J. R. Statist. Soc. A (2006)
169, *Part* 2, *pp.* 305–318

A validation of a computer-assisted randomized response survey to estimate the prevalence of fraud in social security

Gerty J. L. M. Lensvelt-Mulders, Peter G. M. van der Heijden and Olav Laudy

Utrecht University, the Netherlands

and Ger van Gils

Beleids Onderzoek en Advies, Utrecht, the Netherlands

[Received October 2003. Final revision September 2005]

Summary. In the Netherlands, there is a research tradition that measures fraud against regulations by interviewing eligible individuals using a survey. In these studies the sensitive questions about fraud are posed by using a randomized response method. The paper describes the results of a Dutch study into the consequences of replacing home interviews by trained interviewers with Internet-delivered interviews in a survey on fraud in the area of disability benefits. Both surveys used computer-assisted self-interviews with randomized response questions. This study has three goals: first to present the research tradition that makes use of randomized response, second to compare the results of home interviews and the Internet survey and finally to introduce an adapted weighted logistic regression method to test the relationship between the probability of fraud and explanatory variables. The results show that there are no systematic differences between modes of interview, either for estimates of the prevalence of fraud or for the identification of associated variables. These outcomes result in the conclusion that the Internet survey is a useful and cost-effective instrument for measuring fraud in a population, and that it is unlikely that replacing home interviews with the Internet survey will result in a significant break with tradition.

Keywords: Fraud; Logistic regression; Policy making; Randomized response; Social benefit

1. Introduction

In the Netherlands there has been growing political interest in fraud regarding regulations. The reason for this is that, in modern society, there are many rules and laws for which it is not always clear how often individuals follow them. This interest has grown even further because of two major disasters in the Netherlands, which claimed numerous fatalities. In May 2000 a firework explosion blew up part of Enschede, a medium-sized city in the east of the Netherlands, because rules for storing fireworks were not followed. And on New Year's Eve 2000–2001 many people died in a discotheque fire because fire regulations had not been followed.

To investigate fraud regarding regulations three directions can be taken. The first is to audit a random sample of individuals or organizations to which the rules pertain by using a mixture of research methods, like inspections of administrations and visual and physical inspections, thus checking whether individuals are complying with regulations. For many regulations,

Address for correspondence: Gerty J. L. M. Lensvelt-Mulders, Department of Methodology and Statistics, Utrecht University, PO Box 80140, 3508 TC Utrecht, The Netherlands.
E-mail: g.lensvelt@fss.uu.nl

 0964–1998/06/169305

Government officials regularly run such checks to enforce the rules. However, for large scale studies, using these officials is too costly. Moreover, it often does not lead to unbiased estimates of fraud, because these officials tend to focus on those individuals who are more likely not to comply with the rules, thus making the sample selective (Prinsen and Visser, 2000; Veldkamp, 1997). The second direction is to draw a random sample of individuals and to interview them. Here the problem is that violators are known to lie or refuse to answer, which will negatively bias the results. As a result official statistics that are thus obtained tend to underestimate the extent of fraud (Borkenau and Ostendorf, 1992; Buchman and Tracy, 1982; Dalton *et al*., 1997; Turnley and Bolino, 2001). The third direction is to convene a group of different experts and to ask them to estimate the prevalence of fraud on the basis of their expertise in the field. The problem that arises here is that experts tend to overestimate the prevalence of fraud because in their daily environment frauds are overrepresented, and therefore estimates tend to be too high.

Since there is great political pressure to obtain valid estimates for the extent of fraud in society the Dutch Department of Justice has funded a research programme to address these problems. Under this programme, we tried to optimize a tool for asking sensitive questions, called randomized response. The randomized response technique (RRT) was introduced by Warner (1965) to study sensitive topics more validly. The rationale behind Warner's randomized response method is that when a respondent's privacy can be absolutely guaranteed the tendency to refuse co-operation or to give non-incriminating or socially acceptable answers will decrease and thus the trustworthiness of the data will increase. In its original form the respondent must answer one of two statements. For example, statement A is 'I used hard drugs last year' and statement B is the complementary statement 'I did not use hard drugs last year'. The respondents use a randomizing device to decide whether statement A or B must be answered. Since the interviewer does not know the outcome of the randomizer he or she does not know to which statement the answer refers. Thus the respondent's privacy is protected, yet for the sample the probability of hard drug use can be estimated. Since Warner many different (in particular, more efficient) randomized response designs have been developed; see Chaudhuri and Mukerjee (1988) and Umesh and Peterson (1991) for a good overview.

To make the results of the survey more useful for policy makers an instrument explaining why people might perpetrate fraud was developed. This instrument is called the 'Table of 11' (T11) (Ruimschotel *et al*., 1996; Elffers *et al*., 2003). T11 is based on a rational choice framework by the economist Becker (1968) and a social psychological model that explains behaviour in cost–benefit terms (Fishbein and Ajzen, 1975). It consists of 11 dimensions (Table 1), which together identify those factors that promote or oppose fraud with welfare regulations. The T11 factors enable policy makers to initiate policies that reduce fraudulent behaviour. For instance in a study on fraud with the law on housing benefits it became clear that social norms and deterrence explained fraud better than a lack of knowledge of the rules or general conformity to rules (Elffers *et al*., 2003).

