

Honest Answers to Embarrassing Questions: Detecting Cheating in the Randomized Response Model

Stephen J. Clark
Vassar College

Robert A. Desharnais
California State University, Los Angeles

Surveys and questionnaires are frequently used by psychologists, social scientists, and epidemiologists to collect data about behavior, attitudes, emotions, and so on. However, when asked about sensitive topics such as their sexual behavior or illegal activity, some respondents lie or refuse to answer. The randomized response method was developed to reduce these evasive answer biases by guaranteeing subject privacy. However, the method has been criticized as being susceptible to cheaters, that is, respondents who do not answer as directed by the randomizing device. Here the authors show that by splitting the sample into 2 groups and assigning each group a different randomization probability, it is possible to detect whether significant cheating is occurring and to estimate its extent while simultaneously protecting the identity of cheaters and those who may have engaged in sensitive behaviors.

Surveys and questionnaires are standard methods for collecting data about behavior, attitudes, emotions, and so forth. The basic assumption of any interview or survey technique is that the respondents are providing honest information. The validity of this assumption is questionable, however, when researchers ask questions that most would be reluctant to answer publicly. Examples of such questions are those that reveal whether the respondent has engaged in an illegal activity, or an activity that is stigmatized by society, or the query may pertain to a behavior of which the respondent is ashamed or about which the respondent feels is simply too personal to confide in someone else. Faced with such a question, some individuals in a sample will refuse to answer or will lie. Either

type of evasion introduces a bias into survey data. Thus, there are severe methodological obstacles to the use of surveys in studies in which a sensitive behavior is directly related to the phenomenon of interest.

Researchers have tried to reduce evasive-answer bias by assuring respondents that their answers will be anonymous. However, many people are still reluctant to answer, fearing that at least the people conducting the survey will know their responses or that the assurance of anonymity is not genuine, fears that are sometimes well founded (Dawes & Smith, 1985, p. 550). Even when researchers are sincere in their offers of anonymity, they may sometimes be unable to keep their promises in the face of court action (Adler, 1989, as cited in Scheers, 1992). As Scheers has suggested, increasing public awareness of the sharing of information between computerized databases may result in an increase in evasive-answer bias. Indeed, one form of evasion, not responding to the survey, seems to have increased in recent years (Goyder, 1989; Groves, 1989).

To overcome this dilemma, Warner (1965, 1971) developed the randomized response model (RRM). It is based on the premise that cooperation by respondents should improve if their answers would not reveal any information about themselves. In other words, the RRM aims to reduce evasive-answer bias by guaranteeing privacy.

How the RRM guarantees privacy is best understood through an example. For this illustration, we use the form of the RRM used by Dawes and Moore

Stephen J. Clark, Department of Psychology, Biopsychology Program, Cognitive Science Program, Vassar College; Robert A. Desharnais, Department of Biology and Microbiology, California State University, Los Angeles.

We thank Joel E. Cohen and Richard Lowry for helpful comments and encouragement. Thanks to Anne Constantino, Randolph Cornelius, and Jeffrey Cynx for reading and commenting on the manuscript. This work was partially supported by a Creative Leave Award from California State University, Los Angeles.

Correspondence concerning this article should be addressed to Stephen J. Clark, Box 363, Vassar College, Poughkeepsie, New York 12604. Electronic mail may be sent to clark@vassar.edu.

(1979). If a team of researchers was trying to model the spread of the HIV virus among members of the armed forces, the researchers would need to know many demographic variables, including some that would be highly sensitive given the current realities of military life: for example, the proportion of soldiers engaging in homosexual behavior.

Suspecting that many participants might refuse to answer or might lie even with guarantees of confidentiality, the researchers might elect to use an RRM to gather their information. For the purposes of this example, imagine that each participant is given a list of questions that requires him or her to answer dichotomously (e.g., *yes* or *no*, *agree* or *disagree*, etc.). An example of one such question (given to male respondents) might be "Have you ever engaged in unprotected anal intercourse with a member of the same sex?" Suppose also that all the questions would be framed so that answering "yes" is admitting to engaging in a sensitive behavior. In addition to the list of questions, each participant would be given a coin and instructions on how to use that coin in answering the questions. The participants would be told to flip the coin before answering each question and to answer that question according to the outcome of the coin toss and the following rule: "If the coin comes up heads, then answer the question truthfully; if it comes up tails, then ignore the question altogether and just say 'yes' no matter what you would have answered to the question."

