

Repeated Cross-Sectional Randomized Response Data Taking Design Change and Self-Protective Responses into Account

L. E. Frank,¹ A. van den Hout,² and P. G. M. van der Heijden¹

¹Department of Methodology and Statistics, Utrecht University, The Netherlands

²MRC Biostatistics Unit, Institute of Public Health, Cambridge, UK

Abstract. Randomized response (RR) is an interview technique that can be used to protect the privacy of respondents if sensitive questions are posed. This paper explains how to measure change in time if a binary RR question is posed at several time points. In cross-sectional research settings, new insights often gradually emerge. In our setting, a switch to another RR procedure necessitates the development of a trend model that estimates the effect of the covariate time if the dependent variable is measured by different RR designs. We also demonstrate that it is possible to deal with self-protective responses, thus accommodating our trend model with the latest developments in RR data analysis.

Keywords: linear trend, longitudinal data, misclassification, randomized response, repeated cross-sections, self-protective responses

Randomized response (RR) is an interview technique that can be used if sensitive questions are posed and respondents are reluctant to answer directly (Chaudhuri & Mukerjee, 1988; Warner, 1965). Sensitive questions can be about fraud, drinking, or sexual behavior. A recent meta-analysis shows that RR designs lead to more valid answers than other conventional question-and-answer methods (Lensvelt-Mulders, Hox, van der Heijden, & Mass, 2005). RR designs can be defined in various ways, but they all have a specified probability mechanism that protects the privacy of individual respondents. The resulting RR variables represent misclassified responses on categorical variables if conditional misclassification probabilities are fixed by design (Chen, 1989). The true status of individual respondents is not revealed because their observed answers depend on the misclassification design as well as on the true status.

In addition to the RR setting, misclassification probabilities occur in several other fields of research. The one most closely related to RR is the postrandomization method (PRAM, Kooiman, Willenborg, & Gouweleeuw, 1997) that misclassifies values of categorical variables using a computerized process after the data are collected to protect the respondents' privacy. PRAM uses RR after the data collection. Misclassification also plays a role in medicine and epidemiology with the probabilities correctly classified as a case (sensitivity) or noncase (specificity), see Chen (1989), Copeland, Checkoway, McMichael, and Holbrook (1977), Greenland (1980, 1988), and Magder and Hughes (1997). Misclassified data can be analyzed with loglinear models

or with the general framework of latent variable models and latent class models (see e.g., Haberman, 1979; Hagenars, 1990, 1993; Rabe-Hesketh & Skrondal, 2007; Vermunt & Magidson, 2003; Vermunt, 2005; Walter, Irwig, & Glasziou, 1999).

This paper proposes a model to measure changes in time whenever RR is used to pose sensitive questions at several time points cross-sectionally. The model is illustrated with data from a Dutch repeated cross-sectional study on noncompliance to rules regarding social benefits. Data are collected every 2 years since 2000 and given that measures to prevent regulatory noncompliance are intensified during this period, the question arises as to whether the prevalence of regulatory noncompliance changes over the years and how the change can be modeled.

Considering time a covariate, we propose a method to measure the effect of this covariate if the dependent variable is measured by RR. Several aspects of the cross-sectional study at hand make it impossible to use standard analysis methods and necessitate a new approach in the analysis of RR data to deal with research questions of this type. Firstly, the fact that RR variables represent misclassified responses on categorical variables precludes the use of, for example, the *linear logit model* (Agresti, 2002, p. 180), to test for a linear trend. Using the framework of Van den Hout and van der Heijden (2004) and the results obtained by Maddala (1983) and Scheers and Dayton (1988), the trend tests proposed in this paper take the misclassification induced by the RR design into account. Secondly, as a consequence of

increasing knowledge about the efficiency of RR designs, a change in the design occurs during the cross-sectional study. In this paper, we show how to accommodate the trend model for design changes. Thirdly, to account for self-protective responses (SP) being a new development in RR data analysis (Böckenholt & van der Heijden, 2007; Cruyff, van den Hout, van der Heijden, & Böckenholt, 2007), we also present a way to incorporate SP into the trend model.

The outline of this paper is as follows. The following section explains the RR design as a misclassification design and shows how to deal with changes in the RR design over time. The section “Logit Model for Trend” introduces the trend model for RR variables with an additional procedure to account for SP. The section “Application: Prevalence of Regulatory Noncompliance Regarding Social Benefits” shows an application of the model. The last section concludes the paper with a discussion.