Elements of this approach were successfully used in research on the law on using chemicals for agricultural goals and the law on prevention against misuse of chemicals that can be used to manufacture drugs (Ministry of Justice, 1998), laws on the reuse of refuse in the building industry, the taxi driving law, the retail law, the law on consumer durables and the law on housing benefits (Elffers *et al*., 2003). In this long line of research a special place is taken by surveys into social security, and these will be the topic of this paper.

In 1996, a study was undertaken investigating the qualities of the randomized response method in measuring fraud in social security (van der Heijden *et al*., 2000). A sample of people who had already been caught for fraud was interviewed about their rule compliance. Thus every individual stating that he complied with the rules was not answering truthfully. In this study

Table 1. Table of eleven (T11)

T1: questions about the (lack of) knowledge and clarity of the rules
I know the rules that are applicable to my disability insurance benefit (DIB)
The rules about what to do and what not to do in order to receive a DIB are clear to me
T2: questions about costs and benefits of compliance with the rules
I think it is more beneficial to me not to follow the rules connected to my DIB
T3: questions about acceptability of specific rules
The rules connected to my DIB are very reasonable
It is reasonable that I have to inform the DSS† about holidays and stays abroad that last longer than 4 weeks
It is reasonable that I have to inform the DSS about extra income
It is reasonable that I need permission from the DSS to take up studies
It is reasonable that I have to inform my physician about changes in my health that influence my right to receive a DIB
It is reasonable that I have to inform the DSS about being hospitalized
T4: questions about norm conformity in general life with respect to other rules
One should always obey the law
T5: questions about informal control by the social environment
My environment will react very negatively when I transgress the rules
T6: questions about the perceived probability of being captured
The odds that I will be caught when I earn some extra illegal income are very great
The odds that someone will betray me when I earn some extra illegal income are very great
T7: questions about the amount of inspections
How do you assess the possibility that the DSS will inspect if you do not transgress the rules?
How do you assess the possibility that the DSS will inspect your administration?
T8: questions about inspection detail
If you have transgressed the rules, how do you assess the possibility that the DSS will find out about this rule transgression?
T9: questions about inspection activities
T10: questions about certainty of sanctions
How do you assess the possibility that people who transgress the rules will be punished?
T11: questions about severity of sanctions
What do you think is the punishment for not giving full and exact information to the DSS?

†DSS, Department of Social Security; in the Netherlands this department is responsible for the execution of the social security laws, i.e. DIB, unemployment allowance and welfare benefits.

four conditions were compared: face-to-face interviews with questions asked directly, computer-assisted self-interviews with questions asked directly and two randomized response conditions, one using Kuk's (1990) design and one using the forced response design (Boruch, 1971). The last two surveys were administered at respondent's homes by experienced interviewers. For questions about income fraud, the results unequivocally showed that using an RRT led to significantly more valid results than asking the questions directly. When randomized response was used, about 50% of the individuals answered truthfully, whereas for the other conditions this percentage was about 20% (see van der Heijden *et al.* (2000) for details). It was concluded that, although randomized response still led to an underestimation of fraud, it was useful to develop the randomized response procedure further. This decision led to the first wave in a regular survey

on unemployment benefits fraud, disability insurance benefit fraud, and social security benefit fraud in 2000. Kuk's card method was used as the randomized response design (Kuk, 1990; van der Heijden *et al*., 2000). This design is statistically comparable with Warner's RRT design, but psychologically stronger, because the incriminating 'yes' answer is avoided. Instead of answering 'yes' or 'no', the respondent must name the colour of a card.

In 2002 the second wave in this regular survey was to be undertaken, but three problems arose. The first problem was that computer-assisted self-interviews administered at respondents' homes by experienced interviewers were becoming too costly, because European regulations forbid the funding of interviewers on a provisional agreement for too long, making it more expensive to employ them. The second problem was that non-response percentages in surveys are becoming very high (De Heer, 1999), making the results less trustworthy. Finally, Kuk's card method proved to be statistically inefficient, which meant that very large samples were needed. Larger samples mean higher costs and more time to be invested in the fieldwork. These problems made it desirable to use a more efficient randomized response design in the 2002 wave.

To lower the costs per interview a computer-assisted self-administered randomized response Internet survey was developed (Couper and Nichols, 1998; Supple *et al*., 1999). To solve the problem of high non-response rates it was decided to study the possibilities of using a household panel of people who regularly fill in questionnaires for a marketing organization. Finally, it was decided to replace Kuk's card method with the more efficient forced response design (Boruch, 1971).

Such rigorous interventions could easily lead to a break with tradition; therefore it was necessary to investigate the quality of the results thoroughly. This paper therefore has three aims. The first aim is presenting the development of an instrument using the RRT to investigate fraud by using surveys. Then we shall present the statistical method that we used to study the relationship between the extent of fraud and the explanatory variables from T11. Because of the special features of the RRT it is not possible to analyse the data in a standard way. Therefore a weighted logistic regression was developed that took into account the different variance estimations of the RRT. And finally we shall present a comparison of the home interviews and the Internet survey.