Consider the responses. An affirmative response could mean either that the respondent had engaged in the behavior or simply that he had flipped tails. Even if a participant's answer is known, his or her actual behavior could not be deduced from the answer, thus confidentiality would be assured. Nevertheless, Dawes and Moore (1979) showed that an investigator could determine the proportion of the sample that engages in any behavior using the equation $\pi = 2\lambda - 1$, where π is the proportion of the population that would privately admit to having engaged in the behavior and λ is the proportion of affirmative responses given. They derived this equation in the following way. The respondents who engaged in the behavior (π) will have answered "yes" regardless of the outcome of the coin toss, and if fair coins are used, half of the participants who have not engaged in the behavior, $0.5(1 - \pi)$, also will have answered "yes" because the coin came up tails (Table 1). Therefore, $\lambda = \pi + 0.5(1 - \pi)$. Algebraic rearrangement yields the equation arrived at by Dawes and Moore. So in our ex-

Table 1
Types of Respondents and Their Characteristics

Characteristic	Respondent type		
	Honest yes	Honest no	Cheater
Engaged in behavior	Yes	No	Unknown
Proportion in sample	π	β	γ
Response to heads	Yes	No	No
Response to tails	Yes	Yes	No

Note. When there is no cheating, $\beta = 1 - \pi$.

ample if $\lambda = .55$, then $\pi = .1$. Of all the respondents, 50% answered "yes" because the coin came up tails, and 5% did so because the coin came up heads and they answered "yes" truthfully to the sensitive question.

The procedure just described is one of several different forms of the RRM that have been developed to analyze answers to dichotomous questions. Previous literature contains several reviews that provide detailed descriptions of these variants (D. T. Campbell & Joiner, 1973; Chaudhuri & Mukerjee, 1988; Dawes & Smith, 1985; Scheers, 1992). In addition, variants of the RRM have been developed that can be used with quantitative questions that may be phrased, "How many times have you . . .?" (Park & Park, 1987; Pollock & Bek, 1976; Scheers, 1992).

In the decade after their introduction, RRMs generated considerable scientific interest. In recent years, however, the use of RRMs has declined somewhat (Scheers, 1992). It is possible that this is because the procedure is not as straightforward as direct questioning and is perceived as involving additional time and cost for researchers. This is largely a misperception. Most of the time and cost is incurred in the initial study using the RRM. When randomizing devices (e.g., coins, spinners, etc.) have been acquired and data analysis procedures have been developed, their use transfers easily to a variety of questionnaires and surveys.

A more compelling reason for using RRMs is that numerous empirical studies have shown that a higher proportion of respondents acknowledge having engaged in sensitive behaviors when this method is used than when traditional questionnaires with assurances of anonymity are used. This finding has been consistent across a variety of RRM variants and across a wide variety of behaviors including having an abortion (Krotki & Fox, 1974; Schimizu & Bonham, 1978), academic cheating (Scheers & Dayton, 1987, 1988; Stem & Steinhurst, 1984), drug use (Goodstadt & Gruson, 1975), being arrested (Tracy & Fox, 1981),

and a variety of other illegal or private behaviors (Dawes & Moore, 1979; Fiddler & Kleinknecht, 1977; Greenberg, Abul-Ela, Simmons, & Horvitz, 1969; Horvitz, Shah, & Simmons, 1967; Volicer, Cahill, Neuburger & Arntz, 1983; Weissman, Steer, & Lipton, 1986). Finally, nonresponse rates are lower when RRM's are used (Scheers, 1992). The results of all these studies make a compelling case that researchers who use traditional methods to ask participants about sensitive behaviors run the risk of obtaining biased estimates because of evasive answering.

We are hardly alone in championing the increased use of RRM's. Kolata (1987) reported a suggestion by Joel Cohen that the RRM be used to encourage honest reports of the sexual behaviors that spread HIV (Cohen, 1987, noted that this suggestion was made previously by Fiering & Hooper, 1985). In response to this report, A. A. Campbell (1987) presented several reasons why such use of the RRM might be unfeasible; some of these reasons have since been addressed (Golbeck & Molgaard, 1990). One of A. A. Campbell's concerns that inspired this article was the argument that the RRM is not immune to cheating and that respondents especially anxious to avoid being identified with a sensitive behavior will simply answer "no" regardless of the outcome of the randomizing device.

Cheating in this sense is distinct from lying. Lying is answering contrary to fact, but we operationally define cheating as not answering according to the instructions of the RRM. Some cheaters may be liars, but this is not necessarily the case. In the case of an extremely homophobic individual who has never engaged in any homosexual behavior, who is answering the anal sex question from our example, and who has just flipped tails, accordingly to our rule, the person should answer "yes" to the question. It is possible that such an individual might break the rule and answer "no" to avoid even the possibility of anyone thinking he or she has engaged in homosexual behavior. This would be even more likely if the respondent does not understand how the RRM guarantees privacy. Thus, we make no assumption about whether cheaters have actually engaged in the sensitive behavior or not.

