Journal of Statistical Software

June 2018, Volume 85, Issue 2. doi: 10.18637/jss.v085.i02

RRreg: An R Package for Correlation and Regression Analyses of Randomized Response Data

Daniel W. Heck
University of Mannheim

Morten Moshagen
Ulm University

Abstract

The randomized response (RR) technique was developed to improve the validity of measures assessing attitudes, behaviors, and attributes threatened by social desirability bias. The RR removes any direct link between individual responses and the sensitive attribute to maximize the anonymity of respondents and, in turn, to elicit more honest responding. Since multivariate analyses are no longer feasible using standard methods, we present the R package **RRreg** that allows for multivariate analyses of RR data in a user-friendly way. We show how to compute bivariate correlations, how to predict an RR variable in an adapted logistic regression framework (with or without random effects), and how to use RR predictors in a modified linear regression. In addition, the package allows for power analysis and robustness simulations. To facilitate the application of these methods, we illustrate the benefits of multivariate methods for RR variables using an empirical example.

Keywords: randomized response, indirect questioning, survey design, social desirability, sensitive questions, compliance, power analysis.

1. Introduction

A common goal in the social and behavioral sciences is to investigate behaviors, attitudes, and attributes that are associated with certain individual, legal, or social norms. It is well known, however, that responses to direct questions merely reflect what participants tell investigators rather than their true status on the attribute of interest. As a consequence, prevalence estimates of sensitive, incriminating, or illegal attributes are systematically biased towards respondents' perception of what is socially acceptable (e.g., Tourangeau and Yan 2007), in turn obscuring associations between sensitive attributes and other variables.

The randomized response (RR) technique (Warner 1965) was developed as a means to reduce

this response bias by providing a higher degree of anonymity and confidentiality than traditional direct questioning (DQ) formats – based on the idea that respondents should be more willing to respond truthfully in a truly anonymous situation. RR is best understood as a family of techniques that share the common rationale to add random noise to the individual answers of participants such that an individual's true status is not directly identifiable from her observed response. For example, in the historically first RR model proposed by Warner (1965), respondents either receive the sensitive statement (e.g., "I have used cocaine") or its negation ("I have never used cocaine"), depending on the outcome of a randomization process with known probability distribution (such as the roll of fair dice). The randomization outcome is only known to the respondent, so an individual's *yes* or *no* response does not disclose whether she actually has engaged in the sensitive behavior. Although the randomization thereby ensures the anonymity of individual respondents, elementary probability calculations allow for group based prevalence estimates of the attribute under consideration.

Due to this increased anonymity and confidentiality, more honest responding is encouraged. Indeed, since the seminal work of Warner (1965), a variety of variants of the RR technique (RRT) have been proposed and successfully employed to obtain information about attitudes and behaviors as diverse as abortion (Lara, Strickler, Olavarrieta, and Ellertson 2004), academic fraud (Jann, Jerke, and Krumpal 2012; Mortaz Hejri, Zendehdel, Asghari, Fotouhi, and Rashidian 2013), business ethics (Robertson and Rymon 2001), cheating (Moshagen, Hilbig, Erdfelder, and Moritz 2014), cooperative behavior (Moshagen, Hilbig, and Musch 2011), doping in fitness sports (Simon, Striegel, Aust, Dietz, and Ulrich 2006), employee theft (Wimbush and Dalton 1997), hygiene practices (Moshagen, Musch, Ostapczuk, and Zhao 2010), job applicant faking (Donovan, Dwight, and Hurtz 2003), medication non-adherence (Ostapczuk and Musch 2011), poaching (Blank and Gavin 2009; Solomon, Jacobson, Wald, and Gavin 2007), rape (Soeken and Damrosch 1986), smuggle (Nordlund, Holme, and Tamsfoss 1994), social security fraud (Lensvelt-Mulders, Van der Heijden, Laudy, and Van Gils 2006), software piracy (Kwan, So, and Tam 2010), substance abuse (Dietz, Striegel, Franke, Lieb, Simon, and Ulrich 2013; Franke *et al.* 2013), use of poison by farmers (Santangeli, Arkumarev, Rust, and Girardello 2016), voting (Moshagen, Musch, and Erdfelder 2012), and xenophobia (Krumpal 2012; Ostapczuk, Musch, and Moshagen 2009b). Likewise, a meta-analysis confirmed that more valid prevalence estimates of sensitive attributes can be obtained by means of RR compared to traditional data collection techniques (Lensvelt-Mulders, Hox, Van der Heijden, and Maas 2005). Lensvelt-Mulders *et al.* (2005) concluded that "research has not demonstrated the superiority of any data collection method to RRT" (p. 343). Despite this generally positive evaluation, note that some studies provided mixed evidence concerning the validity of the RR technique (Locander, Sudman, and Bradburn 1976; Lamb and Stem 1978; Tracy and Fox 1981; Umesh and Peterson 1991) or pointed to possible limitations and biases of the RR method (Coutts and Jann 2011; Coutts, Jann, Krumpal, and Näher 2011; Kirchner, Krumpal, Trappmann, and von Hermanni 2013; Wolter and Preisendorfer 2013; Höglinger, Jann, and Diekmann 2016; Höglinger and Jann 2016; Hoffmann, De Puiseau, Schmidt, and Musch 2017), indicating that the validity of the RR depends on several moderating variables, including the chosen RR design, randomization device, and the degree of trust and understanding elicited (Landsheer, Van der Heijden, and Van Gils 1999; Hoffmann *et al.* 2017). Nevertheless, the extensive literature on the RR technique indicates that the RR, when properly implemented, can be a very valuable (though not perfect) measurement approach for research on sensitive topics.

Importantly, typical research questions in the social and behavioral sciences go beyond determining the prevalence of a sensitive attribute, but rather refer to associations between several variables. In other words, the prime interest lies in correlations and predictions involving one or more covariates. Applying traditional correlation or regression analyses to data obtained via RR is inappropriate, however, as the randomization component of the RR distorts the response variable. A simple approach to circumvent this problem is to perform a median-split on a continuous covariate and to compare the estimates across the resulting groups (e.g., Latkin and Vlahov 1998; Ostapczuk and Musch 2011; Rosenfeld *et al.* 1991; Soeken and Macready 1986; Hilbig, Moshagen, and Zettler 2015). However, median-splits suffer from several shortcomings, including loss of information, decreased power, and difficulties when considering more than one variable at a time (Royston, Altman, and Sauerbrei 2006). Thus, multivariate analyses of RR data using appropriately adapted correlation and regression procedures are clearly preferable.

Fortunately, correlation and regression models appropriate for the analysis of RR variables have become available. However, for several years only two Stata (StataCorp 2015) routines provided basic implementations for the prediction of binary RR variables for the original Warner model and two other RR designs (i.e., **RRLOGIT** and **RRREG**; Jann 2008, 2005). More recently, logistic RR regression for four RR designs has become available in R (R Core Team 2018) as implemented in the package **rr** (Blair, Zhou, and Imai 2016), which also includes univariate power analyses. In addition, more general-purpose software for the analysis of misclassified data could potentially be adopted for the RR technique (e.g., using the SIMEX method; Lederer and Küchenhoff 2006; Carroll, Küchenhoff, Lombard, and Stefanski 1996). To further facilitate the multivariate analysis of RR data, herein we introduce the R package **RRreg** (Heck and Moshagen 2018). Package **RRreg** improves upon existing implementations by providing a unifying modeling and computational framework and thereby facilitating the application and comparison of a wide range of RR designs for several types of multivariate analysis. In particular, going beyond previous packages, package **RRreg** implements both logistic and linear RR regression models for a large class of RR designs (including those addressing instruction non-adherence as well as RR designs for continuous data; see Table 2), includes logistic regression with random effects (which is particularly important for large-scale surveys with clustered data), allows to compute RR correlations (including correlations between RR variables), and provides easy-to-use Monte Carlo simulation methods to estimate statistical power and to investigate robustness in both univariate and multivariate settings (allowing, for example, to select the most efficient RR design, randomization probabilities, and statistical approach before administering a survey).

In the following, we first describe several variants of the RR technique and their specification in package **RRreg**. Next, we introduce the implemented multivariate methods for RR variables. Finally, we illustrate the benefits and the implementation of these approaches in package **RRreg** via an empirical example. The package is available from the Comprehensive R Archive Network (CRAN) at `https://CRAN.R-project.org/package=RRreg`.

2. Randomized response models

As indicated above, RR models are a family of different techniques that share the common idea of scrambling an individual's response by means of some sort of random process. Here, we restrict ourselves to compulsory RR models (as opposed to optional models, e.g., Chaud-

huri and Mukerjee 1985) that use a single randomization outcome (as opposed to multiple randomization outcomes, e.g., Ambainis, Jakobsson, and Lipmaa 2004; Mangat and Singh 1990) to assess a single attribute (as opposed to multiple attributes, e.g., Christofides 2005; Tamhane 1981). First, we briefly describe five specific RR designs to illustrate the general idea of the RR technique and to show how RR models are specified in package **RRreg** using two arguments, the RR name `model` and the randomization probability `p`. The package vignette, obtainable by typing `vignette("RRreg", package = "RRreg")`, contains a more detailed list of the implemented models including depictions of the RR designs as probability trees. For a comprehensive overview over different RR variants, see Antonak and Livneh (1995), Chaudhuri and Christofides (2013), and Fox (2015).

2.1. Warner's model and crosswise model

According to the original RRT version by Warner (1965), respondents either answer the sensitive question with probability p or its negation with probability $1 - p$. Therefore, the probability of an affirmative response is given by

$$\lambda = \pi p + (1 - \pi)(1 - p), \tag{1}$$

so that the prevalence can be estimated as $\hat{\pi} = (\hat{\lambda} + p - 1)/(2p - 1)$ based on the relative frequency of observed *yes* responses, $\hat{\lambda} = n_{\text{yes}}/N$.