We shall compare the estimates of the level of fraud as well as the identification of associated variables, because the Internet survey can only replace the home interviews when they lead to comparable results in terms of prevalence and in terms of the relationships between fraud and its associated features (T11). It is often argued that household panels consist of a biased sample of the general population (Vella, 1998). Using a household panel could result in higher estimates, because individuals who join a panel might be more willing to give honest answers and have fewer objections to protection of privacy. But it could also result in lower estimates, because individuals who join a panel might be more law abiding. Therefore, even when the Internet survey point estimates are comparable with or higher than the home interviews point estimates, it cannot safely be concluded that the Internet survey can replace the home interviews. There might still be undetected differences in the sample. To address this problem the relationships between the outcomes and various explanatory variables will be compared for the Internet survey and the home interviews. When the relationships between the prevalence of fraud and its explanatory variables are identical, this could be an indication that the samples are comparable.

In Section 2 an introduction will be given to disability insurance benefit, the sample, the randomized response design and the methods of analysis. Here, the survey method will also be presented. In Section 3 the results of the comparison of point estimates between the Internet survey and the home interviews will be presented, and the results of the adapted weighted logistic regression will be given. This paper will be concluded with a discussion.

2. Methods

2.1. Disability insurance allowance

This paper reports on disability insurance benefit fraud. Disability insurance benefit is part of the Dutch welfare system and consists of a monthly benefit for people who have been forced to retire from work because they have become permanently physically or mentally unable to perform their duties. Every employee who cannot perform his or her duties for these reasons for more than a year is entitled to this benefit. The allowance can amount to as much as 70% of the receiver's last regular income, dependent on their level of disability.

To remain entitled to disability insurance benefit, recipients must comply with rules for generating extra income and health-related behaviour. In our survey these rules were expressed in simple, non-incriminating and neutral terms so that all respondents could understand them (Dillman, 1991; Lee, 1993). Eight questions about possible fraudulent behaviour were asked: three questions were about income, four questions about health-related behaviour and the last question was intended to cover the whole range of fraudulent behaviour.

2.2. Respondents

To be recruited for the sample individuals had to receive disability benefit for at least 12 months before the study, so that they had time to become acquainted with benefit rules and have developed an attitude of (non-)compliance. All respondents were invited to participate in an interview about 'their contacts with the Department of Social Services'.

The home interview sample was drawn from the general population. Respondents were selected by using random telephone dialling. In the first telephone round (23 866 calls) individuals who had received disability benefit for over a year were located; 1700 respondents proved eligible (7%). In the second telephone round the 1700 respondents who satisfied the requirements for participation in the sample were invited to take part in a face-to-face interview at their own home. 850 eligible respondents accepted the invitation and joined the sample and this resulted also in 850 completed interviews, and a response rate of 50%.

The Internet survey sample was drawn from recipients of disability insurance who were listed in the Dutch Institute for Public Opinion and Market Research household panel, a well-known Dutch marketing research organization which is based in Amsterdam and works nation wide. The Institute has a household panel for conducting Internet surveys that contains over 35 000 households (100 000 individuals). A panel is always changing; new members enter the panel and old members leave. This makes it very difficult to compute the initial response rate, but in the literature response rates as small as 10% are mentioned (Groves *et al.*, 2004).

In the Netherlands 7.4% of the population of people 15 years old and older receives disability benefit (see Statistics Netherlands at `www.cbs.nl/statline`), compared with 6.1% in the panel. From the household panel we selected a random sample of 900 eligible individuals who obtained the Internet questionnaire. This resulted in 874 completed interviews (Van Gils *et al.*, 2003), and a conditional response rate of 97%.

The final samples had to be weighted for age and gender to make them comparable with the general population of individuals receiving benefit (Table 2). Age and gender distributions were obtained from Statistics Netherlands (`www.cbs.nl/statline`). As can be seen in Table 2 the differences in composition of the samples between the home interviews and the Internet survey are small. In the Internet survey condition women and younger people are a little over-represented.

In the Internet survey condition the panel protocol was followed. Questionnaires were sent to the respondents by using the Internet and respondents were allowed to answer the questionnaire

Table 2. Weighting variables

Variable	*Disability benefit (%)*		*After weighting (%)*
	Original sample, Internet survey	*Original sample, home interviews*	
Gender			
Male	52	55	55
Female	48	45	45
Age			
19–24 years	2	1	3
25–34 years	17	9	11
35–44 years	24	17	18
45–54 years	26	37	29
⩾55 years	32	37	39

at their own time and place, within 7 days of receiving the questionnaire. In the home interview condition an interviewer visited the respondents at home. When the randomized response questions came up the interviewers explained the randomized response method to the respondents and handed the computer to them. The interviewer then took up a position from which he could not see the screen, so that the respondent could use the virtual dice (see the next section) in privacy. The RRT questions were presented on the screen and the respondents entered the answers themselves. An interviewer was available for the respondent if he or she had difficulties with the program and the respondent had to answer the questions on an unfamiliar laptop. In the Internet survey condition respondents could take the survey from the Internet and fill it in on their own personal computer at a convenient moment.

2.3. The randomized response method, design and analysis

Both the Internet survey as well as the home interviews made use of a computer program to elicit answers to the randomized response questions. The first version of this program was tested in a pilot study using instruments from cognitive laboratories for testing survey questions (Snijkers, 2001; Sirken *et al*., 1999). All respondents found the program easy to use, even those with limited or no prior computer experience. Virtual dice that could be manipulated on the computer were preferred to real dice that had to be thrown manually. The RRT was found to be trustworthy and cheating (not following the rules of the RRT when forced to answer yes or no) and lying (not answering according to the truth when asked to) decreased (Boeije and Lensvelt-Mulders, 2002).

