Design and Analysis of the Randomized Response Technique

Graeme BLAIR, Kosuke IMAI, and Yang-Yang ZHOU

About a half century ago, in 1965, Warner proposed the randomized response method as a survey technique to reduce potential bias due to nonresponse and social desirability when asking questions about sensitive behaviors and beliefs. This method asks respondents to use a randomization device, such as a coin flip, whose outcome is unobserved by the interviewer. By introducing random noise, the method conceals individual responses and protects respondent privacy. While numerous methodological advances have been made, we find surprisingly few applications of this promising survey technique. In this article, we address this gap by (1) reviewing standard designs available to applied researchers, (2) developing various multivariate regression techniques for substantive analyses, (3) proposing power analyses to help improve research designs, (4) presenting new robust designs that are based on less stringent assumptions than those of the standard designs, and (5) making all described methods available through open-source software. We illustrate some of these methods with an original survey about militant groups in Nigeria.

KEY WORDS: Power analysis; Randomization; Sensitive questions; Social desirability bias.

1. INTRODUCTION

About a half century ago, Warner (1965) proposed the randomized response method as a survey technique to reduce potential bias due to nonresponse and social desirability when asking questions about sensitive behaviors and beliefs. The method asks respondents to use a randomization device, such as a coin flip, whose outcome is unobserved by the interviewer. Depending on the particular design, the randomization device determines which question the respondent answers (Warner 1965; Greenberg, Abul-Ela, and Horvitz 1969; Mangat and Singh 1990; Mangat 1994), the type of expression the respondent uses to answer the sensitive question (Kuk 1990), or if the respondent should give a predetermined response (Boruch 1971; Fox and Tracy 1986). By introducing random noise, the randomized response method conceals individual responses and protects respondent privacy. As a result, respondents may be more inclined to answer truthfully.

Despite the wide applicability of the randomized response technique and the methodological advances, we find surprisingly few applications. Indeed, our extensive search yields only a handful of published studies that use the randomized response method to answer substantive questions (Madigan et al. 1976; Chaloupka 1985; Wimbush and Dalton 1997; Donovan, Dwight, and Hurtz 2003; St John et al. 2012). In contrast, a vast majority of existing studies apply the randomized response method to empirically illustrate its methodological properties by including some substantive examples (e.g., Abernathy, Greenberg, and Horvitz 1970; Chi, Chow, and Rider 1972; Goodstadt and Gruson 1975; Reinmuth and Geurts 1975; Locander, Sudman, and Bradburn 1976; Fidler and Kleinknecht 1977; Lamb and Stem 1978; Tezcan and Omran 1981; Tracy and Fox 1981; Edgell, Himmelfarb, and Duchan 1982; Volicer and Volicer 1982; van der Heijden and van Gils 1996; van der Heijden et al. 2000; Elffers, Van Der Heijden, and Hezemans 2003; Lensvelt-Mulders, Hox, and Van Der Heijden 2005a; Lara et al. 2006; Cruyff et al. 2007; Himmelfarb 2008; De Jong, Pieters, and Fox 2010; Gingerich 2010; Krumpal 2012). This finding is consistent with previous reviews of the literature. Like Umesh and Peterson (1991), a recent review by Lensvelt-Mulders et al. (2005b) concludes that "there have been very few substantive applications of RRTs [randomized response techniques] and that most papers are published to test a variant or illustrate a statistical problem" (p. 325).

In this article, we fill this gap by providing a suite of methodological tools that facilitate the use of randomized response technique in applied research. We begin by reviewing and comparing the standard designs available to researchers (Section 2). We categorize commonly used designs into four basic groups and discuss identification and practical issues by using examples from existing studies. Building on the results in the literature, we then develop various multivariate regression techniques for substantive analyses (Section 3). In particular, we show how to use the randomized response as a predictor as well as the outcome in regression models. We also propose power analyses to help improve research designs and discuss the pros and cons of each design from a practical perspective. Using an original survey about militant groups in Nigeria, we illustrate some of these methodologies (Section 4).

The dearth of substantive applications is unfortunate because there exists empirical evidence that the randomized response method is an effective technique for studying sensitive topics at least in some settings (e.g., Tracy and Fox 1981; van der Heijden et al. 2000; Lara et al. 2004; Rosenfeld, Imai, and Shapiro

Graeme Blair is Pre-Doctoral Fellow, Evidence in Governance and Politics, Columbia University, New York, NY 10027 (E-mail: *graeme.blair@columbia.edu*). Kosuke Imai (E-mail: *kimai@princeton.edu*) is Professor and Yang-Yang Zhou (E-mail: *yz3@princeton.edu*) is Ph.D. student, Department of Politics, Princeton University, Princeton, NJ 08544. The survey research was approved by the Princeton University Institutional Review Board under Protocol #5358 and was supported by the International Growth Centre (RA.2010.12.013; Blair and Imai). Zhou acknowledges support from the National Science Foundation (SES.1148900).

*The proposed methods presented in this article are implemented as part of the R package, `rr: Statistical Methods for the Randomized Response Technique` (Blair, Zhou, and Imai 2015b), which is freely available for download at http://cran.r-project.org/package=rr. The replication archive is available at *http://dx.doi.org/10.7910/DVN/AIO5BR* (Blair, Imai, and Zhou 2015a).

Journal of the American Statistical Association
September 2015, Vol. 110, No. 511, Review
DOI: 10.1080/01621459.2015.1050028

2015; Stubbe et al. 2014). Some researchers find that not only does randomized response lead to the lowest response distortion compared to other indirect questioning methods, but also it is generally well received by both interviewers and respondents (e.g., Locander, Sudman, and Bradburn 1976). While the validation studies remain quite rare given the difficulty of obtaining the ground truth about sensitive behavior and attitudes, a recent study by Rosenfeld, Imai, and Shapiro (2015) reports that the randomized response method recovers the truth well when compared to other methods (but see Kirchner 2015, for less favorable evidence).

On the other hand, other scholars caution that the randomized response procedures may confuse respondents and yield noncompliance, requiring more experienced interviewers for successful implementation (e.g., Holbrook and Krosnik 2010; Coutts et al. 2011; Wolter and Preisendörfer 2013; Höglinger, Jann, and Diekmann 2014). To address these potential problems, many researchers explain the goal of randomized response methods to respondents (e.g., Gingerich 2010). Nevertheless, Coutts and Jann (2011) found that many respondents do not believe the randomized response technique protects anonymity even when they completely understand the instructions.

Thus, to further assuage concerns of respondent noncompliance with randomized response survey instructions, we propose new robust designs that are based on less stringent assumptions than those of the standard designs (Section 5). For example, we propose a design that allows for an unknown degree of noncompliance to instructions. We then develop the same set of methodological tools for these modified designs so that researchers can fit multivariate regression models and conduct power analyses. The new designs should address concerns, expressed frequently by applied researchers, about the standard randomized response techniques and hence further widen the applicability of the methodology.

All methodologies discussed in this article are made available through open-source software, `rr: Statistical Methods for the Randomized Response Technique` (Blair, Zhou, and Imai 2015b), which is freely available for download at the Comprehensive R Archive Network (*http://cran.r-project.org/package=rr*). Other related software packages for randomized response methods include the Stata module `rrlogit` (Jann 2011), the R package `RRreg` (Heck and Moshagen 2014), and the (MC)SIMEX algorithm in the R package `simex` (Lederer and Küchenhoff 2013). Among these packages, `RRreg` is perhaps the most comprehensive and hence is similar to our software though there are some differences (e.g., our estimation strategy is based on the EM algorithm whereas `RRreg` uses the standard optimization routine).

Finally, in this article, we assume simple random sampling and do not explore various theoretical and practical issues that may arise when adopting different survey sampling methods. We also do not consider how randomized response methods can be used together with direct questioning. Chaudhuri (2011) explored these and other issues.

2. BASIC DESIGNS WITH KNOWN PROBABILITY

In this section, we summarize the basic designs of the randomized response technique that have been proposed in the literature. We classify these designs into four types: *mirrored question*, *forced response*, *disguised response*, and *unrelated question*. For each type, we provide a brief explanation, an example, and a discussion about identification. All four designs make two assumptions: (1) the randomization distribution is known to researchers, and (2) respondents comply with the instructions and answer the sensitive question truthfully when prompted. For some randomized response methods, randomization is not explicitly done by the respondent using an instrument such as coin flip. Instead, they may exploit a random variation that already exists (e.g., phone number or birthday). We refer to all of these methods as randomized response techniques.

2.1 Mirrored Question Design

We begin with the classic design introduced by Warner (1965), which we call the *mirrored question design* (in the literature, this design is sometimes called "Warner's method"). The basic idea is to randomize whether or not a respondent answers the sensitive item or its inverse. As a recent example, Gingerich (2010) used this design to measure corruption among public bureaucrats in Bolivia, Brazil, and Chile. The survey interviewed 2859 bureaucrats from 30 different institutions. Each respondent was provided a spinner and then instructed to whirl the device without letting the interviewers know the outcome. The actual instruction is reproduced here:

```
For each of the following questions, please
spin the arrow until it has made at least one
full rotation. If the arrow lands on region
A for a particular question, respond true or
false in the space indicated only with respect
to statement A. If the arrow lands on region B
for that question, respond true or false in
the space indicated only with respect to
statement B. Do not make any marks to indicate
in which region the arrow fell for each
question. Please remember that if you respond
false to a statement in its negative form that
means that the positive form of the statement
is true.
```

If the spinner landed on region A, the respondent answers the following question.

```
A. I have never used, not even once, the
resources of my institution for the benefit
of a political party.
```

If the spinner landed on region B, the respondent answers its inverse.

```
B. I have used, at least once, the resources
of my institution for the benefit of
a political party.
```

Another example of this mirrored design is an ecological study to examine whether members of marine clubs in Australia collected shells without permits from the protected Great Barrier

Reef (Chaloupka 1985). Other applications include whether respondents are in favor of capital punishment (Lensvelt-Mulders, Hox, and Van Der Heijden 2005a) and legalizing marijuana use (Himmelfarb 2008).

It is straightforward to see that the response probability for the sensitive question is identified. Let Z_i be the latent binary response to the sensitive question for respondent i (i.e., the first statement in the aforementioned example). We use p to denote the probability, determined by a randomization device such as a spinner, that respondents are supposed to answer the sensitive question in the original (rather than mirrored) format. Finally, the observed binary response is denoted by Y_i. The key relationship among these variables is given by the following equation,

$$\Pr(Y_i = 1) = p \Pr(Z_i = 1) + (1 - p) \Pr(Z_i = 0). \quad (1)$$

Solving for $\Pr(Z_i = 1)$ yields,

$$\Pr(Z_i = 1) = \frac{1}{2p - 1} \{\Pr(Y_i = 1) + p - 1\}. \quad (2)$$

Thus, so long as p is not equal to $1/2$, the response distribution to the sensitive question is identified.

As an extension of this design, Mangat and Singh (1990) and Mangat (1994) proposed a two-stage procedure to improve efficiency while preserving the computational ease of the estimator. It asks respondents who actually possess the sensitive trait to answer truthfully. Respondents who do not have the sensitive attribute are instructed to use the randomization device to determine which of the mirrored questions they must answer. Thus, all "no" answers are true negatives and only the "yes" answers are distorted (or vice versa). Under this alternative design, noncompliance among respondents with the sensitive trait may be higher because the privacy protection of respondents with the sensitive attribute is completely dependent on the cooperation of the other set of respondents who do not possess the trait (Lensvelt-Mulders et al. 2005b).