Recently, Yu *et al.* (2008) proposed to refrain from classical randomization devices such as dice throwing and to use instead a second, irrelevant question with known prevalence p (e.g., "Is your birthday in November or December?", $p = 2/12$). Essentially, the crosswise model prompts participants to provide a single response to both questions simultaneously. Participants are to choose option (A) if the response is either *yes* to both questions or *no* to both questions; or option (B) if one response is *yes* while the other is *no*. The instructions of the crosswise model are easy to understand, do not force participants to respond to the sensitive question directly, and lead to improved prevalence estimates in practice (Jann *et al.* 2012; Hoffmann and Musch 2017; Hoffmann *et al.* 2017; Hoffmann, Diedenhofen, Verschuere, and Musch 2015; but see also Höglinger and Jann 2016 for possible limitations and biases). From a statistical point of view the model is identical to Warner's model, because the probability of responding to option (A) is identical to the probability of responding *yes* as given in Equation 1. In all functions of the package **RRreg**, the model is specified either by `model = "Warner"` or `model = "Crosswise"` and the randomization probability `p = p`.

2.2. Unrelated question technique

In the unrelated question technique (UQT; Horvitz *et al.* 1967), participants are randomly assigned (with probability p) to either the sensitive question or to an independent, nonsensitive question (e.g., "Is your birthday in the first half of the year?") with (known or unknown) prevalence π_{UQ}. Under this model, the probability of observing an affirmative response is thus

$$\lambda = p\pi + (1 - p)\pi_{\text{UQ}}. \tag{2}$$

In package **RRreg**, the UQT with known prevalence is specified using the arguments `model = "UQTknown"` and the randomization probabilities `p = c(p, piUQ)`.

If the prevalence of the nonsensitive question π_{UQ} is unknown, two independent random samples are drawn and different randomization probabilities p_1 and p_2 are applied in order to

RR model	Version	Source	`model =`	Groups	`RRcor`	`RRlog`	`RRmixed`	`RRlin`
Warner's		Warner (1965)	`"Warner"`	1	×	×	×	×
Crosswise		Yu, Tian, and Tang (2008)	`"Crosswise"`	1	×	×	×	×
Unrelated question technique	Known prevalence	Horvitz, Shah, and Simmons (1967)	`"UQTknown"`	1	×	×	×	×
	Unknown prevalence	Horvitz *et al.* (1967)	`"UQTunknown"`	2	×	×		×
Mangat's		Mangat (1994)	`"Mangat"`	1	×	×	×	×
Kuk's card playing		Kuk (1990)	`"Kuk"`	1	×	×	×	×
Forced response		Boruch (1972)	`"FR"`	1	×	×	×	×
Cheating detection	Asymmetric	Clark and Desharnais (1998)	`"CDM"`	2	×			×
	Symmetric	Ostapczuk, Moshagen, Zhao, and Musch (2009a)	`"CDMsym"`	2	×			×
Stochastic lie detector		Moshagen *et al.* (2012)	`"SLD"`	2	×	×		×
Continuous RR (mixture)	Normal	Greenberg, Kuebler Jr, Abernathy, and Horvitz (1971)	`"mix.norm"`	1	×			
	Exponential		`"mix.exp"`	1	×			
	Unknown		`"mix.unknown"`	2	×			
RR misclassification		Van den Hout and Van der Heijden (2002)	`"custom"`	1		×	×	×

Table 1: Randomized response models currently available in the R package **RRreg** (version 0.6.7). The function `RRcor()` computes bivariate correlation tables including one or more RR variables, `RRlog()` predicts a dichotomous RR criterion in a modified logistic regression framework, `RRmixed()` includes mixed effects in the logistic regression based on package **lme4** (Bates *et al.* 2015), and `RRlin()` uses RR variables as predictors in a linear regression of a continuous criterion.

estimate both prevalence rates, π and π_{UQ}, simultaneously. This version of the UQT requires the additional argument `group`, which determines the group membership of each respondent (i.e., a vector of length N with values of 1 and 2; see Section 4). The basic structure of the UQT allows for a generalization to continuous sensitive variables (such as income or number of abortions; Greenberg *et al.* 1971).

2.3. Forced response

Unlike the variants described above, the forced response (FR) model (Boruch 1972; Greenberg, Abul-Ela, Simmons, and Horvitz 1969; Peeters, Lensvelt-Mulders, and Lasthuizen 2010) relies on only one question, but prompts a certain proportion of participants to disregard the question entirely and to provide a prespecified response. In the symmetric dichotomous FR format, respondents are asked to respond *yes* with probability p_y, *no* with probability p_n, or truthfully to the sensitive question with probability $1 - p_y - p_n$, depending on the outcome of the randomization device, so that

$$\lambda = p_y + (1 - p_y - p_n)\pi. \tag{3}$$

This FR design easily generalizes to polytomous response options (e.g., *never*, *sometimes*, *often*) by choosing a randomization device that can take values across all of the possible response categories. In package **RRreg**, the responses across the M categories have to be coded using integers between $0, 1, \ldots, M - 1$. For analysis, the FR model is specified by `model = "FR"` and the unconditional randomization probabilities `p = c(p0, p1, p2, ...)`.

2.4. Stochastic lie detector

Even though RR designs have been shown to reduce response biases, some participants still refuse to respond truthfully (e.g., Höglinger *et al.* 2016; Coutts and Jann 2011; Hoffmann *et al.* 2017). Addressing this issue, a core feature of more recently proposed RR variants is to take instruction non-adherence into account. One of these extended models is the stochastic lie detector (SLD; Moshagen *et al.* 2012), which is based on Mangat's RR procedure (Mangat 1994). The SLD has two free parameters, the proportion π of participants endorsing the sensitive question and the proportion of carriers of the sensitive attribute who respond honestly (t). As in Mangat's variant, anonymity is provided by assigning participants not holding the sensitive attribute to either the sensitive question with probability p or its negation with probability $1 - p$. Participants who in fact hold the sensitive attribute are always asked to respond to the sensitive question. Unlike Mangat's variant, the SLD assumes that only a proportion t of participants holding the sensitive attribute responds according to the RR instructions by responding *yes*, whereas the remaining proportion $(1 - t)$ of this subgroup is assumed to choose the innocuous response (*no*). In order to identify the two free parameters π and t, respondents are randomly assigned to two groups that differ only in the applied randomization probabilities p_1 and p_2 ($p_1 \neq p_2$). The probability of responding *yes* in group i is then given by

$$\lambda_i = \pi t + (1 - \pi)(1 - p_i). \tag{4}$$

In package **RRreg**, the SLD is specified by `model = "SLD"`, `p = c(p1, p2)`, and a vector `group` of length N with values of 1 and 2 (see the empirical example in Section 4).

2.5. Randomized response as misclassification

More generally, RR designs for discrete data can be described as misclassification schemes that add random noise to the true responses (Van den Hout and Van der Heijden 2002). Hence, it is possible to specify each RR design by a misclassification matrix $\boldsymbol{P}$, in which the entries P_{ij} are defined as the conditional probabilities to respond i (e.g., *yes* or *no*) given that the true state of a respondent is j (e.g., cocaine user or non-user). By definition, the misclassification matrix $\boldsymbol{P}$ links the true probabilities of latent states $\boldsymbol{\pi}$ to the expected observed proportions of responses $\boldsymbol{\lambda}$ by

$$\boldsymbol{\lambda} = \boldsymbol{P}\boldsymbol{\pi}. \tag{5}$$

Package **RRreg** allows to specify an RR design by the misclassification matrix $\boldsymbol{P}$ using the arguments `model = "custom"` and, for instance, in case of a binary RR variable, a 2×2 matrix `P = matrix(c(p00, p10, p01, p11), nrow = 2, ncol = 2)`. Similarly, an $M \times M$ misclassification matrix can be used for polytomous responses between $0, 1, \ldots, M-1$. To facilitate a comparison of the prespecified models above and the misclassification approach, the function `getPW()` returns the misclassification matrices of all of the implemented RR designs. For instance, the FR model with unconditional randomization probabilities $p_0 = 0.1$ and $p_1 = 0.15$ to respond *no* or *yes*, respectively, is equivalent to an RR design specified by the misclassification matrix

```
R> getPW(model = "FR", p = c(0.1, 0.15))
```

```
     0   1
0 0.85 0.1
1 0.15 0.9
```

Note that for all RR models that require two independent groups (e.g., the UQT with unknown prevalence), the misclassification probabilities depend on the second parameter and on the group membership, passed by the additional arguments `par2` and `group`, respectively. If the RR procedure is independently applied to several questions, the joint misclassification matrix of these variables is given by the Kronecker product of the misclassification matrices of each individual RR design (Bourke 1982; Van den Hout and Van der Heijden 2002). Hence, univariate and multivariate models for a single discrete RR variable can easily be generalized to multiple discrete RR variables.

3. Multivariate methods for randomized response data

3.1. `RRcor()`: Randomized response correlations

Since the randomization component inherent in any RR model distorts the response variable in a principally predictable way (due to the requirement of a known probability distribution of the randomization process), multivariate analyses of RR data can be performed relying on misclassification and measurement error models (Boruch 1972; Bross 1954; Fuller 2006). Based on this approach, Kraemer (1980) derived an unbiased estimator of correlations using binary RR variables for the Warner and the FR model, which was later generalized by Fox and Tracy (1984) to polytomous and continuous RR variables. This approach allows for

estimating the test-retest reliability of an RR measure or the correlation between two RR variables, which can even differ in scale and RR design.

The method by Fox and Tracy (1984) reformulates the probability of an observed RR response y_i for a participant i as a mixture distribution (similar to the UQT). Either the true underlying state x_i or a value u_i from an independent masking distribution is observed with probability p_t and $(1 - p_t)$, respectively:

$$y_i = \begin{cases} x_i \text{ with probability } p_t, \\ u_i \text{ with probability } (1 - p_t). \end{cases} \tag{6}$$

Since the probability p_t to respond truthfully and the mean and variance of the masking distribution $\boldsymbol{u}$ are known by design, it is possible to estimate the correlation of the true states $\boldsymbol{x}$ with a second (directly or indirectly measured) variable by the method of moments (see Fox and Tracy 1984, for details). In addition to the point estimate of the correlation, the function `RRcor()` provides an option to run a parametric or nonparametric bootstrap (Efron and Tibshirani 1997) to obtain standard errors.