The RR Design

The basic idea of RR is that perturbation induced by a misclassification design protects the respondents’ privacy. Since the researcher is familiar with the statistical properties of the perturbation, a correct analysis of the observed data is feasible that takes the misclassification into account. There are several RR designs (cf. Fox & Tracy, 1986, Chap. 2). Two of these designs are used and discussed below. Each design uses a different randomizing device, that is, playing cards and dice.

For Kuk’s RR design (Kuk, 1990), the randomizing device consists of two stacks of cards, conditional on true status. The randomizing device generates binary outcomes, that is, the *yes* and *no* answers, according to two Bernoulli distributions with known parameters. One way to elicit the required binary outcomes is using two stacks of cards with varying numbers of red cards. Assume answering *yes* to the sensitive question is associated with the color red, the Kuk design can be implemented by creating a stack that contains more red than black cards, 8/10 and 2/10, respectively. The other stack, representing the *no* answer, contains more black than red cards, with 2/10 red cards. After shuffling each stack, the respondent is asked to draw a card at random from each stack. Then the sensitive question is posed. If the answer is *yes*, the respondent should name the color of the card from the right stack (with more red cards) and if the answer is *no*, the respondent should name the color of the card from the left stack (with more black cards). More details on using the Kuk design can be found in Van der Heijden, van Gils, Bouts, and Hox (2000).

The forced response design (Boruch, 1971) uses dice as randomizing device. The binary responses are generated according to the known distribution of the sum of the outcomes of two dice. After a sensitive question is posed, the respondent throws two dice and keeps the outcome hidden from the interviewer. If the outcome is 2, 3, or 4, the respondent should answer *yes*. If the outcome is 5, 6, 7, 8, 9, or 10, the respondent should answer truthfully. If the outcome is 11 or 12, the respondent should answer *no* (for details on using

the forced response method, see Lensvelt-Mulders, van der Heijden, & Laudy, 2006).

In the RR design by Kuk, violations are associated with the color *red*. As a result, the probability of a correct classification is 8/10 for the respondents who violate regulations as well as those who do not. The RR matrix that contains the conditional misclassification probabilities

$$p_{ij} = P(\text{category } i \text{ is observed} \mid \text{true category is } j) \quad (1)$$

is therefore given by

$$\mathbf{P}_{\text{Kuk}} = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix} = \begin{pmatrix} 8/10 & 2/10 \\ 2/10 & 8/10 \end{pmatrix}. \quad (2)$$

Similarly, the forced response design yields the following transition matrix:

$$\mathbf{P}_{\text{FR}} = \begin{pmatrix} 11/12 & 2/12 \\ 1/12 & 10/12 \end{pmatrix}. \quad (3)$$

As an illustration, given the forced response design with the transition matrix of Equation 3, the probability of a forced *yes* is equal to $p_{12} = 1/6$, the probability of a forced *no* is $p_{21} = 1/12$, and the probability of a truthful answer is $1 - (p_{12} + p_{21}) = 3/4$. The probability of an observed *yes* response is equal to $\pi_1^* = p_{12} + (1 - (p_{12} + p_{21}))\pi_1$, with π_1 representing the latent *yes* answer. The observed proportion of *yes* answers, denoted by $\hat{\pi}_1^*$, serves as an estimate of π_1^* . An estimate $\hat{\pi}_1$ of the latent probability of a *yes* answer π_1 can be obtained as follows:

$$\hat{\pi}_1 = \frac{\hat{\pi}_1^* - p_{12}}{1 - (p_{12} + p_{21})} \quad (4)$$

and the estimated variance of $\hat{\pi}_1$ is given by (cf. Fox & Tracy, 1986, p. 21):

$$\hat{\sigma}_{\hat{\pi}_1} = \frac{\hat{\pi}_1^*(1 - \hat{\pi}_1^*)}{N(p_{12} + p_{21})^2}. \quad (5)$$

The general form of RR designs is (Chaudhuri & Mukerjee, 1988; Van den Hout & van der Heijden, 2002)

$$\boldsymbol{\pi}^* = \mathbf{P}\boldsymbol{\pi}, \quad (6)$$

where in the event of dichotomous items, $\boldsymbol{\pi}^* = (\pi_1^*, \pi_2^*)'$ is a vector with the probabilities of the observed answers, $\boldsymbol{\pi} = (\pi_1, \pi_2)'$ is the vector of the probabilities of the latent status, and \mathbf{P} is the 2×2 matrix defined in Equation 2 or Equation 3. If \mathbf{P} is nonsingular and the observed proportion of *yes* and *no* answers are unbiased point estimates $\hat{\boldsymbol{\pi}}^*$ of $\boldsymbol{\pi}^*$, then $\boldsymbol{\pi}$ can be estimated by the unbiased moment estimator (ME) (Chaudhuri & Mukerjee, 1988; Kuha & Skinner, 1997):

