.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. Rotterdam: Erasmusuniversiteit, Faculteit der Rechtsgeleerdheid.

Evan, W M & Miller, J R (1969). Differential effect of response bias of computer vs. traditional administration of a social science questionnaire: An exploratory methodological experiment. *Behavioral Science*, 14, 216-227.

Fox, J.A. and Tracy, P.E. (1986). *Randomized response. A method for sensitive surveys.* Beverly Hills: Sage, Quantitative Applications in the Social Sciences.

Hessing, D.J., H. Elffers and R.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,* 405-413.

Hochstim, J.R. (1967). A critical comparison of three strategies of collecting data from households. *Journal of the American Statistical Association, 62*, 976-989.

Hosseini, J.C. & Armacost, (1993). Gathering sensitive information in organizations. *American Behavioral Scientist*, 36, 443-471.

Johnston, J. & Walton, C. (1995). Reducing response effects for sensitive questions: A computer-assisted self interview with audio. *Social Science Computer Review*, 13, 304-319.

Junger, M. (1989). Discrepancies between police and self-report data for Dutch racial minorities. British Journal of Criminology, Vol. 29, nr 3, 273-283.

Junger, M. (1990). Delinquency and Ethnicity. Deventer: Kluwer.

Kiesler, S & Sproull, L S (1986). Response effects in electronic surveys. *Public Opinion Quarterly*, 50, 402-413.

Koopmans, C.C. (1988) *De informele arbeidsmarkt, een onderzoek naar vraag, aanbod en prijzen. OSA voorstudie nr. V 27.* Den Haag: Ministerie van Sociale Zaken en Werkgelegenheid.

Kuk, A.Y.C. (1990). Asking sensitive questions indirectly. *Biometrika., 77,* 436-438.

Landsheer, J.A., P.G.M. van der Heijden and G. van Gils (1997). *Trust and understanding, two psychological aspects of randomized response. A study of a method for improving the estimate of social security fraud.* Utrecht: Department of Methodology and Statistics, Methods Series MS-97-3.

Lee, R.M. (1993). *Doing research on sensitive topics.* Newbury park, CA: Sage.

Locander, W., Sudman, S. & Bradburn, N. (1976). An investigation of interview method, threat and response distortion. *Journal of the American Statistical Association*, 71, 269-275.

Maddala, G.S. (1983). *Limited dependent and qualitative variables in econometrics.* New York: Cambridge University Press.

Martin, C L & Nagao, D H (1989). Some effects of computerized interviewing on job applicant responses. *Journal of Applied Psychology*, 74, 72-80.

Moriarty, M. and F. Wiseman (1976). On the choice of a randomization technique with the randomized response model. *Proceedings of the social statistics section of the American Statistical Association,* 624-626.

Raudenbush, S.W. (1994). Random effects models. Pp. 301-321 in H. Cooper & L.V. Hedges (eds) *The handbook of research synthesis.* New York: Russel Sage Foundation.

Scheers, N.J. (1992). A review of randomized response techniques. *Measurement and Evaluation in Counseling & Development,* 25, 27-41.

Scheers, N.J. and C.M. Dayton (1988). Covariate randomized response models. *Journal of the American Statistical Association, 83,* 969-974.

Siemiatycki, J. (1979). A comparison of mail, telephone, and home interview strategies for household health surveys. *American Journal of Public Health, 69,* 238-245.

Soeken, K.L. and G.B. Macready (1982). Respondents' perceived protection when using randomized response. *Psychological bulletin, 92,* 487-489.

Sudman, S. & Bradburn, N.M. (1974). *Response effects in surveys: A review and synthesis. Chicago: Aldine.*

Tourangeau, R., & Smith, T.W. (1996). Asking sensitive questions; the impact of data collection, question format, and question context. *Public Opinion Quarterly,* 60, 275-304.

Turner, C.F., Lessler, J.T., & Devore, J.W. (1992). Effects of mode of administration and wording on reporting of drug use. In: C.F. Turner et al (eds.) *Survey measurement of drug use: Methodological studies.* Washington DC: Government printing office.

Umesh, U.N. & Peterson, R.A. (1991). A critical evaluation of the randomized response method: Applications, validation, and research agenda. *Sociological Methods & Research,* 20, 104-138.

van der Heijden, P.G.M., and van Gils, G. (1996). Some logistic regression models for randomized response data. In: A. Forcina, G.M. Marchetti, R. Hatzinger and G. Galmatti, *Statistical Modelling. Proceedings of the 11th International Workshop on Statistical Modelling.* (pp. 341-348). Orvieto, Italy, 15-19 July 1996.

Van Eck, R. and B. Kazimier (1990). *Zwarte inkomsten uit arbeid: een nadere analyse van de uitkomsten van enquete-onderzoek.* Voorburg: C.B.S., Statistische onderzoekingen, M. 36.

van Gils, G., P.G.M. van der Heijden and H. Landsheer (1996). *Rapportage van uitkeringsfraude in surveys.* Den Haag: Ministerie van Sociale Zaken en Werkgelegenheid, werkdocument 43.

Warner, S.L. (1965). Randomized response: a survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association, 60,* 63-69.

Weisband, S. & Kiesler, S. (1996). Self disclosure on computer forms: meta-analysis and implications. In R. Bilger, S. Guest & M.J. Tauber (eds.). *CHI'96 Electronic Proceedings.* ACM.

Ontvangen: 01-10-1997
Geaccepteerd: 05-07-1998