In fact, there is evidence that cheating occurs in some circumstances (Locander, Sudman, & Bradburn, 1976; Shimizu & Bonham, 1978; Weisman, Moriarity, & Schafer, 1975). One reason to expect cheating is that the rationale of the RRM is based on statistics. Statistical literacy is low in the general population,

and some participants may not understand how their privacy is protected. Other participants might assume that the randomizing device is not truly random but is being manipulated in some way by the researchers. Whatever its cause, cheating frustrates the very object of using the RRM. This prompted us to develop a modification to the RRM. This modification enables researchers to estimate the proportion of a sample population engaging in a certain behavior and to test whether a significant proportion of the individuals in a sample are violating the assumptions of the RRM by failing to answer as instructed. Furthermore, our method achieves these aims without revealing which members of the sample have or have not engaged in the behavior and without revealing which individuals are not following directions properly. We believe this method could prove valuable for assessing the quality of data obtained by the RRM.

The Model

We consider questions that could be answered with a "yes" or a "no," with the question framed so that a "yes" response is acknowledging having engaged in a sensitive behavior. We also assume the instructions given above: "If heads, answer truthfully; if tails, answer 'yes.'"

We conceptually divide the sample population into three groups. The first two groups consist of those participants who follow the researchers' instructions and, when instructed to answer truthfully, answer either "yes" (honest yes) or "no" (honest no). The third group consists of those participants who ignore the researchers' instructions and answer "no" regardless of the outcome of the randomizing device (cheaters). We make no assumption about whether individuals in the third group would answer "yes" or "no" if they were to answer truthfully. There is a fourth possible type of respondent: one that ignores instructions and always answers "yes." We assume that these respondents are so rare that they can safely be ignored. The proportion of honest-yes respondents in the population is equal to the π of other RRM's. We designate the proportions of the honest-no respondents and cheater respondents in the population as β and γ , respectively (Table 1).

We base our method on splitting the sample group of N individuals into two groups. Individuals in each group are asked the same questions with the same instructions but with different probabilities of getting the two outcomes from the randomizing device. Using

two different probabilities allows the development of a system of equations for estimating π , β , and γ , as explained below. The investigator then tallies the numbers of "yes" and "no" responses in the two groups.

From these tallies, we can make unbiased estimates of the proportions of honest-yes respondents (π), honest-no respondents (β), and cheaters (γ) in the population. Let p_1 and p_2 represent the randomization probabilities for Groups 1 and 2, respectively (i.e., the probability of being instructed to answer "yes") and consider the probability of a "yes" response. All honest-yes respondents will answer "yes" regardless of the outcome of the randomization device, and their frequency in the population is π (Table 1). Additional positive responses will come from honest-no respondents who would have been directed by the randomization device to answer "yes" (Table 1). Their frequency in the population is β , and the probability that they will answer "yes" is p_i for group i , $i = 1, 2$. Therefore, the probability of a "yes" response in group i is $\lambda_i = \pi + p_i\beta$ for $i = 1, 2$. If λ_i is estimated with the observed frequency of "yes" responses, y_i/N_i , where y_i is the number of "yes" responses and N_i is the number of participants in the i th group, $i = 1, 2$, then the result is a system of two equations with two unknowns, π and β . (Because $\pi + \beta + \gamma = 1$, the proportion of cheaters equals $1 - \pi - \beta$.) Solving these equations gave the following unbiased maximum-likelihood estimates of π , β , and γ , which we denoted by $\hat{\pi}$, $\hat{\beta}$, and $\hat{\gamma}$ (when $p_1 \neq p_2$):

$$\hat{\pi} = (p_2y_1/N_1 - p_1y_2/N_2)/(p_2 - p_1), \tag{1}$$

$$\hat{\beta} = (y_2/N_2 - y_1/N_1)/(p_2 - p_1), \text{ and} \tag{2}$$

$$\hat{\gamma} = 1 - (\hat{\pi} + \hat{\beta}). \tag{3}$$

If either $\hat{\pi}$ or $\hat{\beta}$ is negative or exceeds 1.0 then one or two of the maximum-likelihood estimates for π , β , and γ are equal to 0.0. One must compute three sets of maximum-likelihood estimates for the three cases, $\pi = 0$, $\beta = 0$, and $\gamma = 0$ and choose the set of estimates that yields the largest value for the likelihood function. (Details of these computations are available from the authors.) The asymptotic (large-sample) variances (var) and covariances (cov) of these estimates are as follows:

$$\text{var}(\hat{\pi}) = (p_1 - p_2)^{-2} \left(\frac{p_1^2 y_2 n_2}{N_2^3} + \frac{p_2^2 y_1 n_1}{N_1^3} \right),$$

$$\text{cov}(\hat{\pi}, \hat{\beta}) = -(p_1 - p_2)^{-2} \left(\frac{p_1 y_2 n_2}{N_2^3} + \frac{p_2 y_1 n_1}{N_1^3} \right), \text{ and}$$

$$\text{var}(\hat{\beta}) = (p_1 - p_2)^{-2} \left(\frac{y_2 n_2}{N_2^3} + \frac{y_1 n_1}{N_1^3} \right),$$

where n_1 and n_2 are the numbers of "no" responses in the two sample groups.