$$\hat{\boldsymbol{\pi}} = \mathbf{P}^{-1}\hat{\boldsymbol{\pi}}^*. \quad (7)$$

If a binary RR question is posed on several occasions t , the vector with the probabilities of the observed answers in Equation 6 can be expressed as $\boldsymbol{\pi}^* = (\pi_1^*, \pi_2^*, \dots, \pi_k^*)'$, where $k = 2t$ because at each time point t both *yes* and *no* answers are observed. Accordingly, the matrix of conditional misclassification probabilities \mathbf{P} in Equation 6 has the dimensions $k \times k$ and estimates of \mathbf{P} can be obtained with Equation 7. An example of this extension to several time points is given in the following section.

It should be noted that in practice it is possible to obtain estimates that are outside the parameter space $[0,1]^k$ if, for example, the observed proportion of *yes* answers is very small. Van den Hout and van der Heijden (2002) demonstrate that the maximum likelihood estimator is, in general, a good alternative to the ME and, in the event of boundary solutions, the authors propose using the maximum of the likelihood for point estimation and the bootstrap percentile method to obtain confidence intervals for the point estimates.

Repeated Cross-Sectional RR Data and RR Design

Estimation Efficiency and Perceived Privacy Protection

In research settings with repeated measurements, design changes can occur as more knowledge gradually emerges about the design properties. For example, the repeated cross-sectional study on regulatory noncompliance that serves as an illustration in the application section of this paper uses two different RR designs, that is, Kuk's method in 2000 and the forced response design in the other years. The switch to the forced response design in 2002 results from greater insight into its advantages.

The probabilities of the randomization device result from a compromise between estimation efficiency and perceived respondent privacy protection. Fox and Tracy (1986, pp. 25–26) discuss this issue extensively: to provide optimal respondent protection, the probability of giving a truthful response should be as small as possible. However, the smaller the truthful response probability, the larger the variance of the estimator, thus leading to less efficiency.

Although Kuk's design has the advantage that respondents answer by naming the color of a card instead of giving a more self-incriminating *yes* or *no* answer, the forced response design has more advantages. A comparison between the forced response design and Kuk's design shows that they both yield the same estimated prevalences, but the forced response design is more efficient and comparatively easy for respondents to follow (Van der Heijden et al., 2000). Moreover, the probabilities of a forced *yes* or *no* tend to be overestimated by the respondents (Moriarty & Wiseman, 1976).

As regards the perception of the privacy protection provided by the forced response design in our cross-sectional study, the choice of the value $3/4$ for the probability of a truthful answer, as described in the section "The Randomized Response Design", does not seem to be the smallest possible probability, but it follows from the results obtained by Moriarty and Wiseman (1976) and Soeken and Macready (1982) who demonstrate that the probability of a truthful answer can be chosen between .7 and .8 without interfering with the perceived grade of anonymity. Given their results, by choosing .75, there is a probability of .25 to be divided between the forced *yes* and the forced *no* probability.

The *yes* answer represents the acknowledgment of noncompliance and because of the respondents' reluctance to admit noncompliance, the forced *yes* probability is twice as large as the forced *no* answer to make the respondent more comfortable answering *yes*. At the same time the forced *yes* probability is approximately in the same range as the expected prevalence of the sensitive topic in the population, as recommended by Clark and Desharnais (1998).

Accommodating Changes in the RR Designs

Given the switch of RR design, the misclassification probabilities can be arranged in such a way that it is possible to estimate the prevalences of RR variables collected in a repeated cross-section. In our application (see the section "Application: Prevalence of Regulatory Noncompliance Regarding Social Benefits"), RR variables are measured on three time points and the matrix of misclassification probabilities in Equation 6 can be generalized as follows. Firstly, the probabilities of the observed answers need to be restructured. The RR variable with two categories $i = 1, 2$ is observed at three time points ($t = 1, 2, 3$), leading to the probabilities for the observed answers π_{it}^* . The 2×3 table of observed answer probabilities can be represented as a vector $\pi^* = (\pi_{11}^*, \pi_{21}^*, \pi_{12}^*, \pi_{22}^*, \pi_{13}^*, \pi_{23}^*)'$, and similarly, we obtain the vector $\pi = (\pi_{11}, \pi_{21}, \pi_{12}, \pi_{22}, \pi_{13}, \pi_{23})'$. The transition matrix P in Equation 6 can be extended to a block diagonal matrix P composed of blocks P_t for each time point t . The result is the following 6×6 matrix:

$$P = \begin{pmatrix} P_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & P_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & P_3 \end{pmatrix}. \quad (8)$$

To accommodate the various RR designs used in our application, which consist of a combination of Kuk's method and the forced response method, the block diagonal matrix can be changed in the following way: P_1 is defined by the misclassification probabilities of the Kuk design in Equation 2, and P_2 and P_3 are defined by the misclassification probabilities of the forced response design in Equation 3.