Note that RR models other than the UQT need to be reparameterized to fit the structure in Equation 6. For instance, the FR model with the unconditional randomization probabilities p_1 and p_0 of being asked to respond *yes* or *no* regardless of the true state, respectively, can be reparameterized to $p_t = 1 - p_1 - p_0$. Moreover, the masking distribution of the binary FR is a Bernoulli random variable with mean $\mu_u = p_1/(p_1 + p_0)$, which is the conditional probability to answer *yes* given that the respondent disregarded the sensitive question and responded in a predetermined way. Other RR variants are reparameterized by package **RRreg** accordingly.

3.2. `RRlog()`, `RRmixed()`: Predicting RR variables in a logistic regression

Often, the interest is to predict a binary sensitive outcome measured by RR using continuous or categorical predictors. For this purpose, the standard logistic regression model with the link function $\text{logit}(p) = [\log p/(1 - p)]$ needs to be adapted to take the randomization procedure into account (Maddala 1983; Scheers and Dayton 1988; Van den Hout, Van der Heijden, and Gilchrist 2007). In both the standard and the RR logistic regression, the individual expected prevalence rates π_i are modeled by

$$\pi_i = \text{logit}^{-1}(\boldsymbol{X}_i\boldsymbol{\beta}) = \frac{1}{1 + \exp(-\boldsymbol{X}_i\boldsymbol{\beta})}, \tag{7}$$

where $\boldsymbol{\beta}$ is the vector of regression coefficients and $\boldsymbol{X}_i$ the i-th row of the design matrix $\boldsymbol{X}$.

To obtain the likelihood for the standard logistic regression model, it is assumed that the observed responses $\boldsymbol{y}$ are binomially distributed with means given by the underlying prevalence rates $\boldsymbol{\pi}$ as modeled in Equation 7. In binary RR designs, however, the observed responses are binomially distributed with means given by the expected probabilities $\boldsymbol{\lambda}$ of observing an affirmative response. More specifically, in the general case of expressing an RR design by its misclassification matrix $\boldsymbol{P}$ (Equation 5), this yields the likelihood

$$f(\boldsymbol{y} \mid \beta) = \prod_{i=1}^{N} \binom{n_i}{y_i} (P_{10}(1 - \pi_i) + P_{11}\pi_i)^{y_i} (P_{00}(1 - \pi_i) + P_{01}\pi_i)^{n_i - y_i}, \tag{8}$$

where y_i and n_i are the numbers of affirmative and total responses, respectively.

In other words, the true prevalence rate is transformed by the inverse logit function as in a standard logistic regression, whereas the link function for the likelihood changes according to the specific RR design (Van den Hout *et al.* 2007; Van der Heijden and Van Gils 1996; Kerkvliet 1994). `RRlogit()` employs the expectation-maximization (EM) algorithm by Magder and Hughes (1997) to obtain starting values for a gradient-based search by the R function `optim()`. Asymptotic standard errors are estimated by taking the square root of the diagonal elements of the inverted Fisher information matrix. Similarly, likelihood ratio tests allow for testing parameter restrictions.

The logistic regression belongs to the class of generalized linear models and can thus easily be extended to hierarchical or so-called mixed effects models (Pinheiro and Bates 2000). Such extended models are appropriate if participants are clustered in several groups that differ in their overall intercept, slope, or both. Based on an adjusted link function for RR designs (Van den Hout *et al.* 2007), the function `RRmixed()` uses the estimation routines of package **lme4** (Bates *et al.* 2015) and thus allows for fitting a wide range of mixed effects models (e.g., level-1 and level-2 predictors, crossed random effects; see Bates *et al.* 2015). Currently (i.e., using **RRreg** 0.6.7), however, the application of this method is restricted to one-group RR designs that use the same randomization scheme across all responses and participants.

3.3. `RRlin()`: RR predictors in a linear regression

Instead of using RR variables as a criterion in a logistic regression, it is also possible to include RR variables as predictors in a linear regression. This model was proposed by Van den Hout and Kooiman (2006) based on the interpretation of RR designs as misclassification mechanisms on discrete predictors (Van den Hout and Van der Heijden 2002, cf. Section 2.5). It suffices to state the likelihood of the model for a single RR predictor because several independent RR variables can be modeled using the Kronecker product of their misclassification matrices (Van den Hout and Kooiman 2006). In the following, the first entry of the vector of regression coefficients $\boldsymbol{\beta}$ refers to the RR variable, whereas the remaining entries refer to q directly observable predictors (e.g., variables measured by direct questions).

The likelihood of this linear RR regression model consists of two components. First, the likelihood of the latent distribution $\boldsymbol{\pi}$ of the true states is given by the mixture distribution of the observed responses and depends on the misclassification matrix $\boldsymbol{P}$ and the n_j^* observed frequencies of the $j = 1, \ldots, J$ possible RR responses:

$$\log f_1(\boldsymbol{y} \mid \boldsymbol{\pi}) = \sum_{j=1}^{J} n_j^* \log \sum_{k=1}^{J} P_{jk}\pi_k + C, \tag{9}$$

where C is a constant.

Second, it is assumed that the dependent variable $\boldsymbol{y}$ is conditionally normally distributed with standard deviation σ. For each observation y_i, the expected value is determined by a linear combination of the *true, latent* states on the RR predictors and the *observed* values for directly observed variables (DQ). Because the true states of the participants on the RR variable are unknown, the marginal likelihood of the regression coefficients $\boldsymbol{\beta}$ needs to be computed to remove the uncertainty induced by misclassification,

$$\log f_2(\boldsymbol{y} \mid \boldsymbol{\beta}, \sigma, \boldsymbol{\pi}) = \sum_{i=1}^{N} \log \left[\sum_{j=1}^{J} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(\frac{y_i - \boldsymbol{X}_i^{(j)}\boldsymbol{\beta}}{\sigma} \right)^2 \right) \frac{P_{(i)j}\pi_j}{\sum_{k=1}^{J} P_{(i)k}\pi_k} \right], \tag{10}$$

where the design matrix $\boldsymbol{X}^{(j)}$ has the j-th possible RR response in the first column and the directly observable predictor values in the remaining columns; $P_{(i)j}$ gives the probability to observe the RR response of the i-th person if she were in the j-th true state (see Van den Hout and Kooiman 2006, for details). Eventually, the joint loglikelihood of the whole model is obtained by the sum of Equations 9 and 10. The function `RRlin()` maximizes this likelihood and returns asymptotic standard errors based on the observed Fisher information matrix.

As mentioned in Section 2.5, multiple RR predictors are included by using the Kronecker product of their individual misclassification matrices, thereby taking multicollinearity into account. Note, however, that the randomization devices must be independent of each other and also independent of all sensitive attributes. Such requirements have to be carefully considered especially in RR designs that use irrelevant questions to protect anonymity (Moshagen *et al.* 2011). Unlike logistic RR regression, the method is not restricted to dichotomous RR variables, but can also accommodate polytomous RR variables. However, because a finite-dimensional misclassification matrix $\boldsymbol{P}$ has to be specified, continuous RR variables cannot be used in this framework.

4. Empirical example: Attitudes against minarets in Germany

To illustrate the outlined methods[1], we use data from a study involving 1,621 participants (985 female participants; mean age = 23.7 years) who were asked whether they endorse a ban against the construction of minarets in Germany.[2] The study was conducted in 2010, shortly after a national referendum in favor of a respective law yielded a majority in Switzerland. The referendum result came as a surprise to many, because only a month before the referendum, pre-election polls (based on direct questioning) indicated a clear minority in favor of such law (Fetzer and Soper 2012). The unexpected referendum result led to a heated public debate about the prevalence of attitudes against minarets in Switzerland and German-speaking countries in general.

To test whether the RR technique provides higher – and in this case presumably more valid (Hoffmann and Musch 2017) – prevalence estimates, participants were randomly assigned with a ratio of 4:1 to either the RR ($n = 1,256$) or the DQ format ($n = 365$). In the DQ format, the sensitive question read: "A minaret is a tower next to a mosque from which a muezzin calls Muslims to prayer five times a day. Are you in favor of prohibiting the construction of minarets in Germany?" The SLD (Moshagen *et al.* 2012) was chosen as RR variant. As explained above, respondents actually endorsing a minaret ban were asked to respond truthfully to the sensitive statement ("I am in favor of prohibiting the construction of minarets."). In contrast, the remaining participants were prompted to respond to the sensitive statement with probability $p_1 = 2/12$ and $p_2 = 10/12$ within two subgroups, respectively, but to the negation of the sensitive statement ("I am against prohibiting the construction of minarets.") with counter-probabilities $1 - p_1$ and $1 - p_2$, respectively. Respondents' month of birth (unknown to the investigators) served as a randomization device. In addition, an 11-point self-placement scale ranging from -5 (left) to 5 (right) was used to assess political orientation.

[1] Note that the data set does not contain natural groups that call for a mixed effects logistic RR regression, so we refrain from illustrating `RRmixed()`. However, the syntax only differs from `RRlog` with respect to the argument `formula`, which allows to include random slopes and intercepts (cf. Bates *et al.* 2015).

[2] We are grateful to Sarah Radukic and Jochen Musch for providing the data set.

4.1. Prevalence estimate

First, we consider the univariate prevalence estimates of attitudes against minarets obtained by means of RR and DQ, respectively. The code to load the data into R, to select only the subsample of participants in the RR condition, and to obtain the RR prevalence estimate is

```
R> library("RRreg")
R> data("minarets", package = "RRreg")
R> head(minarets)

  age leftRight rrt condition RRdesign leftRight.c     age.c
1  20        -1   0         0       -1  -0.5070944 -3.669957
2  20        -2   0         2        1  -1.5070944 -3.669957
3  30        -4   0         2        1  -3.5070944  6.330043
4  18         1   0         1        1   1.4929056 -5.669957
5  26        -1   0         2        1  -0.5070944  2.330043
6  20        -3   0         1        1  -2.5070944 -3.669957

R> minarets_RR <- subset(minarets, RRdesign == 1)
R> unimod <- RRuni(response = rrt, data = minarets_RR, model = "SLD",
+    p = c(2/12, 10/12), group = condition)
R> summary(unimod)

Stochastic Lie Detector with p1 = 0.1667 , p2 = 0.8333
Sample size: 564 692

   Estimate   StdErr       z  Pr(>|z|)
pi 0.870695 0.041118  21.175 < 2.2e-16 ***
t  0.635806 0.017344 -20.999 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(for the parameter t, i.e. probability of true responding of carriers,
 the test is H0: t=1; H1: t<1 and the one-sided probability value is given)
```

The data frame `minarets` includes the binary response variable `rrt` ($0 = no$, $1 = yes$) and the grouping variable `condition` (values 1 or 2). The latter determines the applied randomization probability. Moreover, the RR design is specified by the argument `model = "SLD"` and the corresponding randomization probabilities in the vector `p`.