2.2 Forced Response Design

We next consider the *forced response* design, which was first introduced by Boruch (1971). Here, we describe a simpler version of this design as introduced by Fox and Tracy (1986). Under this design, randomization determines whether a respondent truthfully answers the sensitive question or simply replies with a forced answer, "yes" or "no." For example, in a study on the prevalence of civilian cooperation with militant groups in southeastern Nigeria, six-sided dice commonly used for games in the region serve as the randomizing device (Blair 2014). The survey interviewed 2457 civilians in villages affected by militant violence. The instructions are reproduced here:

```
For this question, I want you to answer yes
or no. But I want you to consider the number
of your dice throw. If 1 shows on the dice,
tell me no. If 6 shows, tell me yes. But if
another number, like 2 or 3 or 4 or 5 shows,
tell me your own opinion about the question
that I will ask you after you throw the dice.
[TURN AWAY FROM THE RESPONDENT] Now you
throw the dice so that I cannot see what comes
out. Please do not forget the number that
comes out. [ WAIT TO TURN AROUND UNTIL RES-
PONDENT SAYS YES TO: ] Have you thrown the
dice? Have you picked it up?
```

Thus, when the respondent rolls a one, they are forced to respond "no" to the question; when respondents roll a six, they are forced to respond "yes." Finally, when respondents roll two, three, four, or five, they are instructed to truthfully answer the following sensitive question.

```
Now, during the height of the conflict in 2007
and 2008, did you know any militants, like a
family member, a friend, or someone you talked
to on a regular basis. Please, before you
answer, take note of the number you rolled on
the dice.
```

The idea behind the forced response design is straightforward. Because a certain proportion of respondents are expected to respond "yes" or "no" regardless of their truthful response to the sensitive question, the design protects the anonymity of respondents' answers. That is, interviewers and researchers can never tell whether observed responses are in reply to the sensitive question.

As before, let Z_i represent the latent binary response to the sensitive question for respondent i and Y_i represents the observed response (1 for "yes" and 0 for "no"). Suppose further that we use R_i to represent the latent random variable, taking one of the three possible values; $R_i = 1$ ($R_i = -1$) indicating that respondent i is forced to answer "yes" ("no"), and $R_i = 0$ indicating that the respondent is providing a truthful answer Z_i. Then, the forced design implies the following equality,

$$\Pr(Y_i = 1) = p_1 + (1 - p_1 - p_0) \Pr(Z_i = 1), \quad (3)$$

where $p_0 = \Pr(R_i = -1)$ and $p_1 = \Pr(R_i = 1)$. This allows us to derive the probability that a respondent truthfully answers "yes" to the sensitive question,

$$\Pr(Z_i = 1) = \frac{\Pr(Y_i = 1) - p_1}{1 - p_1 - p_0}. \quad (4)$$

This design is the most popular among applied researchers, and numerous examples are found in various disciplines and methodological illustrations. They include a study of xenophobia and anti-Semitism in Germany (Krumpal 2012), fabrication in job applications (Donovan, Dwight, and Hurtz 2003), employee theft (Wimbush and Dalton 1997), social security fraud (van der Heijden and van Gils 1996), sexual behavior and orientation (Fidler and Kleinknecht 1977), vote choice regarding a Mississippi abortion referendum (Rosenfeld, Imai, and Shapiro 2015), illegal poaching among South African farmers (St John et al. 2012), use of performance enhancing drugs (Stubbe et al. 2014), and violation of regulatory laws by commercial firms (Elffers, Van Der Heijden, and Hezemans 2003). Furthermore, De Jong, Pieters, and Fox (2010) expanded this design to allow for ordinal responses (e.g., a Likert scale), which they use to measure the frequency of respondent consumer use of adult entertainment.

2.3 Disguised Response Design

The next design we consider is the *disguised response design*, which was originally proposed by Kuk (1990) (in the literature, this design is sometimes called "Kuk's design"). This design was created to address the problem that under the other randomized response designs some respondents may still feel uncomfortable providing a particular response (e.g., answering "yes") even when interviewers do not know whether they are answering the sensitive question. For example, Edgell, Himmelfarb, and Duchan (1982) used the forced response design to study college students' experiences with homosexuality. By fixing the outcome of the randomization device unbeknownst to the respondents, the researchers found that 25% of the respondents who were forced to reply "yes" by design did not do so. Considering this unsuccessful application of randomized response, van der Heijden and van Gils (1996) suggested that a disguised response design would have been better suited given respondents had difficulties even giving a false "yes" response.

Under the disguised response design, "yes" and "no" are replaced with more innocuous words. This design is best understood with an example. van der Heijden et al. (2000) used the design to study fraud and malingering by employees regarding social welfare provisions in the Netherlands (see also Cruyff et al. 2007). The randomization device consists of two stacks of cards with both black and red cards. In the right or "yes" stack the proportion of red cards is $p = 0.8$ whereas in the left or "no" stack $1 - p = 0.2$. Respondents are asked to draw one card from each stack. Instead of answering "yes" ("no") to a sensitive question, they are instructed to name the color of the card from the right (left) stack. The original instruction reads as follows:

```
I have two stacks of cards and a box behind in
which I place the cards. [GIVE THE BOX TO THE
RESPONDENT AND LOOK AT IT TOGETHER.] In the
box, you find a card on which it is written
what the stack means: the right-hand stack
is the 'yes' stack, and the left-hand stack is
the 'no' stack. [LET INTERVIEWEE LOOK AND GIVE
DIRECTIONS WITH THE NEXT EXPLANATION.] In the
'yes' stack [POINT TO THE RIGHT-HAND STACK]
there are more red cards than in the 'no' stack
[POINT TO THE LEFT-HAND STACK, RESPONDENT MAY
CHECK]. If you want, you may shuffle the two
stacks [SEPARATELY]. Now, please take from each
stack an arbitrary card. You may take the card
on top or from within the stack. [TAKE A CARD
FROM EACH STACK] Nobody but you can see the
colors of your cards; when you mention a card
color, we do not know the stack from which you
took the card. Thus, your privacy is guaranteed:
your answer will always remain a secret. [...]
I propose that we now try out a few questions.
```

Then, respondents answer with "red" or "black" to the set of questions, which include the following:

```
At a social services check-up, have you ever
acted as if you were sicker or
less able to work than you actually are?
Have you ever noticed an improvement in the
symptoms causing your disability, for example
in your present job, in volunteer work, or the
chores you do at home, without informing the
Department of Social Services of this change?
```

The identification strategy for the probability of answering "yes" to a sensitive question is exactly the same as that for the mirrored response design. Let Z_i represent the latent response to the sensitive question with $Z_i = 1$ ($Z_i = 0$) indicating an affirmative (negative) answer. We use p to represent the proportion of red cards in the right or "yes" stack ($1 - p$ the proportion of red cards in the left or "no" stack). Finally, let Y_i denote the observed binary response where $Y_i = 1$ ($Y_i = 0$) represents the reply "red" ("black"). Then, the key relationship between the probability of observing the answer "red" and the probability of affirmative response toward the sensitive item is described by Equation (1), and therefore the latter quantity is given by Equation (2).

2.4 Unrelated Question Design

The final design we consider is the *unrelated question design*, which was developed by Greenberg, Abul-Ela, and Horvitz (1969) and Greenberg et al. (1971). Under this design, randomization determines whether a respondent should answer a sensitive question or an unrelated, nonsensitive question. Unlike the other designs, this design introduces an unrelated question to increase respondents' compliance with survey instruction. Furthermore, Moors (1971) showed that this design is more efficient than the mirror question, and it allows for quantitative responses.

For example, Chi, Chow, and Rider (1972) applied the unrelated question design to study the incidence of induced abortions in Taiwan using pieces from the regionally popular game *Go*. The researchers interviewed a random sample of 2497 women between ages 15 and 49. Census data were used to estimate the proportion of the unrelated, innocuous question about the respondent's year of birth. We reproduce the instructions here:

```
Here is a bag; in it there are stones from the
game 'Go,' some colored black and others white.
Please take one stone out, and see by
yourself what color it is, black or white.
Don't let me know whether it is black or white,
but be sure you know which it is. If you take
a black one, answer the question: ''Have you
ever had an induced abortion?'' If you take a
white one, answer the question:
''Were you born in the lunar year of the
horse?''
```

Similar studies on abortion rates have been conducted in North Carolina (Abernathy, Greenberg, and Horvitz 1970), Mexico (Lara et al. 2004, 2006), and Turkey (Tezcan and Omran 1981). Other applications of the unrelated question include a criminology study of self-reported arrests in Philadelphia (Tracy and Fox 1981), a sociological assessment concerning the concealment of deaths in the household from local authorities in the Philippines (Madigan et al. 1976), and self-reported failure of classes by college students (Lamb and Stem 1978).

Let p denote the probability that respondents receive the sensitive question. This probability is assumed to be known. In the above example, it equals the proportion of black stones in the bag. We use Z_i to denote the binary latent response to the sensitive question with $Z_i = 1$ ($Z_i = 0$) representing the affirmative (negative) answer. Furthermore, let q represent the probability of answering "yes" to the unrelated question. It is assumed that researchers also know this probability: in the aforementioned example, the census data are used to determine it. Then, if we use Y_i to denote the observed binary response, the key estimating equation is given by,

$$\Pr(Y_i = 1) = p \Pr(Z_i = 1) + (1 - p)q. \tag{5}$$

This yields the identification of the response distribution to the sensitive question,

$$\Pr(Z_i = 1) = \frac{1}{p} \{\Pr(Y_i = 1) - (1 - p)q\}.$$

As variants of this design, Yu, Tian, and Tang (2008) and Tan, Tian, and Tang (2009) proposed two designs that do not require a randomizing device: the triangular and crosswise designs. Both designs make use of an unrelated, nonsensitive question (e.g., whether the respondent is born between August and December) that is assumed to be independent of the sensitive item (e.g., whether the respondent is a drug user). The triangular design asks respondents to mark one of two statements: (1) neither characteristics are true, or (2) at least one of the characteristics is true. Relying on the same setup, the crosswise design asks respondents to choose one of the following statements: (1) both or neither characteristics are true, or (2) one of the characteristics is true. Gingerich et al. (2014) used the crosswise design to develop a joint model that combines indirect and direct questioning within the same survey to determine whether a topic is sufficiently sensitive to justify indirect questioning. Other works that study these designs include Coutts et al. (2011), Jann, Jerke, and Krumpal (2011), Höglinger, Jann, and Diekmann (2014), and Korndörfer, Krumpal, and Schmukle (2014).

3. STATISTICAL ANALYSIS OF THE BASIC DESIGNS

In this section, we describe how to analyze data from the randomized response method under the four basic designs reviewed in the previous section. We begin by presenting the likelihood framework for conducting a multivariate regression analysis, an essential tool for researchers who wish to understand the respondent characteristics that are associated with the sensitive attitudes and behavior under investigation. Within this framework, researchers can then generate predicted probabilities for the sensitive item given characteristics. We also show how to use the sensitive attitude or behavior inferred from the multivariate regression analysis as a predictor for an outcome regression model. Finally, we demonstrate how to conduct power analysis for the four basic designs of randomized response method.