Logit Model for Trend

We now return to our research question: *Has the prevalence of noncompliance changed over the years? How can the change be modeled?* This sensitive question is the dependent variable and the time points represent the independent variable with scores $t = 1, 2, 3$. If one expects a monotone trend, this hypothesis can be tested with the *linear logit model* (cf. Agresti, 2002, p. 180):

$$\pi_{1t} = \frac{\exp(\beta_0 + \beta_1 t)}{1 + \exp(\beta_0 + \beta_1 t)}, \quad (9a)$$

$$\pi_{2t} = \frac{1}{1 + \exp(\beta_0 + \beta_1 t)}, \quad (9b)$$

where π_{1t} and π_{2t} refer to the probability of a latent *yes* and *no* answer, respectively, at time point t . The *independence* model is the special case where $\beta_1 = 0$. An expansion of the model to include a quadratic trend is obtained by adding the term $\beta_2 t^2$ to the model $\beta_0 + \beta_1 t$. The log likelihood is given by

$$\ell(\boldsymbol{\beta}|n) = \sum_t n_{1t} \log \pi_{1t} + \sum_t n_{2t} \log \pi_{2t}, \quad (10)$$

which can be expressed more concisely as

$$\ell(\boldsymbol{\beta}|n) = \mathbf{u}(n \log \boldsymbol{\pi}), \quad (11)$$

where \mathbf{u} is the unit vector.

If the dependent variable is an RR variable, the log likelihood should take the resulting misclassification into account. Using the misclassified observed frequencies n_{it}^* and probabilities π_{it}^* , the adaptation of the log likelihood in Equation 10 becomes

$$\ell(\boldsymbol{\beta}|n^*) = \sum_t n_{1t}^* \log \pi_{1t}^* + \sum_t n_{2t}^* \log \pi_{2t}^*.$$

Analogously to Equation 11, the log likelihood can also be expressed in matrix algebra as

$$\ell(\boldsymbol{\beta}|n^*) = \mathbf{u}(n^* \log \boldsymbol{\pi}^*) = \mathbf{u}(n^* \log P\boldsymbol{\pi}), \quad (12)$$

where elements π_{1t} and π_{2t} of vector $\boldsymbol{\pi}$ are defined in Equations 9a and 9b, and P is a block diagonal matrix as in Equation 8. Maximizing the log likelihood in Equation 12 over parameters $\boldsymbol{\beta}$ leads to estimated probabilities for the *yes* and *no* answers on each time point, that is, $\hat{\pi}_{1t}$ and $\hat{\pi}_{2t}$.

A goodness-of-fit measure for the trend models can be obtained with the likelihood ratio statistic using the log likelihood defined in Equation 12. It makes it possible to test the hypotheses of no change (the independence model) and linear or quadratic trend (the latter is only possible, of course, if there are enough time points to leave degrees of freedom). It is well known that using the order in the time points leads to more efficient estimates of π_{it} and more powerful tests (cf. Agresti, 2002, Section 6.4, p. 236). The R-code to fit the models described in this section is available from the authors.

Accounting for SP

Although the respondents' privacy is protected by the RR design, the respondents do not always perceive it this way. Since the RR forces respondents to give a potentially self-incriminating answer about something they did not do, they might give SP, that is, respondents say *no* even if – according to the randomizing device – they should have said *yes* (see e.g., Edgell, Himmelfarb, & Duncan, 1982). In our application, the on-line questionnaires are designed in such a way that the outcome of the dice is not recorded and this is noted in the instructions given to the respondents. As a result, the respondents are free to give a different answer

than the *yes* or *no* based on the dice. Although RR performs relatively well, by eliciting more admissions of fraud than direct questioning or computer-assisted self-interviews (Lensvelt-Mulders et al., 2005), noncompliance prevalence might still be underestimated if SP is not taken into account.

Recently, several studies have focused on the detection or estimation of SP in the RR setting. Clark and Desharnais (1998) show that by splitting the sample into two groups and assigning each group a different randomization probability, it is possible to detect the presence of SP responses and to measure their extent. Böckenholt and van der Heijden (2007) use a multivariate approach to estimate SP by proposing an item randomized-response (IRR) model, with a common sensitivity scale assumed for a set of RR variables. Response behavior that does not follow the RR design is approached by introducing mixture components in the IRR models with the first component consisting of respondents who answer truthfully and follow an item response model, and the second component consisting of respondents who systematically say *no* to every item in a subset of items. A similar approach is adopted by Cruyff et al. (2007) who work out the same idea in the context of loglinear models.