As expected, the RR format yielded a higher prevalence estimate of attitudes against minarets ($\hat{\pi}_{\text{RR}} = 0.87$, SE $= 0.04$) than the DQ format ($\hat{\pi}_{\text{DQ}} = 0.43$, SE $= 0.03$).[3] The difference was significant in a likelihood ratio test restricting the prevalence rate to be equal in both formats ($\chi^2(1) = 84.1$, $p < 0.01$).[4] Moreover, even though the RR procedure secured the

[3]The DQ estimator is simply given by the proportion of affirmative responses, $\hat{\pi} = n_{\text{yes}}/N$.

[4]To test this, we ran a logistic RR regression using `RRlog()` including only the binary predictor variable `RRdesign` (cf. Section 4.3).

full anonymity of respondents, the probability of truthful responding among respondents endorsing a minaret ban was only $\hat{t} = 0.64$ (SE $= 0.02$).

4.2. Correlations: Covariates of attitudes against minarets

The code to estimate the correlation of political left-right orientation and age with the sensitive attribute is:

```
R> cormod <- RRcor(x = minarets_RR[, c("rrt", "age", "leftRight")],
+    models = c("SLD", "d", "d"), p.list = list(c(2/12, 10/12)),
+    group = minarets_RR[, "condition"], bs.type = c("se.n", "pval"),
+    bs.n = 1000, nCPU = 1)
R> cormod

Randomized response variables:
  Variable  RRmodel p
1 rrt       SLD      0.1667 0.8333
2 age       direct
3 leftRight direct

Sample size N = 1256

Estimated correlation matrix:
                rrt       age leftRight
rrt        1.000000 -0.054478  0.593345
age       -0.054478  1.000000 -0.005980
leftRight  0.593345 -0.005980  1.000000

Standard errors from nonparametric bootstrap:
               rrt      age leftRight
rrt             NA 0.118752  0.141630
age       0.118752       NA  0.035393
leftRight 0.141630 0.035393        NA

Two-sided p-values from (separate) parametric bootstrap:
               rrt      age leftRight
rrt             NA 0.623246  0.006006
age       0.623246       NA  0.974975
leftRight 0.006006 0.974975        NA
```

From the data frame `minarets_RR`, we first select three columns, that is, the response to the sensitive question, age, and political left-right orientation. RR models are specified by the vector `models` in order of their appearance (`"d"` indicates directly measured variables). The randomization probabilities for all included RR designs have to be defined in `p.list` in the same order. The remaining two arguments `bs.type` and `bs.n` determine that 1,000 nonparametric bootstrap samples should be drawn to estimate standard errors and p values, respectively. The results show that politically right-oriented participants were more likely to

endorse a ban of minarets ($\hat{\rho} = 0.59$, SE $= 0.14$, $p < 0.01$), whereas no significant relationship to age was observed ($\hat{\rho} = -0.05$, SE $= 0.12$, $p = 0.61$).

4.3. Logistic regression: Predicting attitudes against minarets

In the present example, two substantive research questions can be investigated using logistic regression analyses. First, do the covariates age and left-right orientation independently or interactively predict attitudes against minarets? Second, is the predictive power of any of these variables stronger in the RR compared to the DQ format? The first research question can be investigated by simply fitting the modified logistic regression model for a dependent RR variable including the two (mean-centered) predictors age and left-right orientation:

```
R> mod.log <- RRlog(formula = rrt ~ age.c * leftRight.c, data = minarets_RR,
+    model = "SLD", p = c(2/12, 10/12), group = condition)
R> summary(mod.log)

Call:
RRlog.formula(formula = rrt ~ age.c * leftRight.c, data = minarets_RR,
    model = "SLD", p = c(2/12, 10/12), group = condition)

RR Model:
SLD with p =  0.167,0.833

Model fit:
    n   logLik
 1256 -821.084

                   Estimate    StdErr Wald test Pr(>Chi2,df=1)   deltaG2
(Intercept)         1.89659   0.38104  24.77438        0.00000   74.8273
age.c               0.01364   0.06071   0.05045        0.82227    0.0634
leftRight.c         0.62442   0.15681  15.85642        0.00007   15.9638
age.c:leftRight.c  -0.02039   0.01168   3.04856        0.08081    1.9996
t                   0.65170   0.01800 374.60380        0.00000 1189.5374
                  Pr(>deltaG2)
(Intercept)             <2e-16 ***
age.c                   0.8012
leftRight.c              6e-05 ***
age.c:leftRight.c       0.1573
t                       <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Note that the parameter t is tested against the null hypothesis
        that all participants answered truthfully (i.e., H0: t=1)
```

Similar to the standard regression function `lm()` in R, the argument `formula` defines the model by stating the dependent variable on the left and the independent variables on the

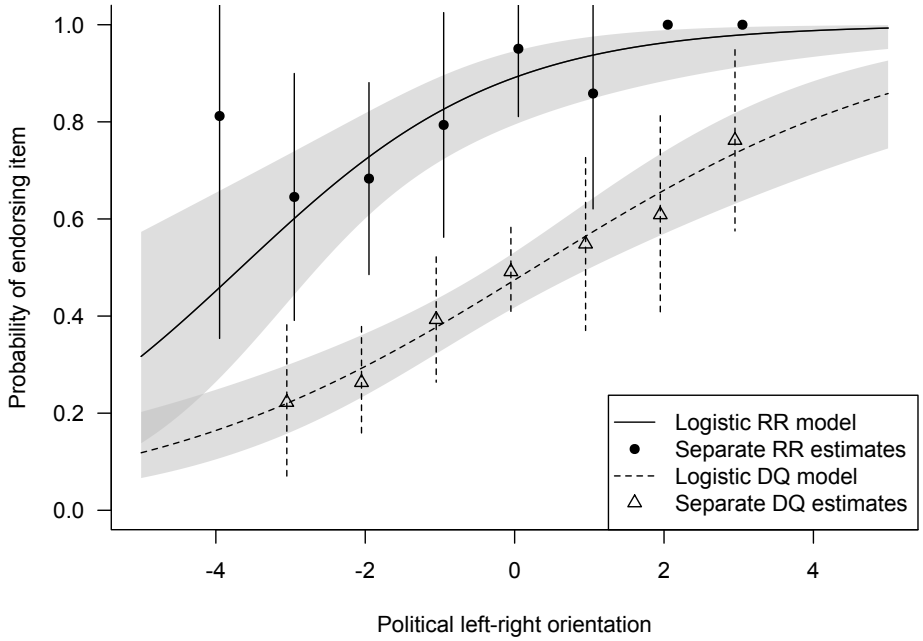

Figure 1: Predicted probability of endorsing a ban against minarets as a function of political left-right orientation (measured on an 11-point self-placement scale) using either RR or DQ. Both logistic regression models include 95% confidence bands. Univariate RR and DQ prevalence estimates are shown for each subgroup of political left-right orientation with $n > 20$ (including 95% confidence intervals).

right, where `age.c * leftRight.c` estimates both main effects and the interaction term. The RR design is specified in the same way as in the case of bivariate correlations.

The results of this analysis are congruent with the correlations obtained in the preceding section. When shifting one point to the right on the political orientation scale, the odds that participants endorsed a ban against the construction of minarets increased by a factor of 1.87 ($\hat{\beta} = 0.62$, SE $= 0.16$, $\chi^2(1) = 16.0$, $p < 0.01$). In contrast, neither age (odds ratio $= 1.01$, $\hat{\beta} = 0.01$, SE $= 0.06$, $\chi^2(1) = 0.1$, $p = 0.80$) nor the interaction between age and left-right orientation (odds ratio $= 0.98$, $\hat{\beta} = -0.02$, SE $= 0.01$, $\chi^2(1) = 2.0$, $p = 0.16$) significantly contributed to the prediction of endorsement. Figure 1 shows the model predictions of the reduced model containing only left-right orientation as a predictor. As a graphical test, we computed prevalence estimates separately for each subgroup of left-right orientation with $n > 20$. Overall, the logistic regression model was in line with these separate estimates and resulted in smaller confidence intervals compared to the univariate analysis as expected.

To investigate the second research question concerning different effects of political left-right orientation in the DQ and the RR response format, we analyzed the complete data set including participants in both questioning conditions. For a comparison of DQ and RR, the function `RRlog()` requires a vector group defining the group membership of each participant; in the present example 0 = DQ, 1 = SLD(p_1), and 2 = SLD(p_2). Moreover, we included an effect-coded variable `RRdesign` for the response format (-1 = DQ, 1 = RR), the mean-centered political left-right orientation, and the interaction of both terms:

```
R> mod.log2 <-RRlog(rrt ~ RRdesign * leftRight.c, data = minarets,
+    model = "SLD", p = c(2/12, 10/12), group = condition)
R> summary(mod.log2)

Call:
RRlog.formula(formula = rrt ~ RRdesign * leftRight.c, data = minarets,
    model = "SLD", p = c(2/12, 10/12), group = condition)

RR Model:
SLD with p =  0.167,0.833 (n=1256) combined with DQ (n=365)

Model fit:
    n    logLik
 1621 -1051.466

                      Estimate    StdErr Wald test Pr(>Chi2,df=1)  deltaG2
(Intercept)            0.76516   0.17804  18.47047        0.00002  34.1689
RRdesign               1.05568   0.17804  35.15931        0.00000  75.5982
leftRight.c            0.47765   0.08001  35.64373        0.00000  29.9496
RRdesign:leftRight.c   0.09676   0.08001   1.46281        0.22648   1.6434
t                      0.65300   0.01805 369.45732        0.00000 237.1835
                     Pr(>deltaG2)
(Intercept)                <2e-16 ***
RRdesign                   <2e-16 ***
leftRight.c                <2e-16 ***
RRdesign:leftRight.c       0.1999
t                          <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Note that the parameter t is tested against the null hypothesis
        that all participants answered truthfully (i.e., H0: t=1).
```

The results show that the odds of endorsing a ban of new minarets increased by a factor of 8.26 in the RR compared to the DQ condition ($\hat{\beta} = 1.06$, SE $= 0.18$, $\chi^2(1) = 75.6$, $p < 0.01$). Across both response formats, a shift of one point to the right on the political orientation scale increased the odds of endorsing the sensitive question by a factor of 1.61 ($\hat{\beta} = 0.48$, SE $= 0.08$, $\chi^2(1) = 29.9$, $p < 0.01$). Moreover, the regression coefficient of the interaction term descriptively supported the contention that left-right orientation is a better predictor in the RR than in the DQ format, but was not statistically significant (odds ratio $= 1.10$, $\hat{\beta} = 0.10$, SE $= 0.08$, $\chi^2(1) = 1.64$, $p = 0.20$). It is evident from Figure 1 that the estimated prevalence of attitudes against minarets was generally higher in the RR than in the DQ condition and increased for politically right-oriented respondents. The (nonsignificant) interaction is visible as the steeper slope in the RR condition.