3.1 Multivariate Regression Model

The goal of multivariate regression analysis is to characterize how a vector of respondent characteristics X_i is associated with the latent response to the sensitive question Z_i. We define this regression model as

$$\Pr(Z_i = 1 \mid X_i = x) = f_\beta(x), \tag{6}$$

where β is a vector of unknown parameters. A popular choice of the parametric model is the logistic regression, $f_\beta(x) = \exp(x^\top \beta)/\{1 + \exp(x^\top \beta)\}$. Using Equations (1), (3), and (5), we can construct the likelihood function as

$$\mathcal{L}(\beta \mid \{X_i, Y_i\}_{i=1}^n) = \prod_{i=1}^{N} \{c f_\beta(X_i) + d\}^{Y_i} \{1 - (c f_\beta(X_i) + d)\}^{1-Y_i}, \tag{7}$$

where c and d are known constants determined by each of the four basic designs. For example, under the mirrored question design, $c = 2p - 1$ and $d = 1 - p$, where p is the probability of answering the sensitive question in the original format. Table 1 summarizes the relationship between the model parameters (c, d) and the design parameters under each of the four basic designs.

Our contribution here is to point out that all four designs can be analyzed under the single likelihood function given in Equation (7). In the literature, van den Hout, van der Heijden, and Gilchrist (2007) showed that the same likelihood function applies to the forced response and mirrored response designs (see also Scheers and Dayton 1988). van der Heijden and van Gils (1996) developed a similar likelihood framework for the forced response and disguised response designs. Additionally, Warner (1965) considered the linear regression model while Winkler and Franklin (1979) and O'Hagan (1987) explored its Bayesian extensions.

Table 1. Correspondence between design and model parameters. The table shows, for each design of the randomized response method, the correspondence between design and model parameters.

Design	Design parameters	Model parameters
Mirrored question	p: probability of receiving the sensitive question in its original format as opposed to its inverse	$c = 2p - 1$ $d = 1 - p$
Forced response	p: probability of answering truthfully p_1: probability of forced "yes" p_0: probability of forced "no" ($p_0 = 1 - p - p_1$)	$c = p$ $d = p_1$
Disguised response	p: probability of selecting a red card from the "yes" stack	$c = 2p - 1$ $d = 1 - p$
Unrelated question	p: probability of receiving the sensitive question as opposed to the unrelated question q: probability of answering "yes" to the unrelated question (assumed to be independent of covariates, otherwise needs to be modeled)	$c = p$ $d = (1 - p)q$

NOTE: The general model, which can be applied to all four designs, is given in the likelihood function of Equation (7).

It is important to emphasize that an additional assumption is made when applying the likelihood function in Equation (7) to the unrelated question design. Specifically, it is assumed that the response to the unrelated question q is independent of the covariates X_i. In the case of the empirical application discussed in Section 2.4, whether a respondent is born in a certain lunar year is assumed to be independent of whatever covariates that will be included in the model $f_\beta(x)$. If this assumption is relaxed, then the design parameter q must be modeled as a function of covariates as $q_\gamma(X_i)$ where γ is a vector of unknown parameters. This in turn implies that the model parameter d needs to be a function of X_i. The likelihood function for the unrelated question design, then, becomes

$$\mathcal{L}(\beta, \gamma \mid \{X_i, Y_i\}_{i=1}^n) = \prod_{i=1}^N \{p f_\beta(X_i) + (1-p) q_\gamma(X_i)\}^{Y_i} \times \left[1 - \{p f_\beta(X_i) + (1-p) q_\gamma(X_i)\}\right]^{1-Y_i}. \quad (8)$$

To avoid this unnecessary modeling assumption about responses to the unrelated question, researchers should choose an unrelated question whose responses are known to be independent of respondent characteristics.

One approach, which guarantees that the required independence assumption is met, is to employ a two-stage randomization process. For instance, in addition to the first randomization device, which determines whether the respondent answers the sensitive question or the unrelated question (e.g., drawing from the bag of black-and-white stones in Chi, Chow, and Rider (1972)), respondents are instructed to flip a coin. Upon selecting a white stone, the respondent is prompted to answer the following unrelated question, "Did you flip heads?" While this process waives the need to model q, it also adds a layer of complexity to the design procedure.

3.2 Estimation

van den Hout, van der Heijden, and Gilchrist (2007) focused on the generalized linear model framework and used iteratively reweighted least squares. For example, if we assume the logistic regression for $f_\beta(X_i)$, we have

$$\mu_i \equiv c f_\beta(X_i) + d \quad \text{and} \quad g(\mu_i) = \log \frac{\mu_i - d}{c + d - \mu_i} = \beta X_i,$$

where $g(\cdot)$ is a monotonic and differentiable link function with its domain equal to $(d, c+d)$. Then, the standard generalized linear model (GLM) routine can be used to obtain the maximum likelihood estimate of β.

As an alternative and more generally applicable estimation method, we develop the expectation-maximization (EM) algorithm below to maximize the likelihood function in Equation (7) (Dempster, Laird, and Rubin 1977). The advantage of the proposed algorithm is that it only requires the estimation routine for the underlying model $f_\beta(X_i)$ and hence is applicable to a wide range of models beyond the GLMs. While we develop a separate EM algorithm for each design, they all maximize the same observed-data likelihood function given in Equation (7).

3.2.1 Mirrored Question Design. We first develop the EM algorithm under the mirrored question design. Let the latent indicator variable $T_i = 1$ ($T_i = 0$) denote the scenario where respondent i answers the sensitive question in the original (mirrored) format. Under this design, the complete-data likelihood function is given as follows:

$$\mathcal{L}_{\text{com}}(\beta \mid \{Y_i, T_i, X_i\}_{i=1}^n) = \prod_{i=1}^N f_\beta(X_i)^{T_i Y_i + (1-T_i)(1-Y_i)} \times \{1 - f_\beta(X_i)\}^{T_i(1-Y_i)+(1-T_i)Y_i}.$$

The E-step consists of calculating the following conditional expectations:

$$\mathbb{E}(T_i \mid X_i = x, Y_i = y) = \frac{p f_\beta(x)^y (1 - f_\beta(x))^{1-y}}{p f_\beta(x)^y (1 - f_\beta(x))^{1-y} + (1-p) f_\beta(x)^{1-y} (1 - f_\beta(x))^y}.$$

Then, the M-step maximizes the following objective function with respect to β,

$$\sum_{i=1}^n \{1 - Y_i - (1 - 2Y_i) w_T(X_i, Y_i)\} \log f_\beta(X_i) + \{Y_i + (1 - 2Y_i) w_T(X_i, Y_i)\} \log(1 - f_\beta(X_i)),$$

where $w_T(X_i, Y_i) = \mathbb{E}(T_i \mid X_i, Y_i)$. Given the starting values for β, the algorithm proceeds by alternating the E-step (using the values of β from the previous iteration) and the M-step. In particular, the M-step can be implemented via a weighted regression fitting routine for $f_\beta(x)$.

Finally, we can use the following equation to calculate the posterior prediction of latent responses to the sensitive question for each respondent in the sample, that is,

$$\Pr(Z_i = 1 \mid X_i = x, Y_i = y) = \frac{p^y (1-p)^{1-y} f_\beta(x)}{p^y (1-p)^{1-y} f_\beta(x) + p^{1-y} (1-p)^y (1 - f_\beta(x))}. \quad (9)$$

3.2.2 Forced Response Design. Next we consider the forced response design. Let R_i denote the latent randomization variable where $R_i = 1$ ($R_i = -1$) indicates that respondents are forced to answer "yes" ("no") and $R_i = 0$ implies that the respondent answers the sensitive question truthfully. Then, the complete-data likelihood function is given by

$$\mathcal{L}_{\text{com}}(\beta \mid \{X_i, Y_i, R_i\}_{i=1}^n) \propto \prod_{i=1}^n \{f_\beta(X_i)^{Y_i} (1 - f_\beta(X_i))^{1-Y_i}\}^{\mathbf{1}\{R_i = 0\}}, \quad (10)$$

where $Y_i = Z_i$ when $R_i = 0$ and the likelihood function is constant in β when $R_i \neq 0$.

The E-step is given by the following conditional expectations:

$$\mathbb{E}(\mathbf{1}\{R_i = 0\} \mid X_i = x, Y_i = y) = \frac{p f_\beta(x)^y (1 - f_\beta(x))^{1-y}}{p f_\beta(x)^y (1 - f_\beta(x))^{1-y} + p_1^y p_0^{1-y}}.$$

Then, the objective function for the M-step is

$$\sum_{i=1}^n w_R(X_i, Y_i) \{Y_i \log f_\beta(X_i) + (1 - Y_i) \log(1 - f_\beta(X_i))\},$$

where $w_R(X_i, Y_i) = \mathbb{E}(\mathbf{1}\{R_i = 0\} \mid X_i, Y_i)$. The algorithm iterates between the E and M steps where the latter is carried out by fitting the weighted regression model.

Finally, we use the following conditional expectation to calculate the posterior prediction of responses to the sensitive question for each respondent in the sample:

$$\Pr(Z_i = 1 \mid X_i = x, Y_i = y) = \frac{(p + p_1)^y p_0^{1-y} f_\beta(x)}{p f_\beta(x)^y (1 - f_\beta(x))^{1-y} + p_1^y p_0^{1-y}}. \tag{11}$$

3.2.3 Disguised Response Design. For the disguised response design, the latent response to the sensitive item, Z_i, determines whether a respondent draws a card from the "yes" stack ($Z_i = 1$) or "no" stack ($Z_i = 0$). In each stack, the probability of drawing a "red" card ($Y_i = 1$) is determined by p and $1 - p$ for the "yes" and "no" stacks, respectively. Thus, the complete-data likelihood function is given by

$$\mathcal{L}_{\text{com}}(\beta \mid \{X_i, Y_i, Z_i\}_{i=1}^n) = \prod_{i=1}^n \{f_\beta(X_i) p^{Y_i} (1 - p)^{1-Y_i}\}^{Z_i} \times \{(1 - f_\beta(X_i)) p^{1-Y_i} (1 - p)^{Y_i}\}^{1-Z_i}.$$

Then, the E-step of the EM algorithm is given by

$$E(Z_i \mid X_i = x, Y_i = y) = \frac{f_\beta(X_i) p^y (1 - p)^{1-y}}{f_\beta(X_i) p^y (1 - p)^{1-y} + (1 - f_\beta(X_i))(1 - p)^y p^{1-y}},$$

which also gives the posterior prediction of response to the sensitive question. Finally, the objective function for the M-step is given by

$$\sum_{i=1}^n w_Z(X_i, Y_i) \log f_\beta(X_i) + (1 - w_Z(X_i, Y_i)) \log(1 - f_\beta(X_i)),$$

where $w_Z(X_i, Y_i) = \mathbb{E}(Z_i \mid X_i, Y_i)$.