To determine whether significant cheating was occurring, we developed a likelihood ratio test for the null hypothesis that there are no cheaters ($H_0: \gamma = 0$). The alternative hypothesis is that cheaters were present ($H_1: \gamma > 0$). When the null hypothesis—that there was no cheating—holds, $\pi + \beta = 1$, and β is estimated as the smaller of 1 or

$$\beta^* = \frac{-b + \sqrt{b^2 - 4ac}}{2a},$$

where $a = -(N_1 a_1 a_2 + N_2 a_2 a_1)$, $b = (N_1 a_1 + n_1 a_2 + N_2 a_2 + n_2 a_1)$, $c = -(n_1 + n_2)$, $a_1 = 1 - p_1$, $a_2 = 1 - p_2$, and * denotes the maximum-likelihood estimate of a variable. With $y_i^* = N_i(\pi^* + \beta^* p_i)$ and $n_i^* = N_i - y_i^*$ for $i = 1, 2$, the following test statistic has an asymptotic chi-square distribution with one degree of freedom (Kendall & Stuart, 1979):

$$G^2 = 2 \sum_{i=1}^2 \left[y_i \log \left(\frac{y_i}{y_i^*} \right) + n_i \log \left(\frac{n_i}{n_i^*} \right) \right].$$

An alternative to the G^2 with the same limiting distribution is Pearson's chi-square (also with one degree of freedom):

$$\chi^2 = \sum_{i=1}^2 \left[\frac{(y_i - y_i^*)^2}{y_i^*} + \frac{(n_i - n_i^*)^2}{n_i^*} \right].$$

Either of these two test statistics can be used to detect cheating.

The ability of these tests to detect cheating when it is actually present is measured by the power of the test. The power is defined as the probability of rejecting the null hypothesis, H_0 , of no cheating when cheating actually exists ($\gamma > 0$). For a measure of the power, we fixed the level of significance at .05, set $N_1 = N_2$, and used the noncentral chi-square distribution (Johnson & Kotz, 1970) to compute the power for two cases of the randomization probabilities: when $p_1 = 1/4$, $p_2 = 3/4$ and when $p_1 = 1/3$, $p_2 = 2/3$. Because the power also depends on the unknown parameter, π , we acted conservatively and chose the smallest values of the power for π in the range 0 to $1 - \gamma$. Power

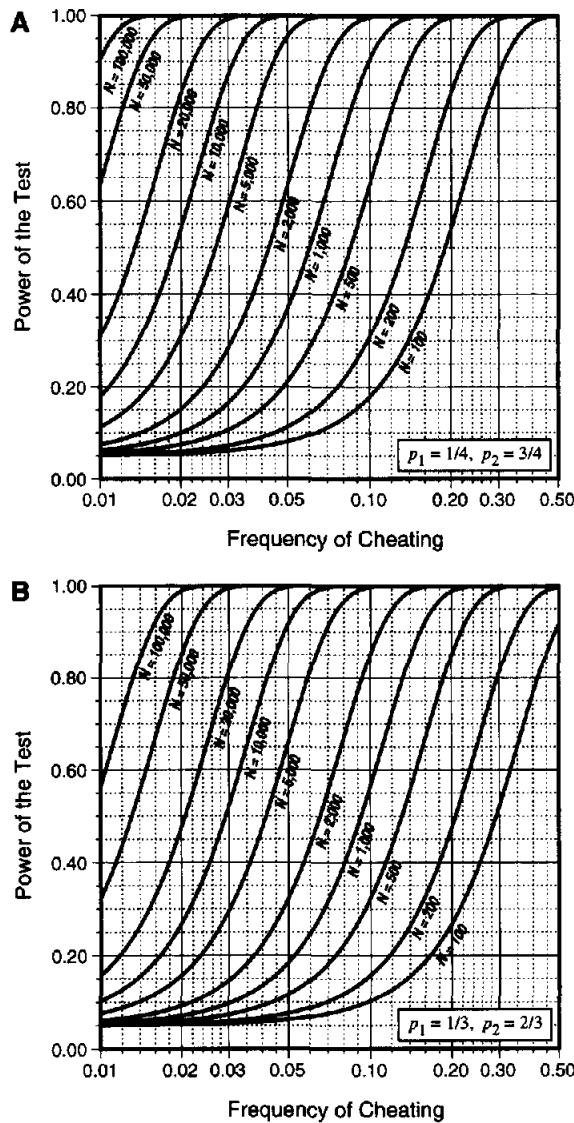


Figure 1. Power curves. The probability of detecting cheating as a function of the true frequency of cheating (γ). Randomization probabilities for Panel A are $p_1 = 1/4$ and $p_2 = 3/4$ and for Panel B are $p_1 = 1/3$ and $p_2 = 2/3$. The level of significance for the null hypothesis of no cheating equals .05. These results assume equal sample sizes in the two groups, $N_1 = N_2 = N/2$. The different curves represent different sample sizes $N = N_1 + N_2$ as indicated in each graph. The power curves are minima for all possible frequencies of honest-yes respondents (π).

curves for a variety of sample sizes, $N = N_1 + N_2$, are presented in Figure 1 as a function of the true frequency of cheating, γ . Increasing the sample size or choosing randomization probabilities p_1 and p_2 that are further apart increases the ability to detect cheating.