4.4. Linear regression: Predicting political left-right orientation

Endorsement of the sensitive question concerning a ban against minarets is not a very plausible

causal predictor for any of the available continuous variables. Nevertheless, for illustration purposes, we will use it to predict individual left-right orientation. Additionally, we included age as a second, continuous predictor. The model is estimated by the function `RRlin()` as follows:

```
R> mod.lin <- RRlin(formula = leftRight ~ rrt + age, data = minarets_RR,
+    models = "SLD", p.list=list(c(2/12, 10/12)), group = condition)
R> summary(mod.lin)


Call:
RRlin(formula = leftRight ~ rrt + age, data = minarets_RR, models = "SLD",
    p.list = list(c(2/12, 10/12)), group = condition, fit.n = 10)

Randomized response variables:
  Variable Model p
1 rrt      SLD   c(0.167, 0.833)

Coefficients (beta):
            Estimate   StdErr Wald test Pr(>Chi2,df=1)
(Intercept) -1.58196  0.31619   25.0314         <2e-16 ***
rrt          1.36380  0.28987   22.1360         <2e-16 ***
age         -0.00150  0.00784    0.0365         0.8485
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error (sigma): 1.755; N=1256
```

The model is again specified by a formula including the dependent and independent variables on the left and right side, respectively. RR predictors have to be included first and are specified by the argument `models` in order of their appearance. If more predictors are included than RR models listed, the remaining variables are treated as being directly measured without error (DQ variables).

Consistent with the analyses above, endorsing a ban against minarets significantly predicted left-right orientation, $\hat{\beta} = 1.36$, SE $= 0.29$, $\chi^2(1) = 22.1$, $p < 0.01$. This regression coefficient can be interpreted as a shift of 1.36 points to a more right-oriented political orientation for participants who endorsed the sensitive question. As in the analyses above, age did not significantly predict political left-right orientation ($\hat{\beta} < -0.01$, SE $= 0.01$, $\chi^2(1) = 0.04$, $p = 0.85$).

4.5. Power-analysis simulations

Package **RRreg** provides functions to assess two major practical issues that have to be considered in any RR application: statistical power and robustness. The RR technique necessarily reduces statistical power, because it adds random noise to the responses by design (Ulrich, Schröter, Striegel, and Simon 2012). Therefore, compared to standard correlation and regression analyses, larger sample sizes are required to obtain the same level of precision. The

statistical power of multivariate RR models depends on several factors such as the prevalence π of the sensitive attribute, the type of RR design, the corresponding randomization probabilities, and, of course, on the magnitude of effect.

To estimate the statistical power by means of simulation, `RRsimu()` generates bivariate data assuming either a binary RR variable and a normally distributed covariate or two binary RR variables (relevant only for `RRcor()`). For instance, the code to estimate the power of multivariate methods for the SLD in case of our empirical example is:

```
R> sim.power <- RRsimu(numRep = 1000, n = 1256, pi = 0.87, cor = 0.3,
+    model = "SLD", p = c(2/12, 10/12), groupRatio = 0.45,
+    complyRates = c(0.64, 1), sysBias = c(0, 0), nCPU = 1)
```

Most of the arguments of this function are self-explaining, that is, the number of bootstrap samples (`numRep`), the sample size (`n`), the prevalence of the sensitive attribute (`pi`), the correlation between the sensitive question and the normally distributed covariate (`cor = 0.3`), the RR model (`model`), the randomization probability (`p`), the ratio of participants in the two SLD conditions (`groupRatio`), and the number of parallel cores to use (`nCPU`).

Moreover, the vector `complyRates` defines the probabilities of adherence in both of the latent subgroups, that is, separately for participants holding the sensitive attribute and for those who do not. If any of these two compliance rates is smaller than one, the assumptions of standard RR designs (such as the Warner model) are violated. In contrast, the SLD explicitly models noncompliance of participants holding the sensitive attribute, and we can thus set the first compliance rate equal to the estimated adherence rate of $t = 0.64$ without violating the SLD assumptions. The second compliance rate is set to 1 because the SLD assumes that all noncarriers comply with the instructions. To specify the type of noncompliance in detail, the vector `sysBias` defines the conditional probability of responding *yes* in case of noncompliance for the two latent subgroups. In the code above, `sysBias = c(0, 0)` specifies that noncomplying participants always respond *no*.

As output, `RRsimu()` shows the mean and standard deviation of the estimated parameters by `RRcor()`, `RRlog()`, and `RRlin()`:[5]

```
R> sim.power

SLD ; n= 1256; randomization probability: 0.167, 0.833

Bootstrapped mean and SD of parameters (1000 replications):
                     Mean      SD
pi.true           0.86998 0.00866
pi.RRuni          0.87035 0.03715
piSE.RRuni        0.04101 0.00031
cor.true          0.30011 0.02483
cor.RRcor         0.30028 0.11331
beta.true         0.94178 0.08979
betaSE.true       0.09468 0.00420
```

[5]The optional argument `method` allows to run the simulation only for a single multivariate RR function, e.g., `method = "RRlog"`.

```
beta.RRlog          1.02092 0.44683
betaSE.RRlog        0.45479 0.22557
lincoef.true        0.93581 0.07973
lincoefSE.true      0.08398 0.00288
lincoef.RRlin       0.91595 0.24465
lincoefSE.RRlin     0.25704 0.06528
beta.deltaG2.RRlog 9.42137 5.60029
par2.true           0.63975 0.01364
par2.RRuni          0.64002 0.01614

Power on alpha=0.05 level:
       RRuni(z-test) RRcor(par_bootstrap)       RRlog(LR-test)
               1.000                 0.730                0.826
    RRlin(Wald-test)
               0.892
```

To facilitate a comparison of the loss of efficiency in RR designs, the output includes parameters based on analyzing the true underlying sensitive attribute directly (e.g., `cor.true`, `beta.true`). The results of this simulation show that the power to detect the medium-sized correlation $\rho = 0.3$ is 73.0%, 82.6%, and 89.2% for the functions `RRcor()`, `RRlog()`, and `RRlin()`, respectively.

Often the interest lies in planning the required sample size to obtain a desired level of statistical power prior to conducting a study (see Ulrich *et al.* 2012, for power analyses of univariate tests). For such purposes, the function `powerplot()` plots the estimated power of the multivariate RR methods for different effect sizes as a function of the sample size. In comparison to printed tables or graphs, this function thus allows for a simple power analysis adjusted to the scenario of interest. To demonstrate such an application, we compare the power of a logistic regression using either the DQ format or the symmetric FR model with randomization probabilities of $p_y = p_n = 0.1$ (with a probability of responding to the sensitive question directly of $p_t = 1 - p_y - p_n = 0.8$; see Section 3.1) or $p_y = p_n = 0.2$ ($p_t = 0.6$). The code to estimate and plot the power for the FR design assuming a true prevalence of the sensitive attribute of $\pi = 0.3$ is

```
R> power <- powerplot(numRep = 1000, model = "FR", p = c(0.2, 0.2), pi = 0.3,
+    n = seq(100, 1000, 100), cor = seq(0, 0.25, 0.05), method = "RRlog",
+    alpha = 0.05, nCPU = 1)
R> plot(power)
```

The only arguments that differ from the functions above are the vectors for the sample sizes `n` (a sequence from 100 to 1,000 in steps of 100) and the true correlations `cor` (a sequence from 0 to 0.25 in steps of 0.05). As expected, Figure 2 reveals that the power increases for larger effect and sample sizes. More importantly, the power is largest for the DQ format and decreases with the amount of noise added in each of the two FR designs. Depending on the topic of interest, an RR design with appropriate randomization probabilities can be selected in order to balance the loss of efficiency and the increase in anonymity.

