ISSN: 1541-745X *www.idrjournal.com*

Comparative Validation of a Partial (Versus Full) Randomized Response Technique: Attempting to Control for Social Desirability Response Bias to Sensitive Questions

Bill Thornton* & Sat Gupta
University of Southern Maine

ABSTRACT - Social desirability response bias is prevalent in interview or questionnaire surveys involving personally sensitive questions. The present research provided for a comparative evaluation of a partial and full randomized response technique (RRT) as a statistical technique to circumvent this tendency to present oneself in a socially desirable manner. Also included for validation purposes were face-to-face, assured anonymity, and bogus pipeline conditions (the latter previously demonstrated as a means to elicit more accurate, truthful responding to sensitive items). Consistent with previous computer simulation research (Gupta et al., 2002), the present data based on actual survey respondents similarly suggests the effectiveness of a partial RRT in providing estimates closer to "accurate" responses (obtained in the anonymous and bogus pipeline conditions) and with less error of estimation than that possible using a full RRT. This advantage was evident on personally sensitive items; on non-sensitive items, there was no apparent advantage in utilizing any strategy intended to circumvent response bias.

Social desirability bias (SDB) – the predisposition to self-report so as to present one's self in a socially desirable manner – is one of the oldest (Crowne & Marlowe, 1960; Edwards, 1957; Meehl & Hathaway, 1946) and most documented response bias in social and behavioral sciences (Maher, 1978; Paulhus, 1991). This undesirable, yet pervasive, influence in research is evident in psychology, sociology, education, business, marketing, economics, and any other endeavor relying on self-report assessments (Fisher & Katz, 2000). In particular, SDB is often apparent in the assessment of personality variables (Mick, 1996), attitudes (Fisher, 1993), and self-reported behaviors (Mensch & Kandel, 1988; Meston, Heimen, Trapnell, & Paulhus, 1988). Considered to be a relatively stable trait, rather than a situationally-determined response (Furnham, 1986), SDB typically reflects the desire to present oneself in a conventional, socially acceptable manner (i.e., impression management concerns) and typically results in an overly favorable self-presentation (Nederhof, 1985; Paulhus, 1991; 1992). As such, SDB may impact on the relationships observed among variables in research and thereby affect the reliability and validity of both experimental and survey research findings (Nederhof, 1985; Zerbe & Paulhus, 1987).

**Bill Thornton; Department of Psychology; University of Southern Maine; P.O. Box 9300; Portland, ME 04104-9300; thornton@usm.maine.edu (email).*

The Marlowe-Crowne Social Desirability Scale (Crowne & Marlowe, 1960) was developed as a general measure of SDB and has been used to identify, and statistically control for, tendencies to respond in a socially desirable way so as to elicit approval of others and avoid negative evaluations. Other assessments with similar intent have since been developed and utilized, each with advantages and disadvantages (Furnham, 1986; Paulhus, 1991). Despite the relative ease of including a SDB scale in questionnaire research, the length of a scale, item appropriateness, and generality of the measure may limit the application and/or usefulness of a scale in controlling for SDB (Fisher, 2000).

Other methods to circumvent SDB reflecting positive self-presentational motives and obtain truthful responses to otherwise sensitive questions do exist but are not easily integrated into traditional data collection techniques. One such method is the "bogus pipeline" technique developed by Jones & Sigall (1971) and involves convincing respondents that the measurement and interpretation of their physiological responses while responding to questionnaire items can determine whether the respondent is providing a truthful or deceptive answer. Although no actual physiological assessment is done, ostensible feedback from the use of several test questions (to which the experimenter knows the subject's true answers) makes it apparent to the subject that analysis of the physiological data is indeed able to detect when the subject provides a true or false response. Believing that they can be detected if they are being deceptive, respondents are presumed motivated to provide more truthful responses to personal, sensitive questions. Indeed, reviews and meta-analyses suggest that the BPL does seem to offer the potential to obtain greater truthfulness in self-reports on personally sensitive issues than otherwise would have been obtained (Aquinis, Pierce, & Quigley, 1993; Roese & Jamieson, 1993).