3.2.4 Unrelated Question Design. Finally, we develop an EM algorithm for the unrelated question design. Bourke and Moran (1988) proposed the EM algorithm for estimating the population proportion of affirmatively answering the sensitive question. Here, we generalize their algorithm for multivariate regression analysis. Let S_i denote the latent binary variable, which indicates whether respondent i answers the sensitive item ($S_i = 1$) or the unrelated question ($S_i = 0$). Then, the complete-data likelihood function is given by

$$\mathcal{L}_{\text{com}}(\beta, \gamma \mid \{X_i, Y_i, S_i\}_{i=1}^n) = \prod_{i=1}^n \{f_\beta(X_i)^{Y_i} (1 - f_\beta(X_i))^{1-Y_i}\}^{S_i} \times \{q_\gamma(X_i)^{Y_i} (1 - q_\gamma(X_i))^{1-Y_i}\}^{1-S_i}.$$

The E-step is given by the following conditional expectations:

$$\mathbb{E}(S_i \mid X_i = x, Y_i = y) = \frac{p f_\beta(x)^y (1 - f_\beta(x))^{1-y}}{p f_\beta(x)^y (1 - f_\beta(x))^{1-y} + (1 - p) q_\gamma(x)^y (1 - q_\gamma(x))^{1-y}}.$$

Given this E-step, the M-step maximizes the following objective function:

$$\sum_{i=1}^n w_S(X_i, Y_i)\{Y_i \log f_\beta(X_i) + (1 - Y_i) \log(1 - f_\beta(X_i))\} + (1 - w_S(X_i, Y_i))\{Y_i \log q_\gamma(X_i) + (1 - Y_i) \log(1 - q_\gamma(X_i))\},$$

where $w_S(X_i, Y_i) = \mathbb{E}(S_i \mid X_i = x, Y_i = y)$. This step is done by fitting the weighted regression models for $f_\beta(X_i)$ and $q_\gamma(X_i)$, separately.

Finally, under the unrelated question design, the posterior prediction of responses to the sensitive question cannot be calculated unless one models the association between responses to the sensitive question and those to the unrelated question, conditional on the respondent characteristics X_i. On the other hand, if we assume the conditional independence between them given X_i, then the posterior probability is given by

$$\Pr(Z_i = 1 \mid X_i = x, Y_i = y) = \frac{\{py + (1 - p) q_\gamma(x)^y (1 - q_\gamma(x))^{1-y}\} f_\beta(x)}{p f_\beta(x)^y (1 - f_\beta(x))^{1-y} + (1 - p) q_\gamma(x)^y (1 - q_\gamma(x))^{1-y}}.$$

3.3 Using Randomized Response as a Predictor

In many cases, researchers wish to use randomized response as a predictor in an outcome regression. Imai, Park, and Greene (2015) developed such a method for the item count technique (or list experiment). Here, we apply the same modeling strategy to the randomized response methods. To illustrate, we consider the forced response design although the same idea can be applied to the other designs. Let V_i represent the outcome variable of interest, and suppose that researchers are interested in fitting the following outcome regression model, $g_\theta(V_i \mid X_i, Z_i)$, where θ is comprised of the parameters of the outcome model. For example, if the outcome model is the normal linear regression, we have $g_\theta(V_i \mid X_i, Z_i) = \mathcal{N}(\alpha + \gamma^\top X_i + \delta Z_i, \sigma^2)$, where $\theta = (\alpha, \gamma, \delta, \sigma^2)$, Z_i is the latent randomized response variable, and the coefficient of interest is δ.

Since Z_i is not directly observed, we develop an EM algorithm to fit this model. We begin by assuming that V_i and Y_i are conditionally independent given X_i. This assumption can be relaxed by modeling their joint distribution, but here we maintain this assumption for the sake of simplicity. Then, the (observed-data) likelihood function for the combined model is given by

$$\mathcal{L}(\theta, \beta \mid \{V_i, X_i, Y_i\}_{i=1}^n) = \prod_{i=1}^N \Big\{ f_\beta(X_i) g_\theta(V_i \mid X_i, 1)(1 - p_0)^{Y_i} p_0^{1-Y_i} + (1 - f_\beta(X_i)) g_\theta(V_i \mid X_i, 0) p_1^{Y_i} (1 - p_1)^{1-Y_i} \Big\}. \tag{12}$$

With R_i denoting the latent randomization variable indicating whether respondents answer the sensitive question, the complete-data likelihood function is given as

$$\mathcal{L}_{\text{com}}(\theta, \beta \mid \{V_i, X_i, R_i, Y_i\}_{i=1}^n) \propto \prod_{i=1}^N \left[\{g_\theta(V_i \mid X_i, 1) f_\beta(X_i)\}^{Y_i} \times \{g_\theta(V_i \mid X_i, 0)(1 - f_\beta(X_i))\}^{1-Y_i}\right]^{\mathbf{1}\{R_i = 0\}},$$

where $Y_i = Z_i$ when $R_i = 0$, and the likelihood function is constant in β when $R_i \neq 0$. The E-step is given by the following conditional expectation:

$$\mathbb{E}(\mathbf{1}\{R_i = 0\} \mid X_i = x, Y_i = y, V_i = v) = \frac{p g_\theta(v \mid x, y) f_\beta(x)^y (1 - f_\beta(x))^{1-y}}{p g_\theta(v \mid x, y) f_\beta(x)^y (1 - f_\beta(x))^{1-y} + p_1^y p_0^{1-y} \{g_\theta(v \mid x, 1) f_\beta(x) + g_\theta(v \mid x, 0)(1 - f_\beta(x))\}}.$$

Finally, the M-step maximizes the following complete-data log-likelihood function:

$$\sum_{i=1}^{n} w_R(X_i, Y_i, V_i) \cdot [Y_i\{\log f_\beta(X_i) + \log g_\theta(V_i \mid X_i, 1)\} + (1 - Y_i)\{\log(1 - f_\beta(X_i)) + \log g_\theta(V_i \mid X_i, 0)\}],$$

where $w_R(X_i, Y_i, V_i) = \mathbb{E}(\mathbf{1}\{R_i = 0\} \mid X_i, Y_i, V_i)$.

3.4 Power Analysis

When choosing among the aforementioned four basic designs and determining the model parameters under each design, one important consideration is statistical efficiency. Here, we show how to conduct power analysis under each design. The literature appears to contain surprisingly few results about efficiency and power analysis. The only relevant work we find is Lakshmi and Raghavarao (1992) who derived a power function to test the probability of respondent noncompliance under a mirrored question design. While others compare efficiency across various designs (Moors 1971; Pollock and Bek 1976; Scheers and Dayton 1988; Umesh and Peterson 1991; Lensvelt-Mulders et al. 2005b), they fall short of providing a unified framework for conducting power analysis to help applied researchers design randomized response surveys. Our analysis fills this important gap in the literature.

Without loss of generality, we consider the likelihood function in Equation (7) with no covariates, that is, $f = f_\beta(1) = \exp(\beta)/\{1 + \exp(\beta)\}$. Again, this f is the probability of possessing the sensitive trait. The unified model introduced in Section 3.1 makes this analysis straightforward. To begin, we derive the Fisher information with respect to f under this unified model,

$$\mathcal{I}(c, d, f) = \mathbb{E}\left[\left(\frac{\partial}{\partial f} \log L\left(f \mid \{Y_i\}_{i=1}^n\right)\right)^2\right] = \frac{c^2}{(cf + d)\{1 - (cf + d)\}}. \tag{13}$$

In addition, the standard error of $\hat{f}$ is given by

$$\sigma(c, d, f, n) = \frac{1}{c\sqrt{n}} \sqrt{(cf + d)\{1 - (cf + d)\}}, \tag{14}$$

where n is the sample size. For each design, we can rewrite both the Fisher information and standard error as the function of design parameters using the relationships between the model and design parameters given in Table 1.

Finally, we derive power functions under all designs. The power function determines the probability that a test procedure will reject a null hypothesis $H_0 : f = f_0$ at significance level α when the true value of f is equal to f^*. We first derive an approximate power function for a one-sided hypothesis test where the null hypothesis is $H_0 : f = f_0$ and the alternative hypothesis is either $H_1 : f > f_0$ or $H_1 : f < f_0$,

$$\psi(c, d, n, f_0, f^*, \alpha) = 1 - \Phi\left[\frac{f_0 - f^* + \Phi^{-1}(1 - \alpha)\sigma(c, d, f_0, n)}{\sigma(c, d, f^*, n)}\right], \tag{15}$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution. Similarly, the power function for a two-sided hypothesis test where the alternative hypothesis is $H_1 : f \neq f_0$ is given by

$$\psi(c, d, f_0, f^*, n, \alpha) = 1 - \Phi\left[\frac{f_0 - f^* + \Phi^{-1}(1 - \alpha/2)\sigma(c, d, f_0, n)}{\sigma(c, d, f^*, n)}\right] + \Phi\left[\frac{f_0 - f^* - \Phi^{-1}(1 - \alpha/2)\sigma(c, d, f_0, n)}{\sigma(c, d, f^*, n)}\right]. \tag{16}$$

Several notable findings follow from these results. First, for a fixed sample size, significance level, and functional form, the power for any two designs that have identical values of c and d will also be identical. This means that the mirrored question and disguised response designs with shared design parameter p will have the same statistical power. In addition, for the forced response and unrelated question designs, the power will be identical when they share the same design parameter p and the forced response parameter p_1 is equal to $(1 - p) \cdot q$ for the unrelated question design.

Now we can compare the power across designs and for different parameter values within each design. Figure 1 displays a comparison for four designs with several realistic design parameter values for each design. There are three notable implications of these comparisons. First, the mirrored and disguised designs have the least power when p is close to one half (note the design cannot be used with $p = 0.5$). For a sample size of 500 and a proportion of "yes" responses to the sensitive item of 0.1, for example, power reaches the typical threshold of 0.8 only when $p \leq 0.25$ or $p \geq 0.75$ (see top left plot, Figure 1).

Second, higher values of p for both the forced response and unrelated question designs yield higher power to detect the sensitive responses. This makes sense: the higher the value of p, the less noise unrelated to the sensitive item responses that is introduced. For example, with a sample size of 1000 and a proportion of "yes" responses to the sensitive item of 0.1, power only reaches the threshold of 0.8 when $p \geq 0.4$.

Third, given a choice of p, it is optimal to either choose small (large) values or large (small) values of p_1 (p_0). That is, the further p_1 and p_0 are from 0.5 in either direction, the higher the power. Figure A.1 in the Appendix displays the power for the forced design with varying values of p_1. For any value of p, the higher the p_1 the lower the power until 0.5, when the relationship reverses. These findings also yield design advice for the unrelated question design. Since $p_1 = (1 - p) \cdot q$, we also know that values of $(1 - p) \cdot q$ closer to 0 and 1 are preferred to values closer to 0.5. For example, for a study with a sample size of 2500, a proportion of "yes" responses to the sensitive item of 0.1, and p set to 0.2, power only reaches the standard threshold of 0.8 when $p_1 < 0.2$ or when $p_1 = 0.8$ (see bottom left plot, Figure A.1 in the Appendix).

3.5 Comparison of the Basic Designs

We now compare the four basic designs of randomized response technique from the point of view of applied researchers. This is summarized in Table 2. While both the mirrored question and forced response designs are the simplest to implement and understand, they have shortcomings. The mirrored question design may suffer from low respondent confidence because both question options—the sensitive item and its complement—are sensitive in nature. Thus, the respondent must understand how the method works, whereas a greater degree of random noise is introduced in the other designs. Confidence may be similarly reduced in the forced response design because although random noise is introduced by the design, the respondent must still respond "yes" in some circumstances, which may still be sensitive depending on the context.

Statistical power provides another metric for choosing between the designs. However, when the design parameters are unconstrained, no design dominates any other in terms of statistical power. There are some values of p, for example, that make the forced response design preferable to the mirrored design and other values of p for which the reverse is true. Typically, practical considerations such as the limited availability or suitability of certain randomization devices places constraints on the feasible values of the design parameters. In such cases, the power comparisons such as Figures 1 and A.1 provided in the Appendix may yield preferable designs. For example, if the researcher can only use values of p below 0.25, the mirrored question design and the disguised response design dominate the forced response design with $p_1 = 0.1$ or $p_1 = 0.5$ and the unrelated question design with $(1 - p) \cdot q = 0.1$ or $(1 - p) \cdot q = 0.5$.