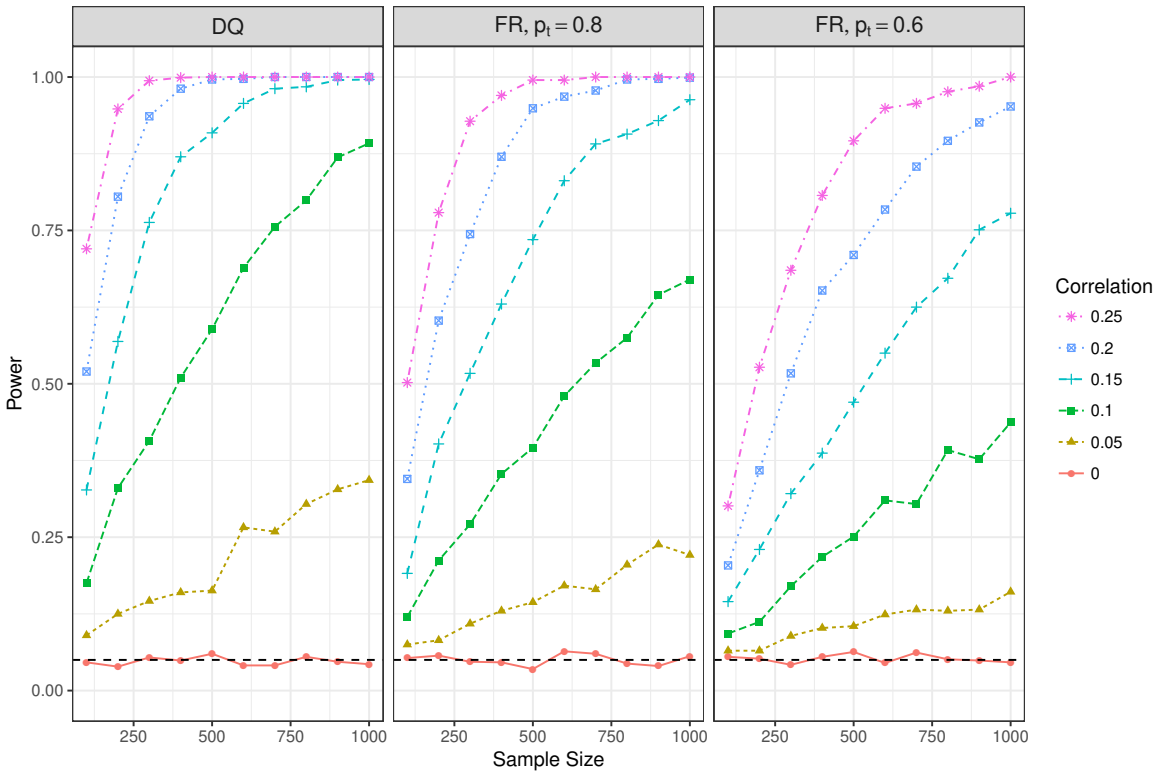

Figure 2: Statistical power of a logistic regression using either the DQ format or the symmetric FR design with different randomization probabilities. Estimates are based on a true prevalence rate of $\pi = 0.3$, and a significance level of $\alpha = 0.05$.

4.6. Robustness simulations for RR designs

Despite the increased anonymity afforded by RR, participants might still engage in self-protecting response behavior, for instance, due to a lack of understanding or mistrust of the protection mechanism (James, Nepusz, Naughton, and Petróczi 2013; Landsheer *et al.* 1999). Therefore, robustness simulations are useful to judge the possible amount of biases in parameter estimates under plausible scenarios of non-adherence.

The function `RRsimu()` can directly be used to assess the robustness of the results given that some respondents do not adhere to the RR instructions in a particular way. For instance, the SLD was derived under the assumption that participants not holding the sensitive attribute always comply with the RR instructions. We can assess how violations of this assumption affect the results, e.g., for the case that only 80% of the participants not holding the sensitive attribute comply, whereas the remaining 20% give self-protecting *no* responses:

```
R> sim.robust <- RRsimu(numRep = 1000, n = 1256, pi = 0.87, cor = 0.3,
+    model = "SLD", p = c(2/12, 10/12), groupRatio = 0.45,
+    complyRates = c(0.64, 0.8), sysBias = c(0, 0), nCPU = 1)
```

Note that, compared to the power simulation above, only the argument `complyRates = c(0.64, 0.8)` changes. Despite this violation of the assumptions of the SLD, the bias in the univariate prevalence estimates is negligible ($\hat{\pi} = 0.9$, SD $= 0.04$, $\hat{t} = 0.62$, SD $= 0.02$),

and the effect on the multivariate statistics is small for correlations ($\rho = 0.3$, $\hat{\rho} = 0.34$, SD $= 0.15$), logistic regression ($\beta_{\log} = 0.94$, $\hat{\beta}_{\log} = 1.07$, SD $= 0.48$), and linear regression ($\beta_{\text{lin}} = 0.94$, $\hat{\beta}_{\text{lin}} = 0.98$, SD $= 0.26$).

To simulate data for more complex scenarios (for example including multiple continuous and/or discrete covariates) appropriate covariance structures can be generated by using any other R package. Given a data set that includes true, latent states, each variable can be passed separately to the function `RRgen()` to simulate the outcome of a specific RR procedure with known compliance rates `complyRates` and systematic biases `sysBias`. For example, the FR design is applied to a vector of true, latent states as follows:

```
R> true <- rbinom(n = 100, size = 1, prob = 0.5)
R> new.dat <- RRgen(trueState = true, model = "FR", p = c(0.1, 0.15),
+    complyRates = c(0.9, 1), sysBias = c(0.5, 0.5))
R> head(new.dat)
```

```
  true comply response
1    0      1        0
2    1      1        1
3    1      1        1
4    1      1        1
5    1      1        1
6    1      1        1
```

Based on the observed RR responses (column `response`), the simulated data can be analyzed using functions such as `RRlog()` or `RRmixed()` to assess the power and robustness of the RR technique in a specific scenario.

5. Discussion

The RR technique may elicit more truthful responding and can yield estimates that are less biased by social desirability (Lensvelt-Mulders *et al.* 2005). Nevertheless, methods to compute correlation and regressions involving RR variables are rarely used in substantive research, probably as a result of their statistical complexity and the scarcity of ready-to-use software. As a remedy, we introduced the package **RRreg** that implements three multivariate methods for the analysis of RR variables: Computing bivariate correlations including one or two RR variables (Fox and Tracy 1984), predicting RR variables in a logistic regression (with or without random effects; Van den Hout *et al.* 2007), and using RR variables as predictors in a linear regression (Van den Hout and Kooiman 2006).

We restricted ourselves to these methods, because these are broadly applicable and yield correlation and regression coefficients that closely resemble those from standard statistical models. However, future versions of package **RRreg** could include multivariate RR methods for more specific situations such as computing sum scores of RR variables (Cruyff 2008), constructing scales comprised of RR items (Himmelfarb 2008), using RR in cross-sectional (Frank, Van den Hout, and Van der Heijden 2009) or cross-national studies (De Jong, Pieters, and Stremersch 2012), combining RR and propensity scoring (Gingerich 2010), or generalizing

single and multidimensional item response theory to RR (Böckenholt and Van der Heijden 2007; Fox 2005; Fox, Entink, and Avetisyan 2014).

Despite the potential superiority of RR over direct questioning when assessing attributes that might be distorted by socially desirable responding, it must be kept in mind that the RR is a large sample technique. This drawback immediately follows from the basic idea of the RR to scramble individual responses by adding random noise, so that the goals of obtaining both efficient and valid estimates (through increased anonymity) necessarily conflict (Ljungqvist 1993). This is even more of an issue considering the advanced, more complex models described herein. The simulation options of the **RRreg** package may be of help in deciding on the required sample size for a particular model. Overall, package **RRreg** broadens the scope of potential applications of the RR technique and provides substantive researchers with a powerful tool to investigate sensitive topics.

References

Ambainis A, Jakobsson M, Lipmaa H (2004). "Cryptographic Randomized Response Techniques." In F Bao, R Deng, J Zhou (eds.), *Public Key Cryptography – PKC 2004*, volume 2947 of *Lecture Notes in Computer Science*, pp. 425–438. Springer-Verlag, Berlin, Germany. doi:10.1007/978-3-540-24632-9_31.

Antonak RF, Livneh H (1995). "Randomized Response Technique: A Review and Proposed Extension to Disability Attitude Research." *Genetic, Social & General Psychology Monographs*, **121**(1), 97–145.

Bates D, Mächler M, Bolker B, Walker S (2015). "Fitting Linear Mixed-Effects Models Using **lme4**." *Journal of Statistical Software*, **67**(1), 1–48. doi:10.18637/jss.v067.i01.

Blair G, Zhou YY, Imai K (2016). ***rr**: Statistical Methods for the Randomized Response Technique*. R package version 1.4, URL https://CRAN.R-project.org/package=rr.

Blank SG, Gavin MC (2009). "The Randomized Response Technique as a Tool for Estimating Non-Compliance Rates in Fisheries: A Case Study of Illegal Red Abalone (Haliotis Rufescens) Fishing in Northern California." *Environmental Conservation*, **36**(2), 112–119. doi:10.1017/s037689290999004x.

Böckenholt U, Van der Heijden PGM (2007). "Item Randomized-Response Models for Measuring Noncompliance: Risk-Return Perceptions, Social Influences, and Self-Protective Responses." *Psychometrika*, **72**(2), 245–262. doi:10.1007/s11336-005-1495-y.

Boruch RF (1972). "Relations among Statistical Methods for Assuring Confidentiality of Social Research Data." *Social Science Research*, **1**(4), 403–414. doi:10.1016/0049-089x(72)90085-3.

Bourke PD (1982). "Randomized Response Multivariate Designs for Categorical Data." *Communications in Statistics – Theory and Methods*, **11**(25), 2889–2901. doi:10.1080/03610928208828430.

Bross I (1954). "Misclassification in 2 × 2 Tables." *Biometrics*, **10**(4), 478–486. doi:10.2307/3001619.

Carroll RJ, Küchenhoff H, Lombard F, Stefanski LA (1996). "Asymptotics for the SIMEX Estimator in Nonlinear Measurement Error Models." *Journal of the American Statistical Association*, **91**(433), 242–250. doi:10.1080/01621459.1996.10476682.

Chaudhuri A, Christofides TC (2013). *Indirect Questioning in Sample Surveys*. Springer-Verlag, Heidelberg, Germany.

Chaudhuri A, Mukerjee R (1985). "Optionally Randomized Response Techniques." *Calcutta Statistical Association Bulletin*, **34**(3–4), 225–229. doi:10.1177/0008068319850311.

Christofides TC (2005). "Randomized Response Technique for Two Sensitive Characteristics at the Same Time." *Metrika*, **62**(1), 53–63. doi:10.1007/s001840400355.

Clark SJ, Desharnais RA (1998). "Honest Answers to Embarrassing Questions: Detecting Cheating in the Randomized Response Model." *Psychological Methods*, **3**(2), 160–168. doi:10.1037/1082-989x.3.2.160.

Coutts E, Jann B (2011). "Sensitive Questions in Online Surveys: Experimental Results for the Randomized Response Technique (RRT) and the Unmatched Count Technique (UCT)." *Sociological Methods & Research*, **40**(1), 169–193. doi:10.1177/0049124110390768.