Another strategy to circumvent SDB in responses to direct questioning on sensitive personal issues has been the use of the randomized response technique (RRT). Originally devised by Warner (1965), respondents are able to disguise their true answers through the use of a randomizing device that may direct the person to answer truthfully, or answer either yes or no regardless of the truth. A "corrected" response at the group level is estimated later by statistically adjusting for the known probabilities inherent in the randomizing device. Subsequent variations of the RRT have generally provided for different and presumably more valid response rates on sensitive questions than otherwise obtained by traditional direct questioning methods (Carr, Marascuilo, & Busk, 1982; Fidler & Kleinknecht, 1977; Himmelfarb & Lickteig, 1982; Horvitz, Greenberg, & Abernathy, 1975; Lamb & Stem, 1978; Scheers, 1992).

The effectiveness of the RRT is well established through simulations (Eichorn & Hayre, 1983; Gupta, Gupta, & Singh, 2002), has demonstrated applicability in actual survey research (Kerkvliet, 1994; Sheers, 1992; Sheers & Dayton, 1987; Weissman, Steer, & Lipton, 1986), and therefore may serve to reduce or eliminate a respondent's evaluation apprehension and circumvent SDB. Despite its potential, Fidler & Kleinknecht (1977) noted that the use of the RRT had been largely underutilized in psychological research, a situation that is evident still. However, the use of RRT in real surveys does introduce an additional source of random error in the measurement of responses. Consequently, subsequent estimates to obtain a "corrected" response may prove to be *less* statistically efficient than if the respondent had been asked directly and responded truthfully in the first place (Himmelfarb & Lickteig, 1982).

Mangat and Singh (1990) introduced a variation of the RRT that would require only a randomly selected proportion of respondents to utilize a randomization device to disguise, or scramble, their individual responses to survey items while the remaining respondents would simply respond as usual. As such, this "partial" RRT does not

introduce as extensive random error since not every respondent will be randomizing compared to the usual "full" RRT. Since less random error is introduced by the randomization procedure, subsequent estimates of a "corrected" response may prove more accurate. Indeed, Gupta et al. (2002) have demonstrated in computer simulations greater predictive accuracy when only a subset of responses is randomized as compared to when full randomization is done. Gupta and Thornton (2002) provide for a comparison of the mathematical models for both full and partial RRTs for qualitative questions eliciting a binary response (e.g., "yes or no" answer) and quantitative questions (e.g., "How often have you..."). Although the partial RRT may be more efficient than the full RRT in estimating true responses, this technique has not been examined in a practical survey situation.

The present study describes research that was undertaken to provide for a comparison of partial and full RRTs in an actual survey conducted among young adults with both personally sensitive and non-sensitive questions. In addition to the partial and full RRT conditions, responses to the same survey questionnaire were obtained under three other conditions – anonymous, face-to-face, and bogus pipeline situations – in order to provide for a comparative evaluation of the partial (versus full) RRT.

Method

Subjects and Procedure

A total of 331 female undergraduates (18-29 years old, $M_{age} = 22.0$) participated in this research for extra credit in introductory psychology and mathematics courses. All subjects responded to the same survey questionnaire, but under one of five survey conditions: anonymous survey (n=68), face-to-face survey (n=65), bogus pipeline (n=69), full randomization (n=69), and partial randomization (n=60).

The questionnaire consisted of 20 items including personally sensitive questions and non-sensitive, innocuous questions which were presented in mixed order. Questions were both qualitative (e.g., "Have you/do you...") and quantitative (e.g., "How often/how many...") for each topical item (see Table 1). *Sensitive items* dealt with drug use (marijuana and cocaine), driving while under the influence, having multiple sexual partners, and being the victim and perpetrator of abuse in a relationship. *Non-sensitive items* dealt with eating breakfast, regular exercise, a full night's sleep, and soda consumption. Non-sensitive items were interspersed among sensitive ones. Data gathered among a separate group of similar subjects (n = 80; 19 to 29 years old, M_{age} = 22.2) indicated that sensitive and non-sensitive questions were indeed differentially perceived (mean item sensitivity ratings are presented in Table 1).

Response Conditions

Anonymous. In a classroom setting, subjects were asked, even implored, to respond to the survey as accurately and truthfully as possible, while pointing out that there was no way in which any individual's responses could be identified. Aside from indicating age and sex, they were to put no other information that would personally identify them. Further, when they were done with the survey, they personally would place it in a box with other surveys at which time they could mix them all up to further insure that there was no way to connect a survey with any individual.

Face-to-Face. In this condition, subjects were questioned individually by a same-sex interviewer following a class. The subject was provided a blank survey to follow along while the interviewer proceeded to ask aloud each survey item. The subject had to verbally provide a response that the interviewer recorded. Other than indicating sex and age, no other personal identification was obtained.