Because of its statistical efficiency, in 2002 Boruch's (1971) forced response design was used. In this design the questions are offered to the respondent together with two virtual dice that can be manipulated by using the 'enter' key. When the dice roll produce 2, 3 or 4, with probability $P_1 = 1/6$, the respondent is 'forced' to answer yes irrespectively of his own true answer to the question. When the dice produce 11 or 12, with probability $P_2 = 1/12$, the respondent is forced to answer no again irrespectively of his own true answer, and when the dice produce 5–10 the respondent is asked to answer the question truthfully ($P_3 = 3/4$). The advantage of using dice is that respondents underestimate the probability of their throwing of 5–10, and therefore feel safer than they really are (Fox and Tracy, 1986). The way in which the randomized response questions were presented to the respondents can be found in Appendix A.

Let π be the unknown probability of observing a yes response to the sensitive question. The probability of a yes response is $\lambda = P_1 + P_3\pi$. So, with the observed proportion as an estimate $\hat{\lambda}$ of λ, we can estimate $\hat{\pi}$ by

$$\hat{\pi} = \frac{\hat{\lambda} - P_1}{P_3} \tag{1}$$

with sampling variance

$$\text{var}(\hat{\pi}) = \frac{\hat{\lambda}(1-\hat{\lambda})}{nP_3^2} \tag{2}$$

(Boruch (1971) and Chaudhuri and Mukerjee (1988), pages 15–16).

To study the differences between the prevalence estimates for the different conditions a one-sided large sample z-test for differences in probabilities ($\hat{\pi}_{\text{Internet}}$ and $\hat{\pi}_{\text{home}}$) was used:

$$z = \frac{\hat{\pi}_{\text{Internet}} - \hat{\pi}_{\text{home}}}{\sqrt{\{\text{var}(\hat{\pi}_{\text{Internet}}) + \text{var}(\hat{\pi}_{\text{home}})\}}}. \tag{3}$$

To study the consequences of using a different sampling frame (the household panel) it was investigated whether the probability that people commit fraud under the Internet survey is related in the same way to explanatory variables as the probability of fraud under the home interviews. An adapted version of weighted logistic regression was developed for this purpose. First, the logistic regression had to be modified by using individual weights so that the outcomes are weighted towards the population characteristics in Table 2 and, second, the logistic regression had to be adapted to take into account that the dependent variable is a randomized response variable.

First, the standard logistic regression models with person weights will be defined. Let $\pi_{1|i}$ be the probability of answer no, and $\pi_{2|i}$ the probability of answer yes, both as a function of covariate vector i. Let the kth covariate value in the covariate vector that is indexed by i be denoted by x_{ik}, where x_{ik} is either a continuous covariate or a dummy coded variable. Let β_k be the regression parameter determining the relationship between covariate k and the response. The standard logistic regression model is defined as

$$\pi_{1|i} = \frac{1}{1 + \exp\left(\sum_k x_{ik}\beta_k\right)}, \tag{4}$$

and $\pi_{2|i} = 1 - \pi_{1|i}$.

Let w_i be the weight for person i and let (n_{i1},n_{i2}) be (1,0) for a no answer and (n_{i1},n_{i2}) be (0,1) for a yes answer. Then the log-likelihood is given by

$$\log(L) = \sum_i w_i n_{i1} \log(\pi_{1|i}) + \sum_i w_i n_{2i} \log(\pi_{2|i}). \tag{5}$$

This model will now be adapted for a dependent randomized response variable. A yes answer to the fraud question will have to be distinguished from a yes answer in the randomized response condition, where yes can be a forced or a true answer. Let $\pi_{2|i}$ be the probability that a person answers yes as the true answer. Then the probability of a yes answer is $\lambda_i = P_1 + P_3\pi_{2|i}$, and the probability of a no answer is $1 - \lambda_i = P_2 + P_3\pi_{1|i}$ (see Maddala (1983), Scheers and Dayton (1988) and van der Heijden *et al.* (2000)). Thus the log-likelihood for the model is

$$\log(L) = \sum_i w_i n_{i1} \log(P_2 + P_3\pi_{1|i}) + \sum_i w_i n_{i2} \log(P_2 + P_3\pi_{2|i}). \tag{6}$$

This log-likelihood is to be maximized over the parameters β_k. The first and second derivatives can be used to estimate model (4)–(5), with for example the Newton–Raphson algorithm. They are

$$\frac{\partial\{\log(L)\}}{\partial\beta_m} = \sum_i P_3 x_{im} \pi_{1|i}\, \pi_{2|i}\, w_i \left(\frac{n_{i2}}{\lambda_i} - \frac{n_{i1}}{1-\lambda i} \right), \tag{7}$$

$$\frac{\partial^2\{\log(L)\}}{\partial\beta_m \partial\beta_n} = \sum_i P_3 x_{im} x_{in} \pi_{1|i}\, \pi_{2|i} \times \left[(\pi_{1|i} - \pi_{2|i}) \left(\frac{w_i n_{i2}}{\lambda_i} - \frac{w_i n_{i1}}{1-\lambda_i} \right) - P_3 \pi_{1|i} \pi_{2|i} w_i \left\{ \frac{n_{i2}}{\lambda_i^2} - \frac{n_{i1}}{(1-\lambda_i)^2} \right\} \right]. \tag{8}$$

There are more statistical methods to handle randomized response data; for these methods we refer to Van den Hout and van der Heijden (2002, 2004).

