Measuring noncompliance in insurance benefit regulations with randomized response methods for multiple items

Ulf Böckenholt[1] and Peter G.M. van der Heijden[2]

[1] Faculty of Management, McGill University, 1001 Sherbrooke Street West, Montreal, QC H3A 1G5, CANADA
[2] Department of Methodology and Statistics, Utrecht University, P.O. Box 80.140, 3508 TC Utrecht, The Netherlands

Abstract: Randomized response (RR) is a well known method for measuring sensitive behavior. Yet it is not often applied. Two possible reasons for this are (i) its lower efficiency and the resulting need for larger sample sizes, making applications of RR expensive, (ii) the notion that in many applications the RR design may not be followed by every respondent ('cheating').
This paper addresses the efficiency problem by proposing item response theory (IRT) models for the analysis of multivariate RR data. In these models a person parameter is estimated based on multiple measures of a sensitive behavior under study which yields a more efficient and powerful analysis of individual differences than available from univariate RR data. Cheating in a RR study is approached by introducing additional mixture components in the IRT models with one component consisting of respondents who answer truthfully and other components consisting of respondents who do not provide truthful responses to all or a subset of the items.
The resulting IRT model is applied to data from a Dutch survey conducted under receivers of disablement insurance benefit (DIB) who are interviewed about their compliance behavior to rules that are a prerequisite for receiving DIB.

Keywords: randomized response; item response theory; cheating; sensitive behavior; efficiency.

1 Introduction

In many RR studies, respondents are asked multiple questions about one or more domains. For example, in the 2002 surveys conducted in the Netherlands on social security regulation infringements of the Occupational Disability Insurance Act, the Unemployment Insurance Act and the National Social Security Assistance Act, each social security recipient was asked about nine randomized–response questions about their compliance with these regulations. The following four questions focussed on health–related issues: (1) Have you been told by your physician about a reduction in your

disability symptoms without reporting this improvement to your social welfare agency? (2) On you last spot–check by the social welfare agency, did you pretend to be in poorer health than you actually were? (3) Have you noticed personally any recovery from you disability complaints without reporting it to the social welfare agency? (4) Have you felt for some time now to be substantially stronger and healthier and able to work more hours, without reporting any improvement to the social welfare agency?
Clearly, these questions are ordered according to their degree of intentional violations of the regulations. A person who does not report the outcome of a medical investigation may also avoid reporting any personally noticed improvements of their health status. In contrast, persons who notice personal improvements may or may not mis-report their health status. Item–response models (van der Linden et al., 1997) are well–suited for studying how individuals differ in their compliance behavior by ordering respondents on a latent continuum that represents their level of compliance.
Although there is much empirical support to indicate that RR methods increase the number of honest responses, there is no guarantee that all respondents provide truthful answers (see van der Heijden et al, 2000). Some respondents might violate the rules set out by the RR procedure. Here the Forced Choice response format is used: respondents are asked to throw two dice, to answer "yes" when the outcome is 2, 3 or 4, to answer "no" when the outcome is 11 or 12, and to answer truthfully when the outcome is between 5 and 10. A typical rule violation is to answer "no" whatever the outcome of the dice (compare van den Hout et al., 2004; Clark et al., 1998).
To accommodate such response behavior, an extension of the item response approach is presented which allows explicitly for a response bias in the sense that it can capture a possible tendency of respondents towards giving a "No" response regardless of the outcome of the randomizing device. These respondents are captured by a latent class that can be identified by an extreme use of "No" responses.

2 RR Models for Multiple Items

We distinguish three classes of RR models for multiple items. The first class assumes that respondents are homogenous in their compliance behavior and have a fixed probability of answering each item. This is the classical RR model and it is used as a benchmark for the models proposed next. The second model class relaxes the homogeneity assumption and allows for individual variability in compliance for the various behaviors under study. The third class of models considers the possibility that a subset of respondents may not follow the randomization instructions and answer "No" regardless of the outcome of the randomization device.
When all respondents have the same probability of endorsing an item, it

is convenient to express the probability of answering affirmatively by the logistic function with

$$\Pr(x_{ij} = 1) = \Pr(\gamma_j) = \frac{1}{1 + \exp(\gamma_j)}$$

Under random sampling of the respondents, the likelihood function of the homogeneous–response model can then be written as

$$L = \prod_{i=1}^{n} \prod_{j=1}^{J} [\frac{1}{6} + .75 \Pr(\gamma_j)]^{x_{ij}} \; [1 - (\frac{1}{6} + .75 \Pr(\gamma_j))]^{1-x_{ij}}. \tag{1}$$

where $\frac{1}{6}$ is the probability of a forced "yes" and .75 is the probability of a truthful answer. Clearly, the assumption that all respondents have an equal probability of answering an item is too strong in most applications although it is the standard assumption for single-item RR studies.

The second class of models assumes that associations among the responses to multiple items are caused by a person–specific compliance parameter. Because typically the number of items is small in a RR study, we adopt the Rasch (1980) model to measure individual differences in compliance behavior. Under this model, the probability that item j is answered affirmatively by person i can be written as

$$\Pr(x_{ij} = 1) = \Pr(\gamma_j, \theta_i) = \frac{1}{1 + \exp(\theta_i - \gamma_j)},$$

where γ_j is called the item location parameter. Typically, the person parameter θ_i is specified to vary according to a normal distribution.