Bogus Pipeline. Participating individually, subjects were met by a same-sex experimenter who showed them computer-interfaced physiological equipment to which they would be attached for measurement and assessment purposes while responding to a series of questionnaire items. Following Jones and Sigall (1971), it was explained that the physiological responses could be analyzed to determine whether someone was providing truthful or deceptive responses. Subjects were seated and had electrodes attached to their non-dominant hand and wrist. On the pretext of "initializing" the computer assessment to the individual, subjects were asked a series of demographic questions to which the experimenter knew the correct answer (e.g., subject's age, sex, major, etc.). Subjects were asked to vary their responses, telling the truth and lying. When a subject lied, the experimenter was able to trigger surreptitiously a computer signal "indicative of a deceptive response." Having successfully "initialized the computer" – and having demonstrated to the subject the ability of the computer to detect deception – the subject was provided a blank survey to follow along while the experimenter proceeded to ask aloud the survey questions and record subject responses.

Full Randomized Responding. Subjects were each provided a deck of 35 cards, five of which (14.3%) had "YES, I have the characteristic" written on them, and 30 (85.7%) had "NO, I do not have the characteristic." Additionally, the cards in the deck had numerical values indicated: 0, 1, 2, 3, 4, 5, 6, 7, and 8 with the corresponding frequencies of 12, 4, 4, 4, 4, 2, 2, 2, 1, respectively. Thus, the frequency distribution of the card values had a mean of 2.40 and a variance of 5.78. After shuffling the cards, a subject was to draw a card and record privately whether it said "YES" or "NO" and the value indicated on the card. They then were to use this information in responding to each of the items on the survey questionnaire. For instance, if a quantitative response was indicated ("How often..."), they were to add the card value to their true response and report the total as their answer. On a qualitative question ("Have you..."), if their true answer (either yes or no) agreed with the statement on the card ("Yes, I have..." or "No, I do not have..."), they were to respond with a "yes" answer; however, if their true answer and the card statement did not agree, they were to respond with a "no" answer. As such, it would not be obvious to anyone looking at an item response whether it was a true response, or whether it had been randomly "scrambled."

Partial Randomized Responding. Each subject was provided with a deck of cards similar to those used in the full RRT and asked to select a card, noting "YES" or "NO" and the value of the card. They were also to make note of whether the card was red or black. In the deck of 35 cards, there were 6 red (17.1%) and 29 black (82.9%) cards. All subjects were provided with the same instructions for using the card to disguise their responses to questionnaire items as in the full RRT described above. An additional, final instruction was provided here – if they had a red card, they were to use the card information to scramble their answer; if their card was black, however, they were to ignore the card information and simply respond to items with their straightforward, truthful response. As such, the majority of subjects would be providing truthful responses. Only a small proportion of the subjects would be employing the card as a randomizing device and introducing random error. As with the full RRT above, it would not be obvious when looking at a response whether it was a "true" response or one that had been scrambled.

Results and Discussion

Descriptive statistics were computed at the group level on each questionnaire item for each of the five survey conditions. Responses in both the full and partial RRT conditions were subjected to statistical adjustments, thereby providing for estimates of "corrected" proportions and mean responses for qualitative and quantitative items, respectively (the

statistical models employed are summarized in Appendix 1 and 2). The data for comparison purposes, along with margins of error, are presented in Table 1.

Table 1

Proportion ("Yes") and Mean Response with Estimated Error (in parentheses) by Survey Condition