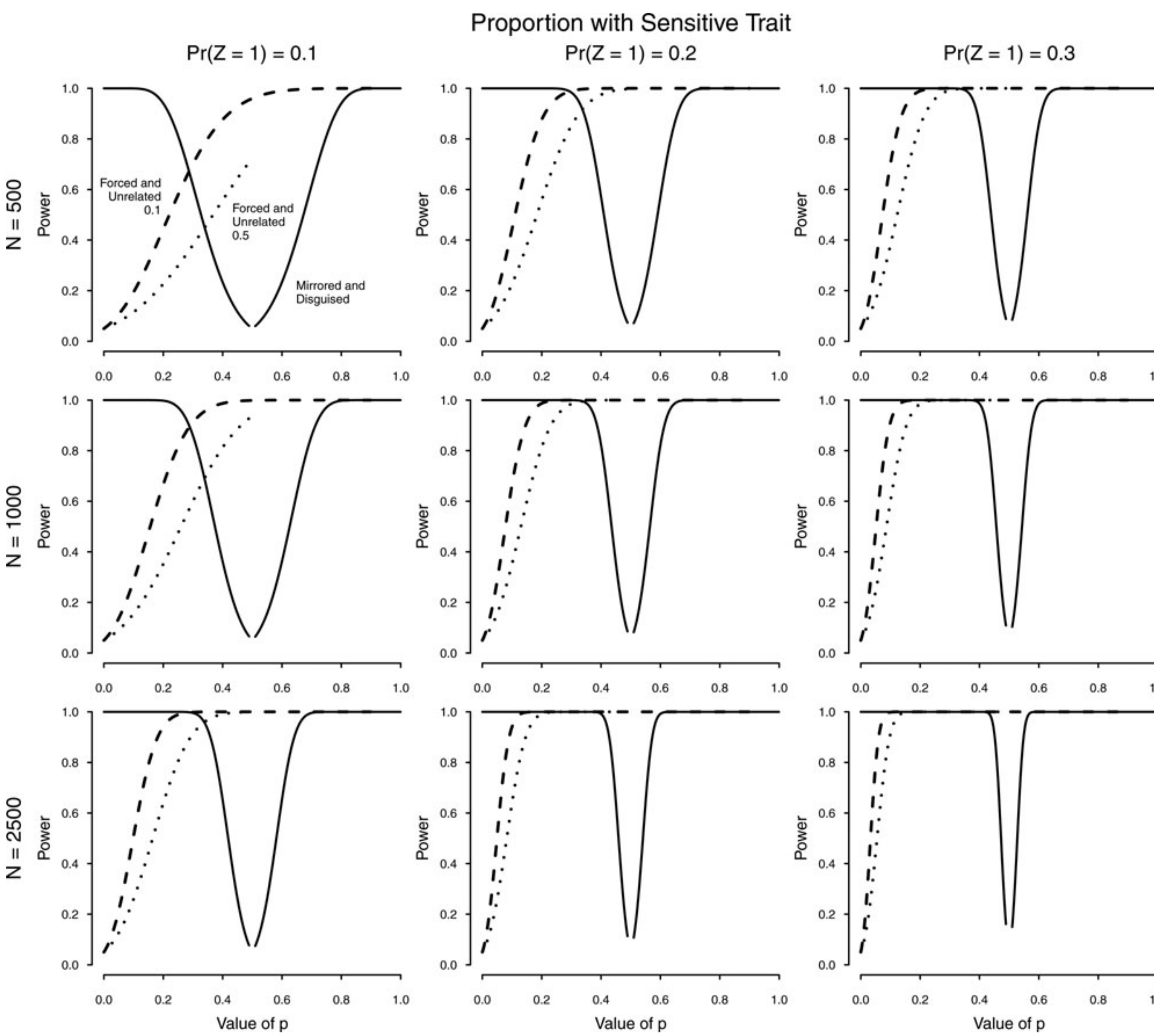

Figure 1. Comparison of power across the four standard designs. First, the power for the forced response and unrelated question design with $p_1 = (1 - p) \cdot q = 0.1$ is displayed (dashed lines). Second, the power for these designs with $p_1 = (1 - p) \cdot q = 0.5$ is displayed (dotted lines). Third, the power for the mirrored question and disguised response designs is displayed (solid lines).

Table 2. Comparison of four basic randomized response designs

Design	Randomization determines	Pros	Cons
Mirrored question	Whether answers sensitive item ("I have the sensitive trait") or its inverse ("I do not have the sensitive trait")	Simple implementation	Low respondent confidence in the answer being hidden
Forced response	Whether answers sensitive item or with forced "yes" or "no"	Simple implementation	Respondents with forced "yes" may fail to say "yes" due to concern that their response might be interpreted as an affirmative admission to the sensitive item
Disguised response	Order of red and black cards in two decks of cards. Respondent states the color chosen from the right deck for "yes" to the sensitive item and the color chosen from the left deck for "no"	Best for items where even saying "yes" out loud is sensitive	Complicated randomization device requires in-person implementation
Unrelated question	Whether answers sensitive item or unrelated, nonsensitive item	High respondent confidence in the answer being hidden	The response to the unrelated question must be either independent of respondent characteristics or modeled

There are also randomization devices that may remove these constraints, such as the use of spinners described in Gingerich (2010).

Ultimately, the choice of the design can be determined by an assessment of the practical constraints in the research context and by careful pilot testing of one or more designs. Pilot testing may help researchers identify the nature of the sensitivity, and, for example, point them to the disguised response design because respondents are unwilling to answer "yes" even with the protections of the mirrored question or forced response design. In addition, the researcher can learn how sensitive the question is for respondents and use this to determine how much protection is needed through the choice of p, for example.

4. EMPIRICAL ILLUSTRATION WITH THE FORCED RESPONSE DESIGN

For empirical illustration, we apply some of the methodologies proposed above to an original survey we conducted in Nigeria in 2013. A goal of the survey is to estimate the proportion of the population who knew or came into regular contact with armed groups. Disclosing social connections with members of armed groups was tremendously sensitive because it could have put the respondent or the former armed group member in danger. When asked such a sensitive question directly, the respondent would likely refuse to answer or lie and respond that they had no social connection regardless of their truthful experience. All of the analyses in this section are carried out with our accompanying open-source software package `rr`.

4.1 Design

To address these concerns of nonresponse and social desirability bias, we used the forced response design of the randomized response method. This technique allowed us to protect the anonymity of each individual-level response about whether respondents held social connections to armed groups. Respondents would thus be more willing to respond honestly to the question. A survey was administered to a random sample of 2457 respondents from 204 communities that are representative of communities affected by armed militancy from 2007 to 2008.

The sensitive question, whose question wording appears in Section 2.2, asked respondents about direct social connections to armed groups. The forced response design was used with a six-sided dice with a dice roll of 2, 3, 4, or 5 corresponding to a truthful response ($p = 2/3$), a 6 corresponding to a forced "yes" ($p_1 = 1/6$), and a 1 to a forced "no" ($p_0 = 1/6$).

4.2 Power Analysis

With funds for approximately 2500 respondents, we examined the power of the design to detect a proportion of affirmative responses ranging from 0% to 15%. Using the expression given in Equation (15), we examined the power under a range of other proportions, depicted in Figure 2. For example, the power of the test to detect a proportion of 10%, our prior expectation, is approximately 1. Based on this analysis, we concluded that there would be sufficient power based on our chosen sample size.

4.3 Multivariate Analysis

We begin by estimating the proportion of respondents who answer "yes" to the sensitive question. Based on the observed responses and the design probabilities, we use Equation (4) to calculate the posterior estimate of the proportion of those who had social connections with a militant. We estimate this proportion to be 26% of respondents with a 95% confidence interval of 23% to 29%.

In addition to estimating the proportion of respondents who hold direct social connections with members of armed groups, it is useful to examine which types of civilians are connected to the groups. To do this, we conduct the multivariate regression analysis described in Section 3.1. In particular, we predict whether respondents hold social connections with armed groups as a function of the assets owned by the respondent (an index of nine assets including radio, T.V., motorbike, car, mobile phone, refrigerator, goat, chicken, and cow), marital status (1 = mar-

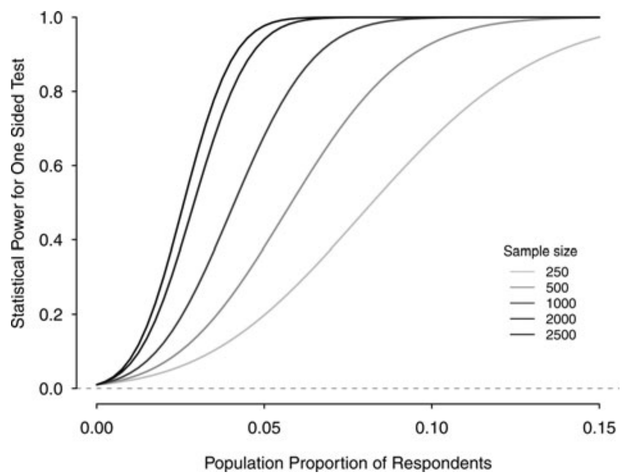

Figure 2. Statistical power analysis for the Nigeria Survey Data. The plot displays the power function for detecting varying proportions answering the sensitive item in the affirmative (the horizontal axis) based on different sample sizes, 250 (lightest line) to 2500 (darkest line).

ried or divorced, 0 = single), age and age squared, education level (from 1 = no schooling to 10 = post-graduate education), and gender (male or female). We use the logistic regression for $f_\beta(x)$ with these covariates as linear predictors.

The estimated coefficients from this model, along with standard errors, are reported in Table 3. The results imply that respondents who have more assets in the household—including radios, televisions, refrigerators—are substantially more likely to be socially connected to armed groups. Women are substantially less likely to be connected, while age holds a curvilinear relationship with militant connections. Marital status and education levels are not strongly associated with social connections to armed groups.

We can also compare the predicted probabilities of a "yes" response to the sensitive item using the fitted model. Based on this logistic regression model and the individual-level posterior predicted probability for the forced response design defined in Equation (11), we estimate that 23% of women shared social connections with members of armed groups, compared to 29% of men with the 95% confidence intervals of 20% to 25% and 27% to 31%, respectively.

Table 3. The estimated logistic regression coefficients from the multivariate regression analysis.

	est.	s.e.
Asset index	0.079	0.041
Married	−0.267	0.255
Age	−3.528	2.642
Age, squared	4.099	2.603
Education level	−0.007	0.046
Female	−0.554	0.162
(Intercept)	−0.340	0.509

NOTE: The model predicts whether the respondent answered the "self-contact" (with militant groups) sensitive item in the affirmative.

Table 4. Multivariate joint model of responses to an outcome regression ("Joined civic group") and a randomized response sensitive item ("Militant connection").

	Joined civic group		Militant connection	
	est.	s.e.	est.	s.e.
Militant connection	0.384	0.160		
Asset index	0.112	0.023	0.087	0.041
Married	0.329	0.133	−0.313	0.240
Age	9.214	1.582	−2.750	2.477
Age, squared	−10.059	1.672	3.417	2.548
Education level	0.086	0.025	−0.014	0.044
Female	−0.025	0.088	−0.562	0.163
(Intercept)	−2.805	0.311	−0.487	0.469

NOTE: Estimated coefficients from logistic regressions with standard errors are reported in each case.