3. Results

3.1. Fraud estimates

Table 3 displays the estimated prevalence of fraud in the population for the Internet survey and the home interviews. Because of the expected underreporting of sensitive behaviour it will be assumed that, in comparing estimates of the two conditions, the higher estimates will be the more valid estimates (Lee, 1993; Himmelfarb and Lickteig, 1982; Lensvelt-Mulders *et al.*, 2005). In Table 3, point estimates between the Internet survey and the home interviews differ significantly for two variables: for variable 3 'have you done some moonlighting over the last twelve months?' the home interviews provide a higher point estimate but for variable 7 'have you felt more capable to work without informing DSS over the last twelve months?' the Internet

Table 3. Estimates of the prevalence of fraud by question and condition†

Question	*Internet survey (N = 874)*		*Home interviews (N = 850)*		*Internet − home*	
	$\hat{\pi}$	*Standard error*	$\hat{\pi}$	*Standard error*	*Z*	*Power*
1. During the last 12 months done an odd job for friends or family and received money for this job	0.13	0.020	0.14	0.021	—	0.093
2. During the last 12 months a regular job next to allowance	0.00	0.017	0.00	0.017	—	0.113
3. During the last 12 months done some moonlighting	0.02	0.018	0.08	0.020	−2.3	0.590
4. Ever complaints less severe without informing DSS	0.02	0.017	0.02	0.018	—	0.494
5. Ever exaggerated gravity of illness	0.00	0.018	0.01	0.018	—	0.110
6. Ever felt recovery without informing DSS	0.06	0.019	0.03	0.019	—	0.192
7. Ever felt more capable to work without informing DSS	0.12	0.020	0.07	0.020	2.0	0.552
8. During the last 12 months transgressed rules connected to the right to receive an allowance	0.00	0.017	0.00	0.018	—	0.201

†DSS, Department of Social Services. (For the phrasing of specific items that were used interested readers can contact the first author). *Post hoc* power estimates are derived from simulation studies.

survey provides a higher estimate. The simulated power of the tests for the other variables is small, but this is not problematic because the differences between the point estimates for which the powers are derived are too small to be relevant. We conclude that there is no systematic bias of one data collection method over the other.

There are two remarks to be made on the size of some estimates in Table 2. First, for question 2 ('do you have a regular job next to your allowance?') estimates of almost 0 are found. This was to be expected because it is well known to recipients of allowances that the computer files of the Inland Revenue are linked to those of the Social Services Department and thus the probability of being caught in cases of fraud is very large. This question therefore lacks the power to discriminate between the home interviews and the Internet survey.

Second, since all questions are about fraud, the estimates for question 8 'Have you ever transgressed the rules?' should be at least as high as the highest estimate for the other questions, being 0.14 for the home interviews and 0.13 for the Internet survey. The very low estimates that were found may be a consequence of the fact that this question lacks validity. In general, sensitive questions should not only be asked in neutral non-incriminating words but also refer to well-defined situations (Dillman, 1991). Questions 1–7 do use neutral wording and refer to fraud against specific rules; however, this does not hold for question 8, which was non-specific and rather bluntly stated.

3.2. Explaining the prevalence of fraud by using logistic regression

Another way to compare the Internet survey with the home interviews is by studying the size and direction of the relationships between committing fraud and the T11 explanatory variables. This is done by using weighted logistic regression, adapted for randomized response questions. Those two fraud questions that have the highest estimates and largest variance in Table 3, namely question 1 and question 7, will be focused on. These questions have the additional advantage

Table 4. Admitting fraud in the context of disability benefits: results for the weighted logistic regression†

T11 variable	*General descriptive T11 values*			*Results for question 1, 'Have you ever done odd jobs?'*		*Results for question 7, 'Have you ever felt more healthy…?'*		
	N	*Mean*	*Standard deviation*	*Model 1 intercept*	b_1	*Model 1 intercept*	b_1	b_2
(1) Knowledge	1703	2.52	(0.950)	−1.981	−0.044	−2.865	−0.341	−0.768
(2) Benefits	1718	3.67	(0.777)	0.765	0.751	0.223	0.624	−0.698
(3) Acceptance	1723	2.23	(0.530)	−3.389	−0.658	−3.266	−0.566	−0.72
(4) Norm conformity	1708	2.53	(0.926)	−3.337	−0.546	−3.754	−0.553	—
(5) Social control	1604	2.74	(1.085)	−3.033	−0.422	−3.543	−0.443	—
(6) Probability of being caught	1661	3.16	(0.817)	−3.039	−0.379	−4.549	−0.714	—
(7) Control density	1724	3.45	(0.725)	−2.326	−0.130	−3.966	−0.567	−0.705
(8) Control detail	1724	2.99	(0.958)	−2.068	−0.063	−2.952	−0.322	−0.765
(10) Certainty of sanction	1703	2.53	(0.950)	−1.941	−0.024	−2.276	−0.119	−0.727
(11) Severity of sanction	1718	3.67	(0.777)	−1.807	−0.016	−0.789	−0.326	−1.774

†In the rows we find the T11 factors (given in full in Table 1). In the middle two columns we find the parameter estimates for fraud question 1. In the last three columns we find the parameter estimates for fraud question 7. Non-significant b_2-weights are not entered.

that they cover the income and the health-related areas. As explanatory variables the factors defined by T11 will be used.