An Example

To give a better idea of how our modification to the RRM would actually be used, we refer again to our imaginary team of researchers. If these researchers wish to know the frequency of various homosexual behaviors among soldiers in the military, a primary concern of the researchers is the accuracy of their data. Because they are aware that they are asking about sensitive behaviors, they decide to use an RRM to reduce evasive-answer bias. They are concerned, however, that some respondents may still decide to cheat, so they adopt our method to determine whether significant cheating is occurring. Assume that the researchers have developed their questions and are now deciding on the sample sizes.

The researchers decide that if the level of cheating exceeds 10%, they want to be fairly certain that they detect it. Looking at Figure 1A, they see that when the frequency of cheating is 0.10, a sample size of $N = 1,000$ will result in the detection of cheating over 90% of the time for randomization probabilities of .25 and .75. (For more on the issues relating to selecting a sample size, see Discussion.)

Having chosen a sample size and randomization probabilities, the researchers now choose a randomizing device. A variety of ingenious methods using spinners, containers of colored balls, dice, poker chips, playing cards, and even the phone book has been used for randomizing devices (for a review, see Scheers, 1992). The researchers construct spinners as shown in Figure 2. The researchers give spinners to participants in both groups. The rule given to members of Group 1 ($p_1 = .75$) is "Before answering each question, spin the arrow. If the arrow stops on an area

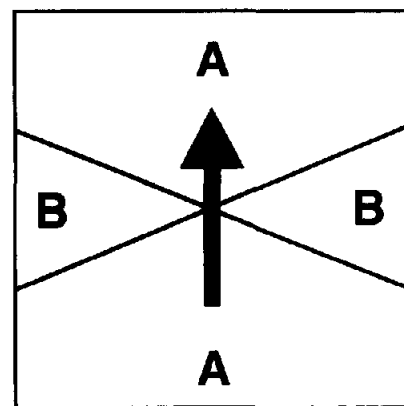


Figure 2. A spinner designed to produce two outcomes, A and B, with $p_A = .75$ and $p_B = .25$.

marked 'B,' then answer the question truthfully; if it stops on an area marked 'A,' then ignore the question altogether and just say 'yes,' no matter what you would have answered to the question." Participants in Group 2 ($p_2 = .25$) are given similar instructions with the actions assigned to areas "A" and "B" reversed.

Table 2 summarizes the results that the hypothetical team of researchers obtain. The data for this model were generated by a stochastic simulation of our method for a pool of 1,000 participants with known parameters ($\pi = .037$, $\beta = .864$, $\gamma = 0.099$). Substituting the values in Table 2 into Equations 1-3 yields the estimates, $\hat{\pi} = .035$, $\hat{\beta} = .876$, and $\hat{\gamma} = 0.089$.

The model suggests that 8.9% of the participants are cheating and are not following the rule when answering questions. The model has done quite well, because the sample actually contains 9.9% cheaters. Of course, the team of researchers has no way of knowing this, and they know that some nonzero estimates of cheating can result from sampling error. To see whether the estimate, $\hat{\gamma} = .089$, could have occurred by chance, the researchers calculate the value of Pearson's chi-square statistic. This calculation is presented in the Appendix and results in a value of 9.01. The researchers consult a chi-square table (e.g., Rohlf & Sokal, 1981) for one degree of freedom and discover that this value is sufficient to reject the null hypothesis that there is no cheating ($H_0: \gamma = 0$) at the 5% level of significance.

Discussion

We have presented a modification to the variant of the RRM developed by Dawes and Moore (1979), which provides a means for detecting and estimating the magnitude of cheating. The modification itself, using two sample groups with different probabilities of getting heads, has been suggested previously (Greenberg et al., 1969; Horvitz et al., 1967). Furthermore, after we developed the modification that we have presented here, we became aware of the work of

Mangat and Singh (1990), who proposed a two-stage RRM with different probabilities at each stage. They showed that when respondents are not completely truthful, their version of the RRM is more efficient than Warner's (1965) original formulation. However, in neither this nor the previously cited studies did the authors propose the application of this modification to the problem of detecting and estimating cheating.

To use our method, researchers must develop the RRM questions, determine how large a sample size is needed, choose a randomizing device, select the probabilities (p_1 and p_2) for the two randomizing devices, randomly assign the respondents to groups, administer the test, and analyze the data. The sample size necessary is affected by many considerations, most of which have nothing to do with the issue of cheating. If one is interested in behaviors—sensitive or not—that might only be practiced by a small proportion of the population, then sample sizes need to be large. This is affected not only by the frequency of the behaviors but on how the question is framed. For example, Kinsey, Pomeroy, and Martin (1948) found that the frequency of homosexual behavior, .37, .10, or .04, depended on whether the respondent was asked about having had (a) a homosexual experience at least once in one's lifetime, (b) only homosexual experiences during at least one 3-year period between ages 16 and 55, or (c) exclusive homosexual behavior postadolescence.