4.4 Using Randomized Response as a Predictor

Finally, we examine whether people with social connections to armed groups are more or less likely to join civic groups in their communities, such as youth groups, women's groups, or community development committees. We accomplish this by using the methodology proposed in Section 3.3. We jointly model the probability of answering "yes" to the sensitive item and the probability of joining a civic group, both using logistic regression with the same set of predictors. The outcome model includes the militant contact as the additional key predictor.

The estimated coefficients from this multivariate joint model are presented, along with standard errors, in Table 4. The first two columns present the results from the outcome regression model while the last two columns show those from the submodel predicting contact with militant groups. The results suggest that respondents who are socially connected to armed groups are more likely to later join civic groups. In particular, 57% of those who are connected to armed groups are predicted to join a civic group (a 95% confidence interval from 52% to 63%), compared to 49% of those who are not connected to the groups (a 95% confidence interval from 46% to 51%). The difference is estimated to be 9% with a 95% confidence interval of 2.7% to 15%.

5. MODIFIED DESIGNS WITH UNKNOWN PROBABILITY

All of the four basic designs explained in Sections 2 and 3 assume that the randomization distribution is known and respondents comply with the instructions. However, such assumptions may be violated in practice. For example, when surveys are conducted via phone rather than in person, respondents may not flip a coin as instructed especially if a coin is not readily available. As a second consideration, the unrelated question design requires researchers to know the response distribution to the unrelated question. But, this information may not be available.

In this section, we introduce the two designs, one new and the other existing, that allow these probabilities to be estimated (see also van den Hout and Klugkist 2009, for a model-based, rather than design-based, strategy). In particular, we modify the forced response and unrelated question designs. The disadvantage of these modified designs, however, is that they require a larger sample size to maintain the same level of statistical power.

5.1 Designs

We consider the forced response and unrelated question designs with unknown probability. We show that both of these modified designs are based on the same identification strategy and hence the identical estimation method is applicable.

5.1.1 Forced Response Design With Noncompliance. Under the standard forced response design, we assume that both the probability of answering the sensitive question p and that of a forced "yes" ("no") p_1 (p_0) are known. However, respondents may not reply "yes" even when forced to do so (Edgell, Himmelfarb, and Duchan 1982). The modified forced response design addresses such noncompliance behavior. The assumption here is that sensitive questions lead to under-reporting though a design similar to the one we propose here can also be used to investigate the instances of over-reporting.

Suppose that we set the probability of a forced "no" to zero. We first randomly split the sample of respondents into two groups. In the first group ($G_i = 1$), respondents are instructed to flip a coin and answer the sensitive question truthfully if they get heads (Pr(heads) $= p$). We assume that this probability p is known and respondents do provide a truthful answer to the sensitive question. If they get tails, then respondents are instructed to answer "yes" but we allow for some noncompliance. That is, the unknown proportion of these respondents may reply "no." Let $1 - q$ denote the probability of such noncompliance (q is the probability of compliance).

In the second group ($G_i = 0$), respondents truthfully answer the sensitive question if they obtain tails whereas they are instructed to reply with "yes" if they get heads. As in the case of the first group, we allow for noncompliance and assume that some of the respondents who are told to say "yes" may answer "no" with the same probability $1 - q$. This assumption holds because the two groups are randomly sampled. Then, the resulting estimating equations are given as follows:

$$\Pr(Y_i = 1 \mid G_i = 1) = p \Pr(Z_i = 1) + (1 - p)q \quad (17)$$
$$\Pr(Y_i = 1 \mid G_i = 0) = (1 - p) \Pr(Z_i = 1) + pq. \quad (18)$$

Solving for $\Pr(Z_i = 1)$ yields

$$\Pr(Z_i = 1) = \frac{1}{2p - 1} \{p \Pr(Y_i = 1 \mid G_i = 1) - (1 - p) \times \Pr(Y_i = 1 \mid G_i = 0)\}. \quad (19)$$

Note that it is possible to use different randomization probabilities for two groups so long as they are known to the researchers.

It is important to note that while this modified design addresses a particular type of noncompliance, in practice respondents may exhibit other types of noncompliance behavior. This design is a special case of the design proposed by Clark and Desharnais (1998) who also considered assigning different probabilities for randomization device across groups. Below, we contribute to this literature by developing a multivariate regression technique and power analysis for this modified design.

5.1.2 Unrelated Question Design With Unknown Probability. Under the standard unrelated question design, we assume that the response probability to the unrelated question is known. However, such information may be unreliable or even nonexistent. The motivation of the modified unrelated question design we consider here is to assume that this response probability is unknown. Specifically, we first randomly split the respondents into two groups. In the first group ($G_i = 1$), the respondents are instructed to flip a coin and answer the sensitive question if they get heads (Pr(heads) $= p$). We assume that this probability is known. The respondents answer the unrelated question if the outcome of the coin flip is tails.

In the second group ($G_i = 0$), we reverse the instructions. That is, assuming that the coin flip has the same randomization distribution, if the respondents get heads (tails), they are told to answer the sensitive (unrelated) question. This modified design has been used in the literature. The applications include studies of drug use (Goodstadt and Gruson 1975), shoplifting (Reinmuth and Geurts 1975), voting (Locander, Sudman, and Bradburn 1976), and compliance with medication (Volicer and Volicer 1982).

If we let q represent the probability of answering "yes" to the unrelated question, the estimating equations are identical to those of the modified forced response design discussed above (i.e., Equations (17) and (18)). Thus, the probability of affirmatively answering the sensitive question is also the same and is given in Equation (19).

5.2 Multivariate Regression Analysis

We first consider an approach that has an important advantage of avoiding the specification of regression function for the unknown probability q. We base our inference on the following moment condition derived using the equality given in Equation (19):

$$f_\beta(X_i) = \frac{1}{2p - 1}\{p \Pr(Y_i = 1 \mid X_i, G_i = 1) - (1 - p) \times \Pr(Y_i = 1 \mid X_i, G_i = 0)\}. \quad (20)$$

In this framework, the regression function of interest, $f_\beta(X_i)$, is obtained as a result of modeling the observed response, $\Pr(Y_i = 1 \mid X_i, G_i)$. An obvious disadvantage of this approach is that one cannot directly specify the latent response to the sensitive question. Rather, we obtain the model specification as a byproduct of the model for the observed response. For example, even if we wish to use the logistic regression for $f_\beta(X_i)$, it is not straightforward to obtain a model for the observed response, which satisfies Equation (20). One exception is the linear probability model, $f_\beta(X_i) = \beta^\top X_i$. In this case, we can also use linear probability models for $\Pr(Y_i = 1 \mid X_i, G_i)$ while satisfying Equation (20). Despite this issue, the proposed approach avoids modeling the unknown probability q and hence rests on the less stringent assumptions.

The second approach we consider follows the modeling and estimation strategy outlined in Sections 3.1 and 3.2 for the standard designs. Unlike the previous one, this approach requires researchers to specify a regression model for the unknown probability $q = q_\gamma(X_i)$ while allowing them to directly model the latent response to the sensitive question. The inference is based on the following likelihood function:

$$\mathcal{L}(\beta, \gamma \mid \{X_i, Y_i, G_i\}_{i=1}^n) = \prod_{i=1}^n \{e(G_i) f_\beta(X_i) + e(1-G_i) \times q_\gamma(X_i)\}^{Y_i} [1 - \{e(G_i) f_\beta(X_i) + e(1-G_i) q_\gamma(X_i)\}]^{1-Y_i},$$

where $e(G_i) = G_i p + (1 - G_i)(1 - p)$.

To maximize this likelihood function, the EM algorithm is useful. Consider the complete-data likelihood function of the following form:

$$\mathcal{L}_{\text{com}}(\beta, \gamma \mid \{X_i, Y_i, G_i, S_i\}_{i=1}^n) = \prod_{i=1}^n \left[f_\beta(X_i)^{Y_i} \times \{1 - f_\beta(X_i)\}^{1-Y_i} \right]^{S_i} \left\{ q_\gamma(X_i)^{Y_i} (1 - q_\gamma(X_i))^{1-Y_i} \right\}^{1-S_i}, \tag{21}$$

where S_i indicates whether respondent i answers the sensitive question ($S_i = 1$) or not ($S_i = 0$) under each modified design. Now, we can derive the E-step as follows:

$$\mathbb{E}(S_i \mid X_i = x, Y_i = y, G_i = g) = \frac{e(g) f_\beta(x)^y (1 - f_\beta(x))^{1-y}}{e(g) f_\beta(x)^y (1 - f_\beta(x))^{1-y} + e(1-g) q_\gamma(x)^y (1 - q_\gamma(x))^{1-y}}.$$

Then, the M-step can be implemented by maximizing the following objective function with respect to β and γ:

$$\sum_{i=1}^n w_S(X_i, Y_i, G_i) \left\{ Y_i \log f_\beta(X_i) + (1 - Y_i) \log(1 - f_\beta(X_i)) \right\} + \{1 - w_S(X_i, Y_i, G_i)\} \left\{ Y_i \log q_\gamma(X_i) + (1 - Y_i) \log(1 - q_\gamma(X_i)) \right\},$$

where $w_S(X_i, Y_i, G_i) = \mathbb{E}(S_i \mid X_i, Y_i, G_i)$. This optimization can be easily done by separately fitting two weighted regressions, $f_\beta(X_i)$ and $q_\gamma(X_i)$.

Although we do not provide details, we can also apply the modeling strategy similar to the one described in Section 3.3 to use the latent response to the sensitive question as an explanatory variable in outcome regression models.

5.3 Power Analysis

To conduct power analysis for these modified designs, we first derive the analytical expression for the standard error. Akin to Section 3.4, we have, without loss of generality, $f = f_\beta(1) = \exp(\beta)/\{1 + \exp(\beta)\}$, which is the probability of possessing the sensitive trait, and $q = q_\gamma(1) = \exp(\gamma)/\{1 + \exp(\gamma)\}$, which is the probability of possessing the unrelated trait. Given the likelihood function in Equation (21), The Fisher information with respect to f is given by

$$\mathcal{I}(p, q, r, f) = \frac{rp^2}{\{pf + (1-p)q\}\{1 - pf - (1-p)q\}} + \frac{(1-r)(1-p)^2}{\{(1-p)f + pq\}\{1 - (1-p)f - pq\}},$$

where $r = \Pr(G_i = 1)$ and each term essentially follows the Fisher information in Equation (13) with $c = p$ and $d = (1-p)q$ for $G_i = 1$, and $c = (1-p)$ and $d = pq$ for $G_i = 0$. Thus, the standard error of $\hat{f}$ under the modified designs is $\sigma(p, q, r, f, n) = 1/\sqrt{n\, \mathcal{I}(p, q, r, f)}$. Using this standard error expression, the power functions under one- and two-sided hypothesis tests are identical to those under the basic designs, given in Equations (15) and (16), respectively.