We first investigated whether the relationships between the T11 factors on the one hand and the fraud questions on the other hand depend on whether the Internet survey or the home interviews are used. This was tested by using a conditional likelihood ratio test comparing the logistic regression model with main effects for each T11 variable and condition (Internet surveys *versus* home interviews; model logit(fraud) = intercept + b_1 T11 variable + b_2 condition) with a model with an additional interaction term between the T11 variable and condition (model logit(fraud) = intercept + b_1 T11 variable + b_2 condition + b_3 T11 variable * condition). For fraud question 1 as well as for question 7 none of these interaction terms turned out to be significant. There is no evidence that the strength of the association between the T11 factors and the fraud questions depends on whether the home interviews or the Internet survey is used and this is again an indication that home interviews and the Internet survey lead to comparable results.

Table 4 illustrates the results of the weighted logistic regression analysis. For fraud question 1 the T11 factors benefits (2), acceptance (3), norm conformity (4), social control (5) and the probability of being caught (6) have a significant association with committing fraud (b_1). Except for benefit, the estimated *b*-values for the T11 factors are all negative: for example, the higher the acceptance of the rules, norm conformity and the larger social control, the lower the probability of fraud. In contrast, the higher the expected benefit, the higher the probability of fraud is. In accordance with Table 3 we did not find significant b_2-weights: the conditions (home interviews *versus* Internet survey) did not affect the probability that respondents admit doing odd jobs.

For fraud question 7 the T11 factors knowledge (1), benefits (2), norm conformity (4), social control (5), probability of being caught (6), density of inspection (7) and detail of inspection (8)

Table 5. Admitting fraud in the context of disability benefits: interpretation of parameters†

T11 variable	*Results for question 1, 'Have you ever done odd jobs?'*			*Results for question 7, 'Have you ever felt more healthy or . . .?'*			*Difference conditions*
	p(fraud) − 1 standard deviation	*p(fraud)*	*p(fraud) + 1 standard deviation*	*p(fraud) − 1 standard deviation*	*p(fraud)*	*p(fraud) + 1 standard deviation*	
(1) Knowledge	NS‡	0.134	NS‡	0.16	0.119	0.09	−0.06
(2) Benefits	0.07	0.12	0.20	0.07	0.112	0.17	−0.05
(3) Acceptance	0.17	0.128	0.07	NS‡	0.119	NS‡	−0.06
(4) Norm conformity	0.19	0.124	0.08	0.14	0.087	0.05	NS‡
(5) Social control	0.19	0.133	0.09	0.14	0.089	0.06	NS‡
(6) Probability of being caught	0.18	0.137	0.10	0.15	0.092	0.05	NS‡
(7) Control density	NS‡	0.133	NS‡	0.17	0.118	0.08	−0.06
(8) Control detail	NS‡	0.132	NS‡	0.16	0.121	0.09	−0.06
(10) Certainty of sanction	NS‡	0.132	NS‡	NS‡	0.122	NS‡	−0.06
(11) Severity of sanction	NS‡	0.148	NS‡	NS‡	0.12	NS‡	−0.1

†Differences in the estimates in *p*(fraud) for the same question are due to differences in weights and valid *N* per T11 variable. The dummy variable for the Internet survey is 0 and for home interviews is 1.
‡NS, not significant.

have a significant association with committing fraud (the direction of these relationships will be discussed below in the context of Table 5). The mode of interview significantly influences the probability of fraud if an estimate for b_2 is displayed in Table 4.

The interpretation of the *b*-estimates is sometimes difficult to explain to policy makers and in this situation it is useful to fill in a logistic regression equation with some specific values of the T11 variable. For this the means, and the means plus or minus 1 standard deviation are used. Table 5 displays the results for those T11 variables that have significant relationships to fraud questions 1 and 7. For 'benefits' the estimated probability of committing fraud is 0.12 when respondents score at the mean for the benefits variable, when respondents expect to obtain more benefits from fraud (1 standard deviation) the probability of committing fraud becomes 0.20, when they expect to obtain fewer benefits from fraud (−1 standard deviation) the probability of committing fraud becomes 0.07. Similar interpretations hold for the relationship between the T11 factors and the probability that respondents do odd jobs. The relationship between T11 variables and the probability that respondents do not inform their physicians about their improved health is comparable, with the difference that the mode influences the magnitude of the estimates. For benefits the mean probability is 0.12. When a respondent expects to gain more benefits (+1 standard deviation) the probability of fraud becomes higher (0.17) and when he expects to obtain fewer benefits from fraud (−1 standard deviation) the probability will become lower (0.07). The estimates that are displayed in the last column of Table 5 (difference conditions) refer to the difference between the Internet and the home interview condition. In the home interviews condition all estimates become approximately 0.05 lower.