Coutts E, Jann B, Krumpal I, Näher AF (2011). "Plagiarism in Student Papers: Prevalence Estimates Using Special Techniques for Sensitive Questions." *Journal of Economics and Statistics (Jahrbücher für Nationalökonomie und Statistik)*, **231**(5–6), 749–760. doi:10.1515/jbnst-2011-5-612.

Cruyff MJLF (2008). *Multivariate Analysis of Randomized Response Data*. Ph.D. thesis, Utrecht University, Utrecht.

De Jong MG, Pieters R, Stremersch S (2012). "Analysis of Sensitive Questions across Cultures: An Application of Multigroup Item Randomized Response Theory to Sexual Attitudes and Behavior." *Journal of Personality and Social Psychology*, **103**(3), 543–564. doi:10.1037/a0029394.

Dietz P, Striegel H, Franke AG, Lieb K, Simon P, Ulrich R (2013). "Randomized Response Estimates for the 12-Month Prevalence of Cognitive-Enhancing Drug Use in University Students." *Pharmacotherapy*, **33**(1), 44–50. doi:10.1002/phar.1166.

Donovan JJ, Dwight SA, Hurtz GM (2003). "An Assessment of the Prevalence, Severity, and Verifiability of Entry-Level Applicant Faking Using the Randomized Response Technique." *Human Performance*, **16**(1), 81–106. doi:10.1207/s15327043hup1601_4.

Efron B, Tibshirani RJ (1997). *An Introduction to the Bootstrap*. Chapman and Hall, New York.

Fetzer JS, Soper JC (2012). "An Ecological Analysis of the 2009 Swiss Referendum on the Building of Minarets." In M Helbling (ed.), *Islamophobia in the West: Measuring and Explaining Individual Attitudes*, pp. 101–111. Routledge, London.

Fox JA (2015). *Randomized Response and Related Methods: Surveying Sensitive Data*, volume 58. 2nd edition. Sage Publications, Beverly Hills.

Fox JA, Tracy PE (1984). "Measuring Associations with Randomized Response." *Social Science Research*, **13**(2), 188–197. doi:10.1016/0049-089x(84)90020-6.

Fox JP (2005). "Randomized Item Response Theory Models." *Journal of Educational and Behavioral Statistics*, **30**(2), 189–212. doi:10.3102/10769986030002189.

Fox JP, Entink RK, Avetisyan M (2014). "Compensatory and Non-Compensatory Multidimensional Randomized Item Response Models." *British Journal of Mathematical and Statistical Psychology*, **67**(1), 133–152. doi:10.1111/bmsp.12012.

Frank LE, Van den Hout A, Van der Heijden PGM (2009). "Repeated Cross-Sectional Randomized Response Data: Taking Design Change and Self-Protective Responses into Account." *Methodology*, **5**(4), 145–152. doi:10.1027/1614-2241.5.4.145.

Franke AG, Bagusat C, Dietz P, Hoffmann I, Simon P, Ulrich R, Lieb K (2013). "Use of Illicit and Prescription Drugs for Cognitive or Mood Enhancement among Surgeons." *BMC Medicine*, **11**(102), 1–9. doi:10.1186/1741-7015-11-102.

Fuller WA (2006). *Measurement Error Models*. John Wiley & Sons, New York.

Gingerich DW (2010). "Understanding Off-the-Books Politics: Conducting Inference on the Determinants of Sensitive Behavior with Randomized Response Surveys." *Political Analysis*, **18**(3), 349–380. doi:10.1093/pan/mpq010.

Greenberg BG, Abul-Ela ALA, Simmons WR, Horvitz DG (1969). "The Unrelated Question Randomized Response Model: Theoretical Framework." *Journal of the American Statistical Association*, **64**(326), 520–539. doi:10.1080/01621459.1969.10500991.

Greenberg BG, Kuebler Jr RR, Abernathy JR, Horvitz DG (1971). "Application of the Randomized Response Technique in Obtaining Quantitative Data." *Journal of the American Statistical Association*, **66**(334), 243–250. doi:10.2307/2283916.

Heck DW, Moshagen M (2018). ***RRreg**: Correlation and Regression Analyses for Randomized Response Data*. R package version 0.6.7, URL https://CRAN.R-project.org/package=RRreg.

Hilbig BE, Moshagen M, Zettler I (2015). "Truth Will Out: Linking Personality, Morality, and Honesty through Indirect Questioning." *Social Psychological and Personality Science*, **6**(2), 140–147. doi:10.1177/1948550614553640.

Himmelfarb S (2008). "The Multi-Item Randomized Response Technique." *Sociological Methods & Research*, **36**(4), 495–514. doi:10.1177/0049124107313900.

Hoffmann A, De Puiseau BW, Schmidt AF, Musch J (2017). "On the Comprehensibility and Perceived Privacy Protection of Indirect Questioning Techniques." *Behavior Research Methods*, **49**(4), 1470–1483. doi:10.3758/s13428-016-0804-3.

Hoffmann A, Diedenhofen B, Verschuere B, Musch J (2015). "A Strong Validation of the Crosswise Model Using Experimentally-Induced Cheating Behavior." *Experimental Psychology*, **62**(6), 403–414. doi:10.1027/1618-3169/a000304.

Hoffmann A, Musch J (2017). "Assessing the Validity of Two Indirect Questioning Techniques: A Stochastic Lie Detector versus the Crosswise Model." *Behavior Research Methods*, **48**(3), 1032–1046. doi:10.3758/s13428-015-0628-6.

Höglinger M, Jann B (2016). "More Is Not Always Better: An Experimental Individual-Level Validation of the Randomized Response Technique and the Crosswise Model." *University of Bern Social Sciences Working Paper 18*, University of Bern, Department of Social Sciences.

Höglinger M, Jann B, Diekmann A (2016). "Sensitive Questions in Online Surveys: An Experimental Evaluation of Different Implementations of the Randomized Response Technique and the Crosswise Model." *Survey Research Methods*, **10**(3), 171–187.

Horvitz DG, Shah BV, Simmons WR (1967). "The Unrelated Question Randomized Response Model." In *Proceedings of the Social Statistics Section*, pp. 65–72. American Statistical Association, Washington, DC.

James RA, Nepusz T, Naughton DP, Petróczi A (2013). "A Potential Inflating Effect in Estimation Models: Cautionary Evidence from Comparing Performance Enhancing Drug and Herbal Hormonal Supplement Use Estimates." *Psychology of Sport and Exercise*, **14**(1), 84–96. doi:10.1016/j.psychsport.2012.08.003.

Jann B (2005). ***RRLOGIT**: Stata Module to Estimate Logistic Regression for Randomized Response Data*. URL https://ideas.repec.org/c/boc/bocode/s456203.html.

Jann B (2008). ***RRREG**: Stata Module to Estimate Linear Probability Model for Randomized Response Data*. URL https://ideas.repec.org/c/boc/bocode/s456962.html.

Jann B, Jerke J, Krumpal I (2012). "Asking Sensitive Questions Using the Crosswise Model: An Experimental Survey Measuring Plagiarism." *Public Opinion Quarterly*, **76**(1), 32–49. doi:10.1093/poq/nfr036.

Kerkvliet J (1994). "Cheating by Economics Students: A Comparison of Survey Results." *The Journal of Economic Education*, **25**(2), 121–133. doi:10.2307/1183278.

Kirchner A, Krumpal I, Trappmann M, von Hermanni H (2013). "Measuring and Explaining Undeclared Work in Germany – An Empirical Survey with a Special Focus on Social Desirability Bias." *Zeitschrift für Soziologie*, **42**(4), 291–314. doi:10.1515/zfsoz-2013-0403.

Kraemer HC (1980). "Estimation and Testing of Bivariate Association Using Data Generated by the Randomized Response Technique." *Psychological Bulletin*, **87**(2), 304–308. doi:10.1037/0033-2909.87.2.304.

Krumpal I (2012). "Estimating the Prevalence of Xenophobia and Anti-Semitism in Germany: A Comparison of Randomized Response and Direct Questioning." *Social Science Research*, **41**(6), 1387–1403. doi:10.1016/j.ssresearch.2012.05.015.

Kuk AYC (1990). "Asking Sensitive Questions Indirectly." *Biometrika*, **77**(2), 436–438. doi:10.2307/2336828.

Kwan SSK, So MKP, Tam KY (2010). "Applying the Randomized Response Technique to Elicit Truthful Responses to Sensitive Questions in IS Research: The Case of Software

Piracy Behavior." *Information Systems Research*, **21**(4), 941–959. doi:10.1287/isre.1090.0271.

Lamb CW, Stem DE (1978). "An Empirical Validation of the Randomized Response Technique." *Journal of Marketing Research*, **15**(4), 616–621. doi:10.2307/3150633.

Landsheer JA, Van der Heijden P, Van Gils G (1999). "Trust and Understanding, Two Psychological Aspects of Randomized Response." *Quality and Quantity*, **33**(1), 1–12. doi:10.1023/a:1004361819974.

Lara D, Strickler J, Olavarrieta CD, Ellertson C (2004). "Measuring Induced Abortion in Mexico: A Comparison of Four Methodologies." *Sociological Methods & Research*, **32**(4), 529–558. doi:10.1177/0049124103262685.

Latkin CA, Vlahov D (1998). "Socially Desirable Response Tendency as a Correlate of Accuracy of Self-Reported HIV Serostatus for HIV Seropositive Injection Drug Users." *Addiction*, **93**(8), 1191–1197. doi:10.1046/j.1360-0443.1998.93811917.x.

Lederer W, Küchenhoff H (2006). "A Short Introduction to the SIMEX and MCSIMEX." *R News*, **6**(4), 26–31. URL https://CRAN.R-project.org/doc/Rnews/Rnews_2006-4.pdf.

Lensvelt-Mulders GJLM, Hox JJ, Van der Heijden PGM, Maas CJM (2005). "Meta-Analysis of Randomized Response Research Thirty-Five Years of Validation." *Sociological Methods & Research*, **33**(3), 319–348. doi:10.1177/0049124104268664.

Lensvelt-Mulders GJLM, Van der Heijden PGM, Laudy O, Van Gils G (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**(2), 305–318. doi:10.1111/j.1467-985x.2006.00404.x.