Under the Forced Choice response format, the item-response model needs to be modified to account for the randomization effect. In this case the likelihood function can be written as:

$$\begin{aligned} L &= \prod_{i=1}^{n} \int \prod_{j=1}^{J} [\frac{1}{6} + .75 \Pr(\gamma_j, \theta_i)]^{x_{ij}} \times \\ & \quad [1 - (\frac{1}{6} + .75 \Pr(\gamma_j, \theta_i))]^{1-x_{ij}} f(\theta; \mu, \sigma) d\theta, \end{aligned} \tag{2}$$

where $f(\theta; \mu, \sigma)$ is the normal density with parameters μ and σ. Note that the mean μ of the population distribution cannot be estimated independently of the item locations. In the reported application, we therefore set $\mu = 0$. It is worthwhile stressing that the normal distribution assumption may not always be appropriate in RR studies and that other distributional forms should be considered to capture more closely the non–compliance variability in the population of interest.

The third class of models allows for the possibility that not all respondents comply with the randomization response format and provide a "No" response regardless of the question asked. Combined with the item–response model given by (2), the likelihood function is specified as:

$$
\begin{aligned}
L &= \prod_{i=1}^{n}(\pi \int \prod_{j=1}^{J}\{[\frac{1}{6}+.75\,\Pr(\gamma_j,\theta_i)]^{x_{ij}}\ [1-(\frac{1}{6}+.75\,\Pr(\gamma_j,\theta_i))]^{1-x_{ij}}\} \\
&\times\ f(\theta;\mu,\sigma)d\theta + (1-\pi)\prod_{j=1}^{J}\{\Pr(\text{“No”})^{x_{ij}}\ [1-\Pr(\text{“No”})]^{1-x_{ij}}\}), \quad (3)
\end{aligned}
$$

where π denotes the probability of a randomly sampled person to answer the questions according to the FC mechanism. In the reported application, we specify that participants who answer “No” regardless of the question asked, give this response with probability 1. It is straightforward to relax this assumption and to estimate the probability of a “No”− response from the data. The crucial assumption of (3) is that members of the “No”-group do not provide any information about the items' location and discrimination parameters.

3 Data Analysis

The aim of the study and the RR design have been described above. We note that 44% of all respondents respondents provide “No” responses to all four items.

The homogeneous model required the estimation of four item location parameters and yielded a goodness-of-fit statistic of $G^2 = 123.8$ with 11 d.f.. Clearly, the assumption of no individual differences does not agree with the data. This result is supported by the fit improvement obtained from Model (2). With one additional parameter, the variance of the normal distribution σ^2, Model (2) provides a major fit improvement ($G^2 = 23.4$ with 10 d.f.). However, despite the better fit, this model does not describe the data satisfactorily. The main reason for the misfit is that the outcome of consistent “No”–responses to the four items is greatly underestimated by (2). Model (3) can address this problem by allowing for the possibility that some respondents select the “No”–response for reasons that are unrelated to the compliance parameter θ. The resulting fit improvement provides support for this specification ($G^2 = 14.3$ with 9 d.f.).

Table 1 contains the corresponding parameter estimates of the three models. We note that the standard errors of Model (1) are too small since this model does not reflect the dependencies among the four responses. In contrast, Model (2) overestimates strongly the degree of heterogeneity in the data since it tries to fit the large percentage of “No”–responses to the four items. Model (3) yields a much reduced but still substantial estimate of the population standard deviation ($\hat{\sigma} = 2.07$). About 196 or 12% of the respondents are classified as consistent “No”–sayers. For the remaining 88% of the respondents, the items are ordered but far away from the mean of the population distribution. Clearly, a positive response to any of the four items is low at the mean of the population distribution.

TABLE 1. Parameter Estimates (and Standard Errors) of RR–Models for Multiple Items

Parameter	Model (1)	Model (2)	Model (3)
$\hat{\gamma}_1$	3.77 (.56)	9.10 (2.74)	4.56 (.88)
$\hat{\gamma}_2$	3.07 (.30)	8.44 (2.73)	3.99 (.80)
$\hat{\gamma}_3$	2.58 (.20)	7.61 (2.72)	3.42 (.67)
$\hat{\gamma}_4$	1.94 (.13)	5.83 (2.01)	2.63 (.53)
$\hat{\sigma}$	–	4.72 (1.61)	2.15 (.47)
$\ln(\frac{\hat{\pi}}{1-\hat{\pi}})$	–	–	2.07 (.34)

By taking into account that about 12% of the respondents give a "No"-response without providing information about their actual compliance behavior, Model (3) renders more accurate estimates about the compliance rate in the population. Under Model (1) the percentage of non-compliant respondents for the four items are 2.2%, 4.5%, 7.0%, and 12.5% respectively. In contrast, under Model (3) the corresponding estimates are 5.2%, 7.7%, 11.0%, and 17.0%. These differences are substantial and demonstrate the value of the proposed models for the analysis of RR data.

References

Clark, S.J. , and Desharnais, R.A. (1998). Honest answers to embarrassing questions: detecting cheating in the randomized response model. *Psychological Methods*, **3**, 160-168.

Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests*. Chicago: The University of Chicago Press. (Original published 1960, Copenhagen: The Danish Institute of Educational Research)

van der Linden, W. J. and Hambleton, R. K. (Eds.) (1997). *Handbook of modern item response theory.* New York: Springer.

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. *Sociological Methods and Research*, **32**, 310-336.

van der Heijden, P.G.M., van Gils, , G., Bouts, J. and Hox, J. (2000). A comparison of randomized response, CASAQ, and direct questioning; eliciting sensitive information in the context of welfare and unemployment benefit. *Sociological Methods and Research*, **28**, 505-537.