Since we feel this new development to correct RR estimates for SP responses is important, we propose the following procedure to incorporate SP into the trend model. In the first step, we estimate the amount of SP at each wave using the Profile Likelihood method proposed by Cruyff et al. (2007). In the second, change in time is modeled using the frequencies adjusted for SP.

The combination of SP estimation with an opportunity to account for changes in RR design, as described in the section “Repeated Cross-Sectional RR Data and RR Design”, presents a new approach in the setting of cross-sectional RR data and has the following advantages. A change in the RR design might lead to differences in the precision of the estimated noncompliance prevalences, which are associated with the amount of trust respondents have in a particular RR design. Less trust leads to more SP. Correcting for SP could lead to better estimates of noncompliance at each time point, and thus to a more valid estimate of the trend. In addition, the estimation of SP at each time point makes it possible to adjust the misclassification probabilities of the RR design at the next time point, leading to a better balance between estimation efficiency and privacy protection.

A drawback of the two-step approach we propose is that the uncertainty about SP estimates in the first step is not automatically taken into account in the theoretical standard errors of the trend model in the second step. Empirical standard errors are thus obtained for the regression coefficients of the trend model using the non-parametric bootstrap (Efron & Tibshirani, 1998). The details of the bootstrap procedure are as follows:

- At each time point, sample B times n respondents with replacement, with n equal to the sample size at each time point.
- At each time point, estimate SP for each bootstrap sample. Adjust the bootstrap sample frequencies for SP at each time point (first step of the two-step approach).

- Fit the independence model and the linear trend model to each of the B bootstrap samples. This results in B estimates of the intercept in the independence model (or intercept only model) and B estimates of the intercept and slope in the linear trend model. The standard deviations of the distributions of these B intercept and slope estimates yield the bootstrap estimates of standard errors and the 95% bootstrap percentile intervals (second step of the two-step approach).

Application: Prevalence of Regulatory Noncompliance Regarding Social Benefits

The Data

It is mandatory for Dutch employees to be insured under the Disability Insurance Act and, provided certain conditions are met, a formerly employed person is entitled to financial benefits amounting to as much as 70% of his last income. The welfare system is rather costly and the Dutch Department of Social Services regularly monitors the prevalence of noncompliance to the rules. After a pilot in 1998, three waves followed in 2000, 2002, and 2004. A detailed description of the 2002 cross-sectional study is given in Lensvelt-Mulders et al. (2006). The Department of Social Affairs intensifies the measures to prevent regulatory non-compliance in these years and is interested in knowing

Table 1. Cross-sectional data consisting of observed weighted frequencies of *yes* (n_1^*) and *no* (n_2^*) answers measured at three time points, and estimated prevalences of noncompliance ($\hat{\pi}_1$) with 95% confidence intervals

	2000 ($n = 1308$)	2002 ($n = 1760$)	2004 ($n = 830$)
n_1^*	388	466	197
n_2^*	920	1294	633
$\hat{\pi}_1$	0.16	0.10	0.07
95% CI	[0.13, 0.19]	[0.07, 0.13]	[0.03, 0.11]

whether the prevalence of regulatory noncompliance changes over the years and how the change can be modeled.

The application focuses on the following sensitive question about the health of the respondent: *For periods of any length at all, do you ever feel stronger and healthier and able to work more hours without informing the Department of Social Services of this change?* If noncompliance is detected, it can lead to sanctions and sometimes even to loss of invalidity insurance benefits. Given the sensitivity of the topic, asking respondents directly whether they violate the rules will not yield valid results (cf. Van der Heijden et al., 2000). A RR design is thus used at each wave to ensure the confidentiality of the answers. For the question just described, Table 1 displays the observed frequencies of *yes* (n_1^*) and *no* (n_2^*) answers at the three time points as well as estimated probabilities of regulatory noncompliance corrected for the RR design (as explained in the section “The Randomized Response Design”). Person weights are used to weight the sample toward population characteristics (cf. Lensvelt-Mulders et al., 2006). A change of the RR design occurred at time point 2002, where Kuk’s design is replaced by the forced response design. Accordingly, the block diagonal matrix of misclassification probabilities in Equation 8 is used to accommodate this RR design change.