4. Conclusions and discussion

Until now, the Dutch Government has used home interviews with computer-assisted self-interviewing and an RRT to survey fraud. The results of individual validation studies have shown that using a randomized response approach resulted in more valid estimates than using conventional data collection methods (van der Heijden *et al*., 2000), although we are aware that the point estimates that are obtained with the randomized response approach are still underestimates of the true probability of fraud in the population.

To lower the overall costs of a regular survey making use of home interviews an Internet survey was developed and tested. The home interviews resulted in a significantly higher estimate when the respondents had to admit moonlighting, but the quality of the question on moonlighting can be questioned. In the Dutch questionnaire the phrase *zwart werken* is used, which can be also translated as 'working in the black market', which in its turn could be more emotionally charged than the English verb 'moonlighting'. So the quality of the question could be at the bottom of the different point estimates. The Internet survey in turn resulted in a significantly higher estimate when respondents had to admit that they felt healthier and more capable of working than they had led the Social Security Department to believe. We conclude that overall the Internet survey resulted in estimates of the extent of fraud that are comparable with those of the home interviews because, in the two questions in which they differed, they differed in opposite directions.

One of the goals of this study was to study the possibility of using an existing household panel instead of random telephone dialling to select a sample of eligible respondents. There were no large differences between the samples to begin with (Table 2). The point estimates for the prevalence of fraud seem comparable across modes and thus across samples. When the household panel members were for instance more law abiding the point estimates should have been lower.

But more important is that the relationships between the explanatory variables and fraud

were comparable across samples. A specially adapted weighted logistic regression was used to investigate this. The results of the logistic regression showed no significant interactions between the T11 variables and the conditions (Internet and home interviews), so both methods seem to trigger the same relationships between the extent of fraud and its explanatory variables. This is a strong indication that, after weighting, the samples are comparable when we want to study fraud in social security benefits. We conclude that using a household panel can be a good instrument to sample eligible respondents, without jeopardizing the external validity of the results.

Summarizing the household panel resulted in point estimates and equivalent relationships between the probability of fraud and explanatory variables that are comparable with those when a random digital dialling design was used to sample the respondents. These outcomes result in the conclusion that the Internet survey is a useful and cost-effective instrument to measure fraud in a population, and that it is unlikely that replacing home interviews with an Internet survey will result in a significant break with tradition.

Appendix A: Instructions for the Internet survey

The following instruction is taken from the Internet survey.

We would like to ask you some questions about your allowance. . ..

From previous research we know that many people find it really hard to answer questions about allowances, because they consider them a violation of their privacy. Some people fear that an honest answer might even have negative consequences for their allowance. But we do not intend to embarrass anyone. Therefore Utrecht University has developed a method to ask these questions in such a way that your privacy is absolutely protected.

You are about to answer the following questions with the help of two dice. With these dice you can throw any number between 2 and 12. Your answer depends on the number you have thrown. In this way your privacy is guaranteed, because nobody, neither the interviewer, nor the researchers, nor the social welfare authorities will ever know the number you have thrown, and thus they can never know why you gave the answer you did.

Now how does this work?
On your screen you see the dice rolling. By pushing the 'enter' button you will stop the dice from rolling. You can then directly see the number you have thrown. If you push the 'enter' button again, the question will appear on your screen.

If you throw 2, 3 or 4 you always push button 1 (= yes).
If you throw 11 or 12 you always push button 2 (= no).
If you throw 5, 6, 7, 8, 9, or 10 you always answer truthfully. You answer 'yes' by pushing button 1 or 'no' by pushing button 2.

Even if you find this technique with the dice a bit strange, it is fun to use and it is useful since it guarantees your privacy. Because nobody but you knows what you threw, nobody knows why you pushed button 1 or button 2. Therefore your true answer really remains a secret. The method is still useful for the researchers of Utrecht University because they can estimate the number of people that pushed button 1 because of the number they threw, and the number of people that pushed button 1 because they had to answer truthfully.

Now follow three exercise questions to acquaint yourself with this method
Please first throw the dice [the virtual dice appear automatically on the screen].
If you throw 2, 3 or 4 please push button 1 (= yes).
If you throw 11 or 12 please push button 2 (= no).
If you throw 5, 6, 7, 8, 9, or 10 please answer the following question truthfully:
Have you ever used public transportation without a valid ticket during the last four weeks?
Push button 1 for yes.
Push button 2 for no.

[The other exercise questions were 'have you read the paper today?' and 'have you driven through a red traffic light during the last week?']

You just answered the practice questions. Maybe you threw 2, 3 or 4 and therefore had to push button 1, while your true answer would have been 'no'. On the other hand you may have thrown 11 or 12 and had to push button 2 while your true answer would have been 'yes'. From previous research we know that people find it strange to answer incorrectly or even dishonestly. You do not need to worry about this. With this dice method you are being honest when you answer according to the rules. It is like a game, when you follow the rules of the game you are playing it honestly.

References

Becker, G. S. (1968) Crime and punishment: an economic approach. *J. Polit. Econ.*, **76**, 169–217.

Boeije, H. R. and Lensvelt-Mulders, G. (2002) Honest by chance: a qualitative interview study to clarify respondents' (non-)compliance with computer assisted randomized response. *Bull. Methodol. Sociol.*, **75**, 24–39.