The sample size needed to reliably detect cheating is a function of three factors: the frequency of cheating, the randomization probabilities assigned to the two groups, and the desired power of the test. The latter is something most researchers never consider when they use statistical tests because the information is not readily available. We have provided Figure 1, however, which enables researchers to determine the power of the test for a variety of sample sizes. As Figure 1 shows, extremely large sample sizes are needed to detect very small levels of cheating. In addition, the closer p_1 and p_2 are to .5, the larger the sample size needed to detect cheating. By using extreme randomization probabilities, say .10 and .90, researchers can increase the power of the test for a given level of cheating and therefore use smaller sample sizes, but extreme probabilities undermine the guarantee of anonymity afforded by the RRM. If the probability of getting a tails is .50, then there is a 50% chance that a respondent who answers affirmatively is only following the rule to answer "yes," but as the probability of a tails drops, the more likely it is that a "yes" answer means that the respondent actually en-

Table 2
Responses Generated by Stochastic Simulation

Group	N_i	p_i	y_i	n_i
1	500	.75	346	154
2	500	.25	127	373

Note. Sample of 1,000 participants from a population consisting of 3.7% honest-yes respondents, 86.4% honest-no respondents, and 9.9% cheaters. N_i = number of participants; p_i = probability of being instructed to answer "yes"; y_i and n_i = numbers of "yes" and "no" responses, respectively, in the i th group.

gaged in the behavior. Under these circumstances, it becomes likely that more people will decide to cheat with the result being that the level of cheating will actually become a function of the probabilities of the randomizing devices. We suggest, therefore, for any given level of cheating to be detected, that researchers use the largest sample size feasible in order to keep p_1 and p_2 close to .5.

In many cases, researchers will have no initial idea about what levels of cheating might occur. In these situations, researchers should use Figure 1 to choose a sample size based on the level of cheating they wish to reliably detect. One might be inclined to say, "I want to detect any level of cheating no matter how small," but an examination of Figure 1 shows that this clearly is not feasible. Even with a sample size of 100,000 and randomization probabilities of $p_1 = .25$ and $p_2 = .75$, one would miss detecting a 1% level of cheating, on average, 10% of the time (Figure 1A). The inability of our method to detect extremely low levels of cheating is not as much a limitation as it may seem. First, low levels of cheating mean that the data are trustworthy; the main purpose of our method is to alert researchers to when data should be discarded because high levels of cheating have occurred. Second, we remind readers to consider the alternative: Not using our method means that researchers will have no idea of what levels of cheating are occurring and are thus accepting any and all levels.

Some may see the value of the method we have proposed when cheating is likely, but will they still wonder if it is more costly than the traditional RRM when no cheating exists? Surprisingly, the answer is no. All versions of the RRM require larger sample sizes than traditional surveys to achieve the same level of reliability. However, compared with versions of the RRM that do not split the sample into two groups, our version is more efficient. Splitting the sample in half and using randomization probabilities that are symmetric around .5 produce an estimate π_0 that has a lower variance than the estimate π_1 , which would be obtained from the traditional method with an equal number of randomized responses. If relative efficiency (Kendall & Stuart, 1979) is defined as $RE = \text{var}(\pi_0)/\text{var}(\pi_1)$, then

$$RE = 1 - \frac{(1 - \pi)(p_1 - p_2)^2}{(1 - \pi)(p_1 - p_2)^2 + 2\pi + 2(1 - \pi)^2 p_1 p_2},$$

where π is the true population value. Given that $p_1 \neq p_2$, then $RE < 1$ for all $\pi < 1$. Thus, we would advo-

cate the use of this procedure whenever one feels that the use of any RRM is desirable.

What should researchers do if the null hypothesis, that there is no cheating, is rejected? The best estimate of the magnitude of cheating is given by the equation for $\hat{\gamma}$ (Equation 3). We caution researchers against trying to use the value of $\hat{\gamma}$ as a correction factor for $\hat{\pi}$. One cannot assume that all cheaters actually engaged in the behavior. No such assumption went into our model, and neither this method nor any other method is capable of indicating the true behavior of cheaters. If $\hat{\gamma}$ is significant but small relative to $\hat{\pi}$, we suggest reporting both values for $\hat{\pi}$ and $\hat{\gamma}$. If $\hat{\gamma}$ is large, as was the case in our example, the data have been hopelessly contaminated by cheating and are unusable. As undesirable as the latter outcome would be, it is still preferable to know that sample data are contaminated than to draw (and publish) erroneous conclusions, which is what researchers implicitly agree to do when using direct questioning or the RRM without our modification.