Results

Two models, the independence model and the linear trend model, are fitted to the RR data. The goodness-of-fit of the models is evaluated with the likelihood ratio statistic. Figure 1 shows that the estimated regulatory noncompliance prevalences decrease monotonically over the time points, and the estimated logistic regression parameter for the linear trend thus has a negative value (see Table 2). The value of the likelihood ratio statistic L^2 in Table 2 indicates that the independence model does not fit ($L^2 = 10.415$, $df = 2$, $p = .001$), whereas the linear model produces a good fit: $L^2 = 0.004$, $df = 1$, $p = .95$. Testing the linear model against the independence model (see the ΔL^2 -values in the lower part of Table 2) leads to the conclusion that the linear trend model is a significant improvement. In addition, both parameters of the linear trend model depart significantly from zero. This means that of the people entitled to social benefits, the proportion of respondents who do not obey the rule of informing the Department of Social Services about any improvement in their health significantly decreased during the period 2000–2004.

Table 2. Results trend analyses (the model of choice is in bold typeface)

Model	L^2	df	p	$\hat{\beta}_0 \left(\hat{\sigma}_{\hat{\beta}_0} \right)$	$\hat{\beta}_1 \left(\hat{\sigma}_{\hat{\beta}_1} \right)$
[1] Independence	10.415	2	.001	-2.09 (0.10)	
[2] Linear	0.004	1	.95	-1.17 (0.28)	-0.49 (0.16)
	ΔL^2	Δdf	p		
[1]–[2]	10.411	1	.001		

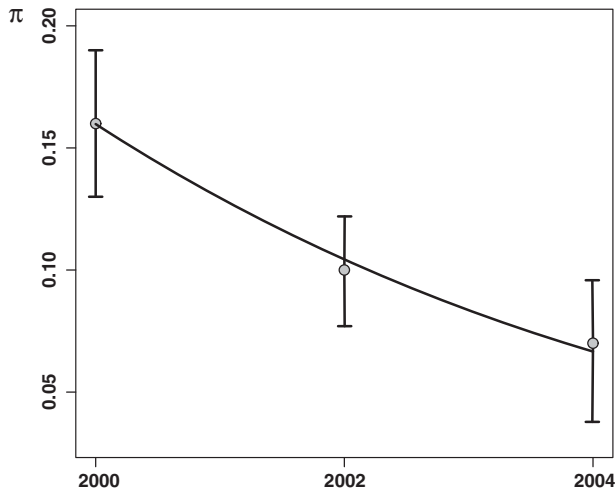


Figure 1. Univariately estimated regulatory noncompliance prevalences (represented by the dots) with 95% confidence intervals, and fitted trend line on the RR question *For periods of any length at all, do you ever feel stronger and healthier and able to work more hours without informing the Department of Social Services of this change?*

Accounting for SP

We now present the results of the trend analysis that takes SP into account using the two-step approach. It should be noted that simply because multivariate RR data are needed to estimate the probability of SP, it is not possible to simultaneously model change in time and SP behavior, whereas we have repeated univariate data here (note that at each time point a distinct sample is used, see Table 1). At the first step, we estimate the amount of SP at each wave using multivariate data consisting of three additional RR questions about health, which are part of the full data set. Given the estimates of SP for each wave, a correction for SP is carried out by adjusting the sampling weights of the data in accordance with the estimated amount of SP. In the second step, we use the estimates of SP as external information in our trend analyses. Using this approach, the SP proportion is estimated regarding the data in Table 1 with the Profile Likelihood method proposed by Cruyff et al. (2007). The resulting proportion of SP-answers in the three waves are .13, .15, and .11, in 2000, 2002, and 2004, respectively.

Adjusting the observed frequencies for the SP proportions yields the following estimates of regulatory noncompliance prevalences $\hat{\pi}$: .24, .17, and .11 for 2000, 2002, and 2004, respectively. Comparing these SP-corrected noncompliance prevalences with the uncorrected probabilities in Table 1 clearly shows that the SP-corrected probabilities are higher and that not accounting for SP leads to an underestimation of noncompliance prevalences. Since the estimated proportions of SP-answers do not change considerably over the years and the findings by Van der Heijden et al. (2000) show that Kuk's design and the forced respond method yield the same estimates of noncompliance preva-

Table 3. Results trend analyses on univariately estimated regulatory noncompliance prevalences corrected for SP with likelihood ratio statistics (L^2), bootstrap estimates of standard errors (\hat{s}_B), and bootstrap percentile confidence intervals

Model	$\hat{\beta}$	\hat{s}_B	95% CI
[1] Independence ($L^2 = 17.62$, $df = 2$, $p = .00$)			
β_0	-1.61	0.10	[-1.75, -1.41]
[2] Linear ($L^2 = 1.27$, $df = 1$, $p = .26$)			
β_0	-0.67	0.30	[-1.07, -0.09]
β_1	-0.50	0.16	[-0.81, -0.26]

lences, no adjustments are made to the misclassification probabilities of the RR design in the cross-sectional study.