Ljungqvist L (1993). "A Unified Approach to Measures of Privacy in Randomized Response Models: A Utilitarian Perspective." *Journal of the American Statistical Association*, **88**(421), 97–103. doi:10.1080/01621459.1993.10594299.

Locander W, Sudman S, Bradburn N (1976). "An Investigation of Interview Method, Threat and Response Distortion." *Journal of the American Statistical Association*, **71**(354), 269–275. doi:10.1080/01621459.1976.10480332.

Maddala GS (1983). *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge University Press, Cambridge.

Magder LS, Hughes JP (1997). "Logistic Regression When the Outcome Is Measured with Uncertainty." *American Journal of Epidemiology*, **146**(2), 195–203. doi:10.1093/oxfordjournals.aje.a009251.

Mangat NS (1994). "An Improved Randomized Response Strategy." *Journal of the Royal Statistical Society B*, **56**(1), 93–95.

Mangat NS, Singh R (1990). "An Alternative Randomized Response Procedure." *Biometrika*, **77**(2), 439–442. doi:10.1093/biomet/77.2.439.

Mortaz Hejri S, Zendehdel K, Asghari F, Fotouhi A, Rashidian A (2013). "Academic Disintegrity among Medical Students: A Randomised Response Technique Study." *Medical Education*, **47**(2), 144–153. doi:10.1111/medu.12085.

Moshagen M, Hilbig BE, Erdfelder E, Moritz A (2014). "An Experimental Validation Method for Questioning Techniques That Assess Sensitive Issues." *Experimental Psychology*, **61**(1), 48–54. doi:10.1027/1618-3169/a000226.

Moshagen M, Hilbig BE, Musch J (2011). "Defection in the Dark? A Randomized-Response Investigation of Cooperativeness in Social Dilemma Games." *European Journal of Social Psychology*, **41**(5), 638–644. doi:10.1002/ejsp.793.

Moshagen M, Musch J, Erdfelder E (2012). "A Stochastic Lie Detector." *Behavior Research Methods*, **44**(1), 222–231. doi:10.3758/s13428-011-0144-2.

Moshagen M, Musch J, Ostapczuk M, Zhao Z (2010). "Reducing Socially Desirable Responses in Epidemiologic Surveys: An Extension of the Randomized-Response Technique." *Epidemiology*, **21**(3), 379–382. doi:10.1097/ede.0b013e3181d61dbc.

Nordlund S, Holme I, Tamsfoss S (1994). "Randomized Response Estimates for the Purchase of Smuggled Liquor in Norway." *Addiction*, **89**(4), 401–405. doi:10.1111/j.1360-0443.1994.tb00913.x.

Ostapczuk M, Moshagen M, Zhao Z, Musch J (2009a). "Assessing Sensitive Attributes Using the Randomized Response Technique: Evidence for the Importance of Response Symmetry." *Journal of Educational and Behavioral Statistics*, **34**(2), 267–287. doi:10.3102/1076998609332747.

Ostapczuk M, Musch J (2011). "Estimating the Prevalence of Negative Attitudes Towards People with Disability: A Comparison of Direct Questioning, Projective Questioning and Randomised Response." *Disability & Rehabilitation*, **33**(5), 399–411. doi:10.3109/09638288.2010.492067.

Ostapczuk M, Musch J, Moshagen M (2009b). "A Randomized-Response Investigation of the Education Effect in Attitudes Towards Foreigners." *European Journal of Social Psychology*, **39**(6), 920–931. doi:10.1002/ejsp.588.

Peeters CFW, Lensvelt-Mulders GJLM, Lasthuizen K (2010). "A Note on a Simple and Practical Randomized Response Framework for Eliciting Sensitive Dichotomous and Quantitative Information." *Sociological Methods & Research*, **39**(2), 283–296. doi:10.1177/0049124110378099.

Pinheiro JC, Bates DM (2000). *Mixed-Effects Models in S and S-PLUS*. Statistics and Computing. Springer-Verlag, New York. doi:10.1007/b98882.

R Core Team (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Robertson DC, Rymon T (2001). "Purchasing Agents' Deceptive Behavior: A Randomized Response Technique Study." *Business Ethics Quarterly*, **11**(3), 455–479. doi:10.2307/3857849.

Rosenfeld P, Giacalone RA, Knouse SB, Doherty LM, Mitchell Vicino S, Kantor J, Greaves J (1991). "Impression Management, Candor, and Microcomputer-Based Organizational Surveys: An Individual Differences Approach." *Computers in Human Behavior*, **7**(1–2), 23–32. doi:10.1016/0747-5632(91)90026-w.

Royston P, Altman DG, Sauerbrei W (2006). "Dichotomizing Continuous Predictors in Multiple Regression: A Bad Idea." *Statistics in Medicine*, **25**(1), 127–141. doi:10.1002/sim.2331.

Santangeli A, Arkumarev V, Rust N, Girardello M (2016). "Understanding, Quantifying and Mapping the Use of Poison by Commercial Farmers in Namibia – Implications for Scavengers' Conservation and Ecosystem Health." *Biological Conservation*, **204**(B), 205–211. doi:10.1016/j.biocon.2016.10.018.

Scheers NJ, Dayton CM (1988). "Covariate Randomized Response Models." *Journal of the American Statistical Association*, **83**(404), 969–974. doi:10.1080/01621459.1988.10478686.

Simon P, Striegel H, Aust F, Dietz K, Ulrich R (2006). "Doping in Fitness Sports: Estimated Number of Unreported Cases and Individual Probability of Doping." *Addiction Biology*, **101**(11), 1640–1644. doi:10.1111/j.1360-0443.2006.01568.x.

Soeken KL, Damrosch SP (1986). "Randomized Response Technique: Applications to Research on Rape." *Psychology of Women Quarterly*, **10**(2), 119–125. doi:10.1111/j.1471-6402.1986.tb00740.x.

Soeken KL, Macready GB (1986). "Application of Setwise Randomized Response Procedures for Surveying Multiple Sensitive Attributes." *Psychological Bulletin*, **99**(2), 289–295. doi:10.1037//0033-2909.99.2.289.

Solomon J, Jacobson SK, Wald KD, Gavin M (2007). "Estimating Illegal Resource Use at a Ugandan Park with the Randomized Response Technique." *Human Dimensions of Wildlife*, **12**(2), 75–88. doi:10.1080/10871200701195365.

StataCorp (2015). *Stata Statistical Software: Release 14*. College Station.

Tamhane AC (1981). "Randomized Response Techniques for Multiple Sensitive Attributes." *Journal of the American Statistical Association*, **76**(376), 916–923. doi:10.1080/01621459.1981.10477741.

Tourangeau R, Yan T (2007). "Sensitive Questions in Surveys." *Psychological Bulletin*, **133**(5), 859–883. doi:10.1037/0033-2909.133.5.859.

Tracy PE, Fox JA (1981). "The Validity of Randomized Response for Sensitive Measurements." *American Sociological Review*, **46**(2), 187–200. doi:10.2307/2094978.

Ulrich R, Schröter H, Striegel H, Simon P (2012). "Asking Sensitive Questions: A Statistical Power Analysis of Randomized Response Models." *Psychological Methods*, **17**(4), 623–641. doi:10.1037/a0029314.

Umesh UN, Peterson RA (1991). "A Critical Evaluation of the Randomized Response Method: Applications, Validation, and Research Agenda." *Sociological Methods & Research*, **20**(1), 104–138. doi:10.1177/0049124191020001004.

Van den Hout A, Kooiman P (2006). "Estimating the Linear Regression Model with Categorical Covariates Subject to Randomized Response." *Computational Statistics & Data Analysis*, **50**(11), 3311–3323. doi:10.1016/j.csda.2005.06.008.

Van den Hout A, Van der Heijden PGM (2002). "Randomized Response, Statistical Disclosure Control and Misclassification: A Review." *International Statistical Review*, **70**(2), 269–288. doi:10.1111/j.1751-5823.2002.tb00363.x.

Van den Hout A, Van der Heijden PGM, Gilchrist R (2007). "The Logistic Regression Model with Response Variables Subject to Randomized Response." *Computational Statistics & Data Analysis*, **51**(12), 6060–6069. doi:10.1016/j.csda.2006.12.002.

Van der Heijden PGM, Van Gils G (1996). "Some Logistic Regression Models for Randomized Response Data." In A Forcina, GM Marchetti, R Hatzinger, G Falmacci (eds.), *Statistical Modelling: Proceedings of the 11th International Workshop on Statistical Modelling*, pp. 341–348.

Warner SL (1965). "Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias." *Journal of the American Statistical Association*, **60**(309), 63–69. doi:10.2307/2283137.

Wimbush JC, Dalton DR (1997). "Base Rate for Employee Theft: Convergence of Multiple Methods." *Journal of Applied Psychology*, **82**(5), 756–763. doi:10.1037//0021-9010.82.5.756.

Wolter F, Preisendorfer P (2013). "Asking Sensitive Questions: An Evaluation of the Randomized Response Technique versus Direct Questioning Using Individual Validation Data." *Sociological Methods & Research*, **42**(3), 321–353. doi:10.1177/0049124113500474.

Yu JW, Tian GL, Tang ML (2008). "Two New Models for Survey Sampling with Sensitive Characteristic: Design and Analysis." *Metrika*, **67**(3), 251–263. doi:10.1007/s00184-007-0131-x.

Affiliation:

Daniel W. Heck
Department of Psychology
School of Social Sciences
University of Mannheim
68131 Mannheim, Germany
E-mail: heck@uni-mannheim.de
URL: https://dwheck.de/

Morten Moshagen
Research Methods
Institute of Psychology and Education
Ulm University
Albert-Einstein-Allee 47
89081 Ulm, Germany
E-mail: morten.moshagen@uni-ulm.de
URL: http://www.uni-ulm.de/in/psy-pfm/

Journal of Statistical Software http://www.jstatsoft.org/
published by the Foundation for Open Access Statistics http://www.foastat.org/
June 2018, Volume 85, Issue 2 *Submitted:* 2015-12-14
doi:10.18637/jss.v085.i02 *Accepted:* 2017-06-29