References

- Campbell, A. A. (1987, May 29). *Science*, 236, 1049.
- Campbell, D. T., & Joiner, B. L. (1973). How to get the answer without being sure you've asked the question. *American Statistician*, 27, 229-231.
- Chaudhuri, A., & Mukerjee, R. (1988). *Randomized response: Theory and techniques*. New York: Marcel Dekker.
- Cohen, J. E. (1987, June 19). Sexual behavior and randomized response. *Science*, 236, 1503.
- Dawes, R. M., & Moore, M. (1979). Die Guttman-skalierung orthodoxer und randomisierter reaktionen [Traditional Guttman-scaling and randomized response]. In F. Peterman (Ed.), *Attitude measurement* (pp. 117-132). Göttinger, Germany: Verlag für Psychologie.
- Dawes, R. M., & Smith, T. L. (1985). Attitude and opinion measurement. In G. Lindzey & E. Aronson (Eds.), *Handbook of social psychology* (Vol. 1, 3rd ed., pp. 509-566). New York: Random House.
- Fiddler, D. S., & Kleinknecht, R. E. (1977). Randomized response versus direct questioning: Two data-collection methods for sensitive information. *Psychological Bulletin*, 84, 1045-1049.
- Fiering, M. B., & Hooper, R. (1985). Analysis of disclosure avoidance procedures. *Civil Engineering Systems*, 2, 12-19.
- Golbeck, A. L., & Molgaard, C. A. (1990). Methodological issues concerning the sensitive query in AIDS/alcohol research: Sample size estimates for randomized response procedures. *Progress in Clinical Biological Research*, 325, 1-10.

- Goodstadt, M. S., & Gruson, V. (1975). The randomized response technique: A test on drug use. *Journal of the American Statistical Association*, 70, 814-818.
- Goyder, J. (1989). *The silent minority: Non-respondents on sample surveys*. Boulder, CO: Westview Press.
- Greenberg, B. C., Abul-El, A.-L. A., Simmons, W. R., & Horvitz, D. G. (1969). The unrelated question randomized response model, theoretical framework. *Journal of the American Statistical Association*, 64, 520-539.
- Groves, R. M. (1989). *Survey errors and survey costs*. New York: Wiley.
- Horvitz, D. G., Shah, B. V., & Simmons, W. R. (1967). The unrelated question randomized response model. *Proceedings of the Social Statistics Section, American Statistical Association*, 65-72.
- Johnson, N. L., & Kotz, S. (1970). *Continuous univariate distributions*. Boston: Houghton Mifflin.
- Kendall, M. K., & Stuart, A. (1979). *The advanced theory of statistics* (Vol. 2). New York: Macmillan.
- Kinsey, A. C., Pomeroy, W. B., & Martin, C. E. (1948). *Sexual behavior in the human male*. Philadelphia, PA: W. B. Saunders.
- Kolata, G. (1987, April 24). How to ask about sex and get honest answers. *Science*, 236, 382.
- Krotki, K., & Fox, B. (1974). The randomized response technique, the interview and the self-administered questionnaire: An empirical comparison of fertility reports. *Proceedings of the Social Statistics Section, American Statistical Association*, 367-371.
- Locander, W., Sudman, S., & Bradburn, N. (1976). An investigation of interview method, threat and response distortion. *Journal of the American Statistical Association*, 71, 269-275.
- Mangat, N. S., & Singh, R. (1990). An alternative randomized response procedure. *Biometrika*, 77, 439-442.
- Park, J. W., & Park, H. N. (1987). A new randomized response model for continuous quantitative data. *Proceedings of the College of Natural Science (Seoul)*, 12, 33-44.
- Pollock, K. H., & Bek, Y. (1976). A comparison of three randomized response models for quantitative data. *Journal of the American Statistical Association*, 71, 884-886.
- Rohlf, F. J., & Sokal, R. R. (1981). *Statistical tables* (2nd ed.). New York: Freeman.
- Scheers, N. J. (1992). Methods, plainly speaking: A review of randomized response techniques. *Measurement and Evaluation in Counseling and Development*, 25, 27-41.
- Scheers, N. J., & Dayton, C. M. (1987). Improved estimation of academic cheating behavior using the randomized response technique. *Research in Higher Education*, 26, 61-69.
- Scheers, N. J., & Dayton, C. M. (1988). Covariate randomized response models. *Journal of the American Statistical Association*, 83, 969-974.
- Schimizu, I. M., & Bonham, G. S. (1978). Randomized response technique in a national survey. *Journal of the American Statistical Association*, 73, 35-39.
- Stem, D., & Steinhurst, R. (1984). Telephone interview and mail questionnaire applications of the randomized response model. *Journal of the American Statistical Association*, 79, 555-564.
- Tracy, P., & Fox, J. (1981). The validity of randomized response for sensitive measurements. *American Sociological Review*, 46, 187-199.
- Volicer, B. J., Cahill, M. H., Neuburger, E., & Arntz, G. (1983). Randomized response estimates of problem use of alcohol among employed females. *Alcoholism*, 7, 321-326.
- Warner, S. L. (1965). Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, 60, 63-69.
- Warner, S. L. (1971). The linear randomized response model. *Journal of the American Statistical Association*, 66, 884-888.
- Weisman, F., Moriarity, M., & Schafer, M. (1975). Estimating public opinion with a randomized response model. *Public Opinion Quarterly*, 39, 507-513.
- Weissman, A. N., Steer, R. A., & Lipton, D. S. (1986). Estimating illicit drug use through telephone interviews and the randomized response technique. *Drug and Alcohol Dependence*, 18, 225-233.