In the second step, change in time is modeled by fitting the trend model to the observed frequencies adjusted for SP and leads to the results shown in Table 3. The likelihood ratio statistic L^2 is equal to 17.62 ($df = 2$, $p = .00$) for the independence model and for the linear trend model $L^2 = 1.27$ ($df = 1$, $p = .26$). The model fit clearly improves after adding a parameter to account for linear trend. Following the bootstrap set-up explained in the section "Accounting for Self-protective Responses", the variability of the logistic regression parameters is estimated by bootstrap standard deviations. The results are based on the observed frequencies of *yes* or *no* responses in Table 1 and three additional RR questions about health in each of the three waves. For each of the years 2000, 2002 and 2004, 1,000 times n respondents are sampled with replacement, with n equal to the sample size 1308, 1760, and 830 for the years 2000, 2002, and 2004, respectively.

The 95% bootstrap confidence intervals (percentile method) show that both of the parameters of the linear model depart significantly from zero, leading to the conclusion that the SP-corrected noncompliance prevalences decrease linearly on the logit scale over the three time points. This means the proportion of people who do not obey the rule about informing the authorities if their health improves decreases in the period from 2000 to 2004.

Discussion

This paper shows how to measure the effect of the covariate time on repeated cross-sectional RR data taking RR design changes and SP into account. Due to the misclassification design, traditional trend models cannot be used. A key element of the proposed method is the construction of a block diagonal matrix with conditional classification probabilities for each time point, making it possible to use different RR designs over time. The block diagonal matrix can easily be extended to include one or more additional categorical covariates. For example, one might wish to know whether noncompliance behavior differs between men and women or whether the difference remains constant over time.

Although the RR method protects the respondents' privacy, it does not entirely exclude an evasive response bias. As a result, noncompliance prevalences might still be underestimated if SP is not taken into account. We show that it is possible to correct RR estimates for SP in the trend model using a two-step procedure with the amount of SP estimated in the first step and the trend model fitted on frequencies corrected for the SP estimates in the second step.

The two-step procedure has the disadvantage of being computationally demanding, but it seems inevitable in the setting of repeated univariate data, with a distinct sample used at each time point. In this situation, it is not possible to simultaneously model change in time and SP behavior, as is shown in the recently developed methods to correct RR estimates for SP responses. It should be noted that the two-step procedure uses the same data twice, that is, to estimate the SP probabilities and fit the model with associated standard errors for the model parameters. A possible solution would be to use cross-validation, estimating the SP and fitting the model on the training set and obtaining the variability estimates for the model parameters in the test set, although this would lead to a very complex simulation set-up. Estimating SP is a recent development and requires further research.

It is the combination of taking design changes and SP into account that constitutes a new approach to cross-sectional RR data. The possible effects of RR design changes on the trend in the estimated prevalences can be accommodated by taking the presence of SP into account. In addition, it is now possible to adjust the misclassification probabilities of the RR design according to the SP estimates at the previous time point, thus providing a better balance between estimation efficiency and privacy protection.