Borkenau, P. and Ostendorf, F. (1992) Social desirability scales as moderator and suppresser variables. *Eur. J. Personlty*, **6**, 199–214.

Boruch, R. F. (1971) Assuring confidentiality of responses in social research: a note on strategies. *Am. Sociol.*, **6**, 308–311.

Buchman, T. A. and Tracy, J. A. (1982) Obtaining reponses to sensitive questions: conventional questionnaire versus randomized response technique. *J. Accountng Res.*, **20**, 263–271.

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

Couper, M. P. and Nichols, J. D. I. (1998) The history and development of computer assisted survey information collection methods. In *Computer Assisted Survey Information Collection* (ed. J. M. O'Reilly), pp. 1–21. New York: Wiley.

Dalton, D. R., Daily, C. M. and Wimbush, J. C. (1997) Collecting sensitive data in business ethics: a case of the unmatched count technique. *J. Bus. Eth.*, **16**, 1049–1057.

De Heer, W. (1999) International response trends: results of an international survey. *J. Off. Statist.*, **15**, 129–142.

Dillman, D. A. (1991) The design and administration of mail surveys. *A. Rev. Sociol.*, **17**, 225–249.

Elffers, H., van der Heijden, P. and Hezemans, M. (2003) Explaining regulatory non-compliance: a survey study of rule transgression for two Dutch instrumental laws, applying the randomized response method. *J. Quant. Crim.*, **19**, 409–439.

Fishbein, M. and Ajzen, I. (1975) *Belief, Attitude, Intention and Behavior*. Reading: Addison-Wesley.

Fox, J. A. and Tracy, P. E. (1986) *Randomized Response: a Method for Sensitive Surveys*. Beverly Hills: Sage.

Groves, R. M., Fowler, F. J., Couper, M. C., Lepkowski, J. M., Singer, E. and Tourangeau, R. (2004) *Survey Methodology*. New York: Wiley Interscience.

van der Heijden, P. G. M., Van Gils, G., Bouts, J. and Hox, J. J. (2000) A comparison of randomized response, computer-assisted self-interview, and face-to-face direct questioning. *Sociol. Meth. Res.*, **28**, 505–537.

Himmelfarb, S. and Lickteig, C. (1982) Social desirability and randomized response technique. *J. Personlty Socl Psychol.*, **43**, 710–717.

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

Lee, R. M. (1993) *Doing Research on Sensitive Topics*. London: Sage.

Lensvelt-Mulders, G. J. L. M., Hox, J. J., van der Heijden, P. G. M. and Maas, C. J. M. (2005) Meta-analysis of randomized response research: 35 years of validation studies. *Sociol. Meth. Res.*, **33**, 319–348.

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

Ministry of Justice (1998) Monitoring van beleidsinstrumentele wetgeving. *Report*. Ministry of Justice, The Hague.

Prinsen, H. M. and Visser, R. A. (2000) Eindrapport en advies verbetering naleving taxiregels. Department of Justice and Expertise Centre for Law Enforcement, The Hague.

Ruimschotel, D., van Reenen, P. and Klaasen, H. M. (1996) De Tafel van Elf (the Table of 11). *Beleidsanalyse*, **3**, 4–13.

Scheers, N. J. and Dayton, M. C. (1988) Covariate randomized response models. *J. Am. Statist. Ass.*, **83**, 969–974.

Sirken, M. G., Herrmann, D. J., Schechter, S., Schwarz, N., Tanur, J. M. and Tourangeau, R. (1999) *Cognition and Survey Research*. New York: Wiley.

Snijkers, G. (2001) *Cognitive Laboratory Experiences*. Heerlen: Statistics Netherlands.

Supple, A. J., Aquilino, W. S. and Wright, D. L. (1999) Collecting sensitive self-report data with laptop computers: impact on the response tendencies of adolescents in a home interview. *J. Res. Adolesc.*, **9**, 467–488.

Turnley, W. H. and Bolino, M. C. (2001) Achieving desired images while avoiding undesired images: exploring the role of self monitoring in self impression management. *J. Appl. Pychol.*, **86**, 351–360.

Umesh, U. N. and Peterson, R. A. (1991) A critical evaluation of the randomized response method: applications, validation, and research agenda. *Sociol. Meth. Res.*, **20**, 104–138.

Van Gils, G., van der Heijden, P. G. M., Laudy, O. and Ross, R. (2003) Regelovertreding in de sociale zekerheid. *Technical Report*. Interuniversitaire Onderzoeksschod Psychometrie en Sociometrie, Utrecht.

Van den Hout, A. and van der Heijden, P. G. M. (2002) Randomized response, statistical disclosure control and misclassification: a review. *Int. Statist. Rev.*, **70**, 269–288.
Van den Hout, A. and van der Heijden, P. G. M. (2004) The analysis of multivariate misclassified data with special attention to randomized response data. *Sociol. Meth. Res.*, **32**, 310–336.
Veldkamp (1997) Naleving en handhaving van de beleidsinstrumentele wetgeving. Intomart, Amsterdam.
Vella, F. (1998) Estimating models with sample selection bias: a survey. *J. Hum. Resour.*, **33**, 127–169.
Warner, S. L. (1965) Randomized response: a survey technique for eliminating evasive answer bias. *J. Am. Statist. Ass.*, **60**, 63–69.