5.4 Possible Extensions

The idea of randomly splitting the sample into two groups can be applied in variety of ways to make the standard designs robust to a certain deviation from the assumptions. We illustrate this by introducing another modified forced response design. Under the standard forced response design, the probability of answering the sensitive question p as well as the probability of forced "yes" and "no" responses, p_0 and p_1, respectively, are assumed to be known. In Section 5.1, we address possible noncompliance to forced response by allowing some respondents to answer "no" when they are supposed to say "yes." Alternatively, we could assume that such noncompliance does not exist but the coin flip probability p is unknown. For example, if the survey is conducted over phone, respondents may not have access to a coin and hence p may not be equal to the assumed probability of a coin flip. Under this alternative assumption, we have $q = 1$ in Equations (17) and (18). Thus, solving for $\Pr(Z_i = 1)$ gives $\Pr(Z_i = 1) = \Pr(Y_i = 1 \mid G_i = 0) + \Pr(Y_i = 1 \mid G_i = 1) - 1$. Given this identification strategy, we can follow the modeling strategies described in Section 5.2 and conduct a multivariate regression analysis. We can also derive the power analysis as done in Section 5.3.

6. CONCLUDING REMARKS

Since its inception a half century ago, the literature on the randomized response technique has focused primarily on theoretical improvements, extensions, and variations in procedures (Chaudhuri 2011). Scholars have assessed the method's efficiency within various designs (e.g., Moors 1971; Dowling and Shachtman 1975; Pollock and Bek 1976) and compared them to estimates from direct questioning (e.g., Lensvelt-Mulders, Hox, and Van Der Heijden 2005a; Krumpal 2012; Gingerich et al. 2014; Rosenfeld, Imai, and Shapiro 2015). The design originally outlined by Warner (1965) has been extended to incorporate multiple sensitive traits (Abul-Ela, Greenberg, and Horvitz 1967; Christofides 2005), multiple sensitive questions (Raghavarao and Federer 1979; Tamhane 1981), responses on a Likert scale (Himmelfarb 2008; De Jong, Pieters, and Fox 2010), and quantitative answers (Eichhorn and Hayre 1983; Fox and Tracy 1984). Recent work has explored flexibility in sampling procedures (Chaudhuri 2001; Chaudhuri and Saha 2005; Chaudhuri 2011).

While immense methodological progress has been made, the lack of substantive applications suggests the need for a practical guide regarding the basic aspects of the randomized response methodology. In this article, we describe commonly used designs with examples, show how to conduct multivariate regression analyses under each design with the sensitive item as outcome or predictor, develop power analyses, and propose new designs that address certain deviations from standard design protocols. Finally, we offer open-source software to facilitate the use of these methods. Taken together, we hope this article enables the effective use of the randomized response technique across disciplines as well as further methodological development.

APPENDIX: ADDITIONAL POWER ANALYSES

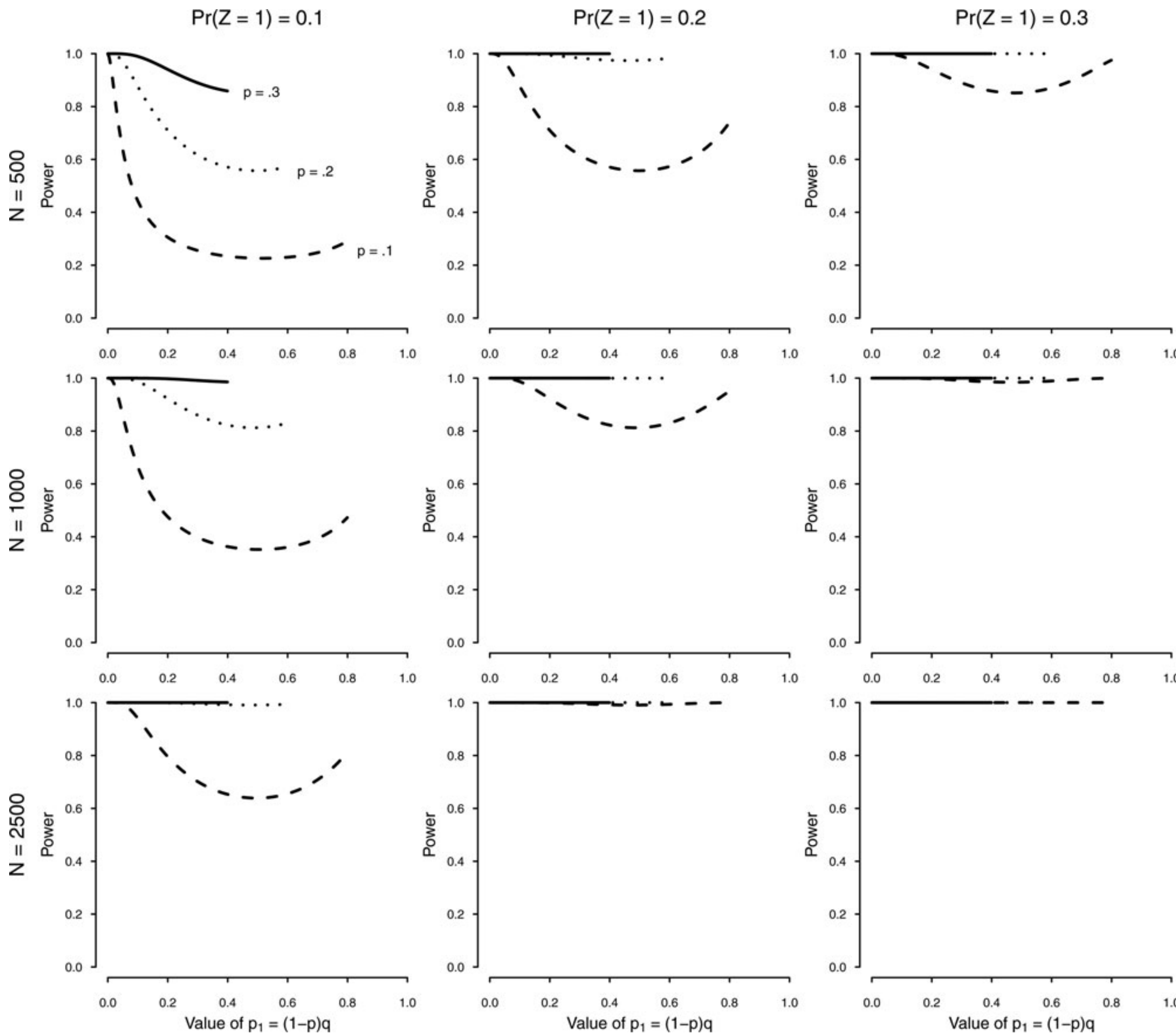

Figure A.1. Comparison of power for the forced response and unrelated question designs. For three typical values of the probability of a truthful response to the sensitive item (p), power is displayed for across values of the probability of a forced "yes" response (p_1) for the forced response design and, equivalently, the known proportion of "yes" responses to the unknown question multiplied by the probability of answering the unknown question ($(1 - p)q$).

[Received September 2014. Revised April 2015.]

REFERENCES

Abernathy, J. R., Greenberg, B. G., and Horvitz, D. G. (1970), "Estimates of Induced Abortion in Urban North Carolina," *Demography*, 7, 19–29. [1304,1307]

Abul-Ela, A.-L. A., Greenberg, G. G., and Horvitz, D. G. (1967), "A Multi-Proportions Randomized Response Model," *Journal of the American Statistical Association*, 62, 990–1008. [1316]

Blair, G. (2014), "Who Informs on their Neighbors in Civil War? Evidence from Nigeria." Unpublished manuscript. [1306]

Blair, G., Imai, K., and Zhou, Y.-Y. (2015a), "Replication Data for: Design and Analysis of the Randomized Response Technique." Available at *http://dx.doi.org/10.7910/DVN/AIO5BR*. The Dataverse Network. [1304]

Blair, G., Zhou, Y.-Y., and Imai, K. (2015b), "rr: Statistical Methods for the Randomized Response," Comprehensive R Archive Network (CRAN). Available at *http://CRAN.R-project.org/package=rr*. [1304,1305]

Boruch, R. F. (1971), "Assuring Confidentiality of Responses in Social Research: A Note on Strategies," *American Sociologist*, 6, 308–311. [1304,1306]

Bourke, P. D., and Moran, M. A. (1988), "Estimating Proportions From Randomized Response Data Using the em Algorithm," *Journal of the American Statistical Association*, 83, 964–968. [1310]

Chaloupka, M. Y. (1985), "Application of the Randomized Response Technique to Marine Park Management: An Assessment of Permit Compliance," *Environmental Management*, 9, 393–398. [1304,1306]

Chaudhuri, A. (2001), "Using Randomized Response From a Complex Survey to Estimate a Sensitive Proportion in a Dichotomous Finite Population," *Journal of Statistical Planning and Inference*, 94, 37–42. [1316]

——— (2011), *Randomized Response and Indirect Questioning Techniques in Surveys*, Boca Raton, FL: Chapman & Hall/CRC Press. [1305,1316]

Chaudhuri, A., and Saha, A. (2005), "Optional Versus Compulsory Randomized Response Techniques in Complex Surveys," *Journal of Statistical Planning and Inference*, 135, 516–527. [1316]

Chi, I.-C., Chow, L., and Rider, R. V. (1972), "The Randomized Response Technique as Used in the Taiwan Outcome of Pregnancy Study," *Studies in Family Planning*, 3, 265–269. [1304,1307,1309]

Christofides, T. C. (2005), "Randomized Response Technique for Two Sensitive Characteristics at the Same Time," *Metrika*, 62, 53–63. [1316]

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. [1315]

Coutts, E., and 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, 169–193. [1305]

Coutts, E., Jann, B., Krumpal, I., and Näher, A.-F. (2011), "Plagiarism in Student Papers: Prevalence Estimates Using Special Techniques for Sensitive Questions," *Journal of Economics and Statistics (Jahrbücher für Nationalökonomie und Statistik)*, 231, 749–760. [1305,1308]

Cruyff, M. J., van den Hout, A., van der Heijden, P. G., and Böckenholt, U. (2007), "Log-Linear Randomized-Response Models Taking Self-Protective Response Behavior Into Account," *Sociological Methods & Research*, 36, 266–282. [1304,1307]

De Jong, M. G., Pieters, R., and Fox, J.-P. (2010), "Reducing Social Desirability Bias Through Item Randomized Response: An Application to Measure Underreported Desires," *Journal of Marketing Research*, 47, 14–27. [1304,1306,1316]

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977), "Maximum Likelihood From Incomplete Data via the EM Algorithm" (with discussion), *Journal of the Royal Statistical Society,* Series B, 39, 1–37. [1309]

Donovan, J. J., Dwight, S. A., and Hurtz, G. M. (2003), "An Assessment of the Prevalence, Severity, and Verifiability of Entry-Level Applicant Faking Using the Randomized Response Technique," *Human Performance*, 16, 81–106. [1304,1306]

Dowling, T. A., and Shachtman, R. H. (1975), "On the Relative Efficiency of Randomized Response Models," *Journal of the American Statistical Association*, 70, 84–87. [1316]

Edgell, S. E., Himmelfarb, S., and Duchan, K. L. (1982), "Validity of Forced Responses in a Randomized Response Model," *Sociological Methods & Research*, 11, 89–100. [1304,1307,1315]

Eichhorn, B. H., and Hayre, L. S. (1983), "Scrambled Randomized Response Methods for Obtaining Sensitive Quantitative Data," *Journal of Statistical Planning and Inference*, 7, 307–316. [1316]

Elffers, H., Van Der Heijden, P., and Hezemans, M. (2003), "Explaining Regulatory Non-Compliance: A Survey Study of Rule Transgression for Two Dutch Instrumental Laws, Applying the Randomized Response Method," *Journal of Quantitative Criminology*, 19, 409–439. [1304,1306]