References

- Agresti, A. (2002). *Categorical data analysis*. New Jersey: John Wiley & Sons.
- Böckenholt, U., & van der Heijden, P. G. M. (2007). Item randomized-response models for measuring noncompliance: Risk-return perceptions, social influences, and self-protective responses. *Psychometrika*, 72, 245–262.
- Boruch, R. F. (1971). Assuring confidentiality of responses in social research: a note on strategies. *The American Sociologist*, 6, 308–311.
- Chaudhuri, A., & Mukerjee, R. (1988). *Randomized response: Theory and techniques*. New York: Marcel Dekker.
- Chen, T. T. (1989). A review of methods for misclassified categorical data in epidemiology. *Statistics in Medicine*, 8, 1095–1106.
- Clark, S. J., & Desharnais, R. A. (1998). Honest answers to embarrassing questions: Detecting cheating in the randomized response model. *Psychological Methods*, 3, 160–168.
- Copeland, K. T., Checkoway, H., McMichael, A. J., & Holbrook, R. H. (1977). Bias due to misclassification in the estimation of relative risk. *American Journal of Epidemiology*, 105, 488–495.
- Cruyff, M. J. L. F., van den Hout, A., van der Heijden, P. G. M., & Böckenholt, U. (2007). Log-linear randomized-response models taking self-protective response behavior into account. *Sociological Methods & Research*, 36, 266–282.
- Edgell, S. E., Himmelfarb, S., & Duncan, K. L. (1982). Validity of forced response in a randomized response model. *Sociological Methods and Research*, 11, 89–110.
- Efron, B., & Tibshirani, R. (1998). *An introduction to the bootstrap*. London: Chapman and Hall.
- Fox, J. A., & Tracy, P. E. (1986). *Randomized response. A method for sensitive surveys* (No. 58). Newbury Park: Sage Publications.
- Greenland, S. (1980). The effect of misclassification in the presence of covariates. *American Journal of Epidemiology*, 112, 564–569.
- Greenland, S. (1988). Variance estimation for epidemiologic effect estimates under misclassification. *Statistics in Medicine*, 7, 745–757.
- Haberman, S. J. (1979). *Analysis of qualitative data: New developments* (Vol. 2). New York: Academic Press.
- Hagenaars, J. A. (1990). *Categorical longitudinal data. Loglinear analysis of panel, trend and cohort data*. Newbury Park: Sage.
- Hagenaars, J. A. (1993). *Loglinear models with latent variables*. Newbury Park: Sage.
- Koiman, P., Willenborg, L. C. R. J., & Gouweleeuw, J. M. (1997). *PRAM: A method for disclosure limitation of microdata*. Voorburg/Heerlen: Statistics Netherlands (Research paper No. 9705).
- Kuha, J., & Skinner, C. (1997). Categorical data analysis and misclassification. In L. Lyberg (Ed.), *Survey measurement and process quality*. New York: Wiley.
- Kuk, A. Y. C. (1990). Asking sensitive questions indirectly. *Biometrika*, 77, 436–438.
- Lensvelt-Mulders, G. J. L. M., Hox, J. J., van der Heijden, P. G. M., & Maas, C. J. M. (2005). Meta-analysis of randomized response research: Thirty-five years of validation. *Sociological Methods and Research*, 33, 319–348.
- Lensvelt-Mulders, G. J. L. M., van der Heijden, P. G. M., & Laudy, O. (2006). A validation of a computer-assisted randomized response survey to estimate the prevalence of fraud in social security. *Journal of the Royal Statistical Society A*, 169, 305–318.
- Maddala, G. S. (1983). *Limited dependent and qualitative variables in econometrics*. Cambridge: Cambridge University Press.
- Magder, L. S., & Hughes, J. P. (1997). Logistic regression when the outcome is measured with uncertainty. *American Journal of Epidemiology*, 146, 195–203.
- Moriarty, M., & Wiseman, F. (1976). On the choice of a randomization technique with the randomized response model. In *Proceedings of the social statistics section of the American statistical association* (pp. 624–626). Washington, DC: American Statistical Association.
- Rabe-Hesketh, S., & Skrondal, A. (2007). Multilevel and latent variable modeling with composite links and exploded likelihoods. *Psychometrika*, 72(2), 123–140.
- Scheers, N. J., & Dayton, C. M. (1988). Covariate randomized response models. *Journal of the American Statistical Association*, 83, 969–974.
- Soeken, K. L., & Macready, G. B. (1982). Respondents' perceived protection when using randomized response. *Psychological Bulletin*, 92, 487–489.
- Van den Hout, A., & van der Heijden, P. G. M. (2002). Randomized response, statistical disclosure control and misclassification: A review. *International Statistical Review*, 70, 269–288.
- Van den Hout, A., & van der Heijden, P. G. M. (2004). The analysis of multivariate misclassified data with special attention to randomized response. *Sociological Methods and Research*, 32, 384–410.

- Van der Heijden, P. G. M., van Gils, G., Bouts, J., & Hox, J. J. (2000). A comparison of randomized response, computer-assisted self-interview, and face-to-face direct questioning. Eliciting sensitive information in the context of welfare and unemployment benefit. *Sociological Methods and Research*, 28, 505–537.
- Vermunt, J. K. (2005). Mixed-effects logistic regression models for indirectly observed outcome variables. *Multivariate Behavioral Research*, 40, 281–301.
- Vermunt, J. K., & Magidson, J. (2003). Latent class models for classification. *Computational Statistics and Data Analysis*, 41, 531–537.
- Walter, S. D., Irwig, L., & Glasziou, P. P. (1999). Meta-analysis of diagnostic tests with imperfect reference standards. *Journal of Clinical Epidemiology*, 10, 943–951.
- Warner, S. L. (1965). Randomized response: A survey technique for eliminating answer bias. *Journal of the American Statistical Association*, 60, 63–69.

Laurence E. Frank

Department of Methodology and Statistics
Universiteit Utrecht
Heidelberglaan 1
P.O. Box 80140
3508 TC Utrecht
The Netherlands
Tel. +31 30 253 4438
Fax +31 30 253 5797
E-mail L.E.Frank@uu.nl