(Appendix follows)

Appendix

Calculation of Pearson's χ^2

The formula for Pearson's chi-square is

$$\chi^2 = \sum_{i=1}^2 \left[\frac{(y_i - y_i^*)^2}{y_i^*} + \frac{(n_i - n_i^*)^2}{n_i^*} \right].$$

As stated in the text, y_i and n_i are the number of "yes" and "no" responses in the i th group, and * denotes the maximum-likelihood estimate as a variable. $y_i^* = N_i(\pi^* + \beta^*p_i)$, $n_i^* = N_i - y_i^*$, and $\pi^* = 1 - \beta^*$ for $i = 1, 2$, where N_i is the number of participants in the i th group. β^* is defined as the smaller of 1 or

$$\beta^* = \frac{-b + \sqrt{b^2 - 4ac}}{2a},$$

where $a = -(N_1a_1a_2 + N_2a_2a_1)$, $b = (N_1a_1 + n_1a_2 + N_2a_2 + n_2a_1)$, $c = -(n_1 + n_2)$, $a_1 = 1 - p_1$, $a_2 = 1 - p_2$. By substituting the values from Table 2 into these formulas, we calculate the value of chi-square as follows:

$$a_1 = 1 - p_1 = .25$$

$$a_2 = 1 - p_2 = .75$$

$$\begin{aligned} a &= -(N_1a_1a_2 + N_2a_2a_1) \\ &= -[(500 \times .25 \times .75) + (500 \times .75 \times .25)] \\ &= -187.5 \end{aligned}$$

$$\begin{aligned} b &= N_1a_1 + n_1a_2 + N_2a_2 + n_2a_1 \\ &= (500 \times .25) + (154 \times .75) + (500 \times .75) \\ &\quad + (373 \times .25) \\ &= 708.75 \end{aligned}$$

$$\begin{aligned} c &= -(n_1 + n_2) \\ &= -527 \end{aligned}$$

$$\begin{aligned} \beta^* &= \frac{-b + \sqrt{b^2 - 4ac}}{2a} \\ &= \frac{-708.75 + \sqrt{708.75^2 - 4(-187.5) \times (-527)}}{2(-187.5)} \\ &= 1.02 \end{aligned}$$

Because this is larger than 1, we set $\beta^* = 1$. The reason for setting $\beta^* = 1$ is that $\beta^* > 1$ is equivalent to $\pi^* < 0$. This condition occurs when, under the null

hypothesis of no cheating, there are fewer "yes" responses than one would expect to get from the randomization device alone even if the behavior were nonexistent in the population. In this example, if there were no homosexual behavior and everyone were honest, we should have gotten 500 "yes" responses due to random chance. We got only 473 (346 from Group 1 and 127 from Group 2). If one assumes no cheating (null hypothesis), then the equations try to account for the dearth of "yes" responses by setting $\pi^* < 0$. In these cases, the best one can do to maximize the likelihood of the data under the null hypothesis is to set $\pi^* = 0$. This will occur with high probability when the true frequency of the behavior is low relative to the level of cheating.

$$\pi^* = 1 - \beta^* = 0$$

$$\begin{aligned} y_1^* &= N_1[\pi^* + \beta^*p_1] \\ &= 500[0 + (1 \times .75)] \\ &= 375 \end{aligned}$$

$$\begin{aligned} y_2^* &= N_2[\pi^* + \beta^*p_2] \\ &= 500[0 + (1 \times .25)] \\ &= 125 \end{aligned}$$

$$\begin{aligned} n_1^* &= N_1 - y_1^* \\ &= 500 - 375 \\ &= 125 \end{aligned}$$

$$\begin{aligned} n_2^* &= N_2 - y_2^* \\ &= 500 - 125 \\ &= 375 \end{aligned}$$

$$\begin{aligned} \chi^2 &= \sum_{i=1}^2 \left[\frac{(y_i - y_i^*)^2}{y_i^*} + \frac{(n_i - n_i^*)^2}{n_i^*} \right] \\ &= \left[\frac{(346 - 375)^2}{375} + \frac{(154 - 125)^2}{125} \right] - \left[\frac{(127 - 125)^2}{125} \right. \\ &\quad \left. + \frac{(373 - 375)^2}{375} \right] \\ &= 9.01 \end{aligned}$$

Received November 25, 1996
Revision received July 24, 1997
Accepted September 11, 1997 ■