Fidler, D. S., and Kleinknecht, R. E. (1977), "Randomized Response Versus Direct Questioning: Two Data-Collection Methods for Sensitive Information," *Psychological Bulletin*, 84, 1045–1049. [1304,1306]

Fox, J. A., and Tracy, P. E. (1984), "Measuring Associations With Randomized Response," *Social Science Research*, 13, 188–197. [1316]

——— (1986), *Randomized Response: A Method for Sensitive Surveys*, Beverly Hills, CA: Sage. [1304,1306]

Gingerich, D. W. (2010), "Understanding Off-The-Books Politics: Conducting Inference on the Determinants of Sensitive Behavior With Randomized Response Surveys," *Political Analysis*, 18, 349–380. [1304,1305,1313]

Gingerich, D. W., Corbacho, A., Oliveros, V., and Ruiz-Vega, M. (2014), "When to Protect? Using the Crosswise Model to Integrate Protected and Direct Responses in Surveys of Sensitive Behavior." [1308,1316]

Goodstadt, M. S., and Gruson, V. (1975), "The Randomized Response Technique: A Test on Drug Use," *Journal of the American Statistical Association*, 70, 814–818. [1304,1315]

Greenberg, B. G., Abul-Ela, A.-L. A., and Horvitz, D. G. (1969), "The Unrelated Question Randomized Response Model: Theoretical Framework," *Journal of the American Statistical Association*, 64, 520–539. [1304,1307]

Greenberg, B. G., Kuebler Jr, R. R., Abernathy, J. R., and Horvitz, D. G. (1971), "Application of the Randomized Response Technique in Obtaining Quantitative Data," *Journal of the American Statistical Association*, 66, 243–250. [1307]

Heck, D. W., and Moshagen, M. (2014), "RRreg: Correlation and Regression Analyses for Randomized Response Data," Comprehensive R Archive Network (CRAN). Available at *http://cran.r-project.org/package=RRreg*. [1305]

Himmelfarb, S. (2008), "The Multi-Item Randomized Response Technique," *Sociological Methods & Research*, 36, 495–514. [1304,1306,1316]

Höglinger, M., Jann, B., and Diekmann, A. (2014), "Sensitive Questions in Online Surveys: An Experimental Evaluation of the Randomized Response Technique and the Crosswise Model," Technical Report, Department of Social Sciences, University of Bern. [1305,1308]

Holbrook, A. L., and Krosnik, J. A. (2010), "Measuring Voter Turnout by Using the Randomized Response Technique: Evidence Calling Into Question the Method's Validity," *Public Opinion Quarterly*, 74, 328–343. [1305]

Imai, K., Park, B., and Greene, K. F. (2015), "Using the Predicted Responses From List Experiments as Explanatory Variables in Regression Models," *Political Analysis*, 23, 180–196. [1310]

Jann, B. (2011), "RRLOGIT: Stata Module to Estimate Logistic Regression for Randomized Response Data," available at *https://ideas.repec.org/c/boc/bocode/s456203.html*. [1305]

Jann, B., Jerke, J., and Krumpal, I. (2011), "Asking Sensitive Questions Using the Crosswise Model: An Experimental Survey Measuring Plagiarism," *Public Opinion Quarterly*, doi: 10.1093/poq/nfr036. [1308]

Kirchner, A. (2015), "Validating Sensitive Questions: A Comparison of Survey and Register Data," *Journal of Official Statistics*, 31, 31–59. [1305]

Korndörfer, M., Krumpal, I., and Schmukle, S. C. (2014), "Measuring and Explaining Tax Evasion: Improving Self-Reports Using the Crosswise Model," *Journal of Economic Psychology*, 45, 18–32. [1308]

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, 1387–1403. [1304,1306,1316]

Kuk, A. Y. (1990), "Asking Sensitive Questions Indirectly," *Biometrika*, 77, 436–438. [1304,1307]

Lakshmi, D. V., and Raghavarao, D. (1992), "A Test for Detecting Untruthful Answering in Randomized Response Procedures," *Journal of Statistical Planning and Inference*, 31, 387–390. [1311]

Lamb, C. W., and Stem, D. E. (1978), "An Empirical Validation of the Randomized Response Technique," *Journal of Marketing Research (JMR)*, 15, 616–621. [1304,1308]

Lara, D., García, S. G., Ellertson, C., Camlin, C., and Suárez, J. (2006), "The Measure of Induced Abortion Levels in Mexico Using Random Response Technique," *Sociological Methods & Research*, 35, 279–301. [1304,1307]

Lara, D., Strickler, J., Olavarrieta, C. D., and Ellertson, C. (2004), "Measuring Induced Abortion in Mexico: A Comparison of Four Methodologies," *Sociological Methods & Research*, 32, 529–558. [1304,1307]

Lederer, W., and Küchenhoff, H. (2013), "simex: Simex- and Mcsimex-Algorithm for Measurement Error Models," Comprehensive R Archive Network (CRAN). Available at *http://cran.r-project.org/package=simex*. [1305]

Lensvelt-Mulders, G. J., Hox, J. J., and Van Der Heijden, P. G. (2005a), "How to Improve the Efficiency of Randomised Response Designs," *Quality and Quantity*, 39, 253–265. [1304,1306,1316]

Lensvelt-Mulders, G. J., Hox, J. J., Van der Heijden, P. G., and Maas, C. J. (2005b), "Meta-Analysis of Randomized Response Research Thirty-Five Years of Validation," *Sociological Methods & Research*, 33, 319–348. [1304,1306,1311]

Locander, W., Sudman, S., and Bradburn, N. (1976), "An Investigation of Interview Method, Threat and Response Distortion," *Journal of the American Statistical Association*, 71, 269–275. [1304,1315]

Madigan, F. C., Abernathy, J. R., Herrin, A. N., and Tan, C. (1976), "Purposive Concealment of Death in Household Surveys in Misamis Oriental Province," *Population Studies*, 30, 295–303. [1304,1308]

Mangat, N. S. (1994), "An Improved Randomized Response Strategy," *Journal of the Royal Statistical Society,* Series B, 56, 93–95. [1304,1306]

Mangat, N. S., and Singh, R. (1990), "An Alternative Randomized Response Procedure," *Biometrika*, 77, 439–442. [1304,1306]

Moors, J. (1971), "Optimization of the Unrelated Question Randomized Response Model," *Journal of the American Statistical Association*, 66, 627–629. [1307,1311,1316]

O'Hagan, A. (1987), "Bayes Linear Estimators for Randomized Response Models," *Journal of the American Statistical Association*, 82, 580–585. [1308]

Pollock, K., and Bek, Y. (1976), "A Comparison of Three Randomized Response Models for Quantitative Data," *Journal of the American Statistical Association*, 71, 884–886. [1311,1316]

Raghavarao, D., and Federer, W. T. (1979), "Block Total Response as an Alternative to the Randomized Response Method in Surveys," *Journal of the Royal Statistical Society,* Series B, 41, 40–45. [1316]

Reinmuth, J. E., and Geurts, M. D. (1975), "The Collection of Sensitive Information Using a Two-Stage, Randomized Response Model," *Journal of Marketing Research (JMR)*, 12, 402–407. [1304,1315]

Rosenfeld, B., Imai, K., and Shapiro, J. (2015), "An Empirical Validation Study of Popular Survey Methodologies for Sensitive Questions," *American Journal of Political Science.* doi: 10.1111/ajps.12205. [1305,1306,1316]

Scheers, N., and Dayton, C. M. (1988), "Covariate Randomized Response Models," *Journal of the American Statistical Association*, 83, 969–974. [1308,1311]

St John, F. A., Keane, A. M., Edwards-Jones, G., Jones, L., Yarnell, R. W., and Jones, J. P. (2012), "Identifying Indicators of Illegal Behaviour: Carnivore Killing in Human-Managed Landscapes," *Proceedings of the Royal Society B: Biological Sciences*, 279, 804–812. [1304,1306]

Stubbe, J. H., Chorus, A. M., Frank, L. E., Hon, O., and Heijden, P. G. (2014), "Prevalence of Use of Performance Enhancing Drugs by Fitness Centre Members," *Drug Testing and Analysis*, 6, 434–438. [1305,1306]

Tamhane, A. C. (1981), "Randomized Response Techniques for Multiple Sensitive Attributes," *Journal of the American Statistical Association*, 76, 916–923. [1316]

Tan, M. T., Tian, G.-L., and Tang, M.-L. (2009), "Sample Surveys With Sensitive Questions: A Nonrandomized Response Approach," *The American Statistician*, 63, 9–16. [1308]

Tezcan, S., and Omran, A. R. (1981), "Prevalence and Reporting of Induced Abortion in Turkey: Two Survey Techniques," *Studies in Family Planning*, 12, 262–271. [1304,1307]

Tracy, P. E., and Fox, J. A. (1981), "The Validity of Randomized Response for Sensitive Measurements," *American Sociological Review*, 46, 187–200. [1304,1307]

Umesh, U., and Peterson, R. A. (1991), "A Critical Evaluation of the Randomized Response Method: Applications, Validation, and Research Agenda," *Sociological Methods & Research*, 20, 104–138. [1304,1311]

van den Hout, A., and Klugkist, I. (2009), "Accounting for Non-Compliance in the Analysis of Randomized Response Data," *Australian & New Zealand Journal of Statistics*, 51, 353–372. [1314]

van den Hout, A., van der Heijden, P. G., and Gilchrist, R. (2007), "The Logistic Regression Model With Response Variables Subject to Randomized Response," *Computational Statistics & Data Analysis*, 51, 6060–6069. [1308,1309]

van der Heijden, P. G., and van Gils, G. (1996), "Some Logistic Regression Models for Randomized Response Data," in *Proceedings of the 11th International Workshop on Statistical Modeling*, pp. 15–19. [1304,1306,1307,1308]

van der Heijden, P. G., van Gils, G., Bouts, J., and Hox, J. J. (2000), "A Comparison of Randomized Response, Computer-Assisted Self-Interview, and Face-to-Face Direct Questioning Eliciting Sensitive Information in the Context of Welfare and Unemployment Benefit," *Sociological Methods & Research*, 28, 505–537. [1304,1307]

Volicer, B. J., and Volicer, L. (1982), "Randomized Response Technique for Estimating Alcohol Use and Noncompliance in Hypertensives," *Journal of Studies on Alcohol and Drugs*, 43, 739–750. [1304,1315]

Warner, S. L. (1965), "Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias," *Journal of the American Statistical Association*, 60, 63–69. [1304,1305,1308,1316]

Wimbush, J. C., and Dalton, D. R. (1997), "Base Rate for Employee Theft: Convergence of Multiple Methods," *Journal of Applied Psychology*, 82, 756–763. [1304,1306]

Winkler, R. L., and Franklin, L. A. (1979), "Warner's Randomized Response Model: A Bayesian Approach," *Journal of the American Statistical Association*, 74, 207–214. [1308]

Wolter, F., and Preisendörfer, P. (2013), "Asking Sensitive Questions: An Evaluation of the Randomized Response Technique Versus Direct Questioning Using Individual Vatextlidation Data," *Sociological Methods & Research*, 42, 321–353. [1305]

Yu, J.-W., Tian, G.-L., and Tang, M.-L. (2008), "Two New Models for Survey Sampling With Sensitive Characteristic: Design and Analysis," *Metrika*, 67, 251–263. [1308]