	Anonymous Survey	Face-to-Face Interview	Bogus Pipeline	Full RRT	Partial RRT	F
SENSITIVE ITEMS						
1. Have you ever used marijuana? (Sensitivity = 4.65)	0.85 (±0.09)	0.80 (±0.10)	0.82 (±0.09)	0.71 (±0.16)	0.88 (±0.09)	1.50
2. How many days have you used marijuana in the last 30 days? (Sensitivity = 5.22)	3.60 (±.2.24)	1.74 (±.1.06)	3.37 (±.1.10)	4.08 (±.1.97)	3.42 (±.1.32)	1.04
3. Have you ever used cocaine? (Sensitivity = 5.55)	0.19 (±0.10)	0.15 (±0.09)	0.18 (±0.09)	0.26 (±0.16)	0.18 (±0.11)	0.62
4. How often have you used cocaine in the last 30 days? (Sensitivity = 5.55)	0.04 (±0.06)	0.06 (±.0.09)	0.00 (±.0.04)	0.37 (±.0.60)	0.07 (±.0.34)	1.26
5. In the last year, have you had multiple sexual partners? (Sensitivity = 7.15)	0.31 (±.0.11)	0.25 (±.0.11)	0.27 (±.0.11)	0.30 (±.0.16)	0.25 (±.0.12)	0.26
6. In the last year, how many sexual partners have you had? (Sensitivity = 7.78)	1.49 (±.0.30)	1.35 (±.0.28)	1.36 (±.0.22)	1.59 (±.0.62)	1.51 (±.0.40)	0.26
7. In the last 30 days, have you driven while legally "under the influence?" (Sensitivity =4.95)	0.22 (±.0.10)	0.06 (±.0.06)	0.18 (±.0.09)	0.30 (±.0.16)	0.29 (±.0.12)	3.66**
8. In the last 30 day, how often have you driven while legally "under the influence?" (Sensitivity = 5.07)	0.63 (±.0.28)	0.26 (±.0.16)	0.54 (±.0.32)	0.70 (±.0.66)	0.49 (±.0.20)	0.67
9. Have you ever been in a relationship where you were the victim of physical abuse by your partner? (Sensitivity = 6.67)	0.22 (±.0.10)	0.15 (±.0.09)	0.19 (±.0.10)	0.33 (±.0.16)	0.15 (±.0.10)	2.48**
10. In how many relationships have you been the victim of physical abuse by your partner? (Sensitivity = 6.63	0.34 (±.0.18)	0.23 (±.0.12)	0.25 (±.0.14)	0.61 (±.0.60)	0.26 (±.0.36)	0.89
11. Have you ever been in a relationship where you physically abused your partner? (Sensitivity = 6.63)	0.03 (±.0.04)	0.08 (±.0.06)	0.04 (±.0.05)	0.25 (±.0.16)	0.08 (±.0.08)	4.93**
12. In how many relationships have you physically abused your partner? (Sensitivity = 6.65)	0.08 (±.0.09)	0.14 (±.0.10)	0.12 (±.0.07)	0.40 (±.0.60)	0.16 (±.0.34)	0.69

LESS SENSITIVE ITEMS

1. Do you typically exercise vigorously?						
(Sensitivity = 2.00)	0.50	0.55	0.48	0.47	0.53	0.39
	(±.0.12)	(±.0.12)	(±.0.12)	(±.0.17)	(±.0.13)	
2. How many days in a typical week do you exercise vigorously?						
(Sensitivity = 2.22)	2.53	2.42	2.31	2.51	2.41	0.18
	(±.0.48)	(±.0.48)	(±.0.52)	(±.0.82)	(±.0.62)	
3. On a typical day, do you drink some kind of soda?						
(Sensitivity = 0.97)	0.54	0.65	0.66	0.60	0.54	0.82
	(±.0.12)	(±.0.12)	(±.0.12)	(±.0.17)	(±.0.13)	
4. In a typical day, how many sodas do you drink?						
(Sensitivity = 0.63)	0.78	1.26	1.00	1.27	1.07	0.76
	(±.0.24)	(±.0.34)	(±.0.26)	(±.0.72)	(±.0.58)	
5. Do you typically eat breakfast?						
(Sensitivity = 0.50)	0.71	0.45	0.70	0.61	0.67	3.24*
	(±.0.11)	(±.0.12)	(±.0.11)	(±.0.17)	(±.0.13)	
6. In a typical week, how many mornings do you eat breakfast?						
(Sensitivity = 0.52)	4.04	2.89	4.01	4.03	3.99	5.02**
	(±.0.52)	(±.0.60)	(±.0.44)	(±.0.90)	(±.0.61)	
7. Do you typically get at least 7 hours of sleep a night?						
(Sensitivity = 0.88)	0.74	0.62	0.75	0.53	0.71	1.53
	(±.0.10)	(±.0.12)	(±.0.10)	(±.0.17)	(±.0.13)	
8. In a typical week, how many days do you get at least 7 hours of sleep?						
(Sensitivity = 0.80)	4.88	4.02	4.70	4.43	4.41	0.13
	(±.0.48)	(±.0.58)	(±.0.40)	(±.0.80)	(±.0.56)	

Note. *df*s = 4,324 $^{*}p < .05$ $^{**}p < .01$

Analyses of variance were conducted on responses to the 20 items by survey condition. Among the 12 sensitive items, significant differences were indicated on only three qualitative items requiring a yes/no response: having driven under the influence, $F(4,324) = 3.66$, $p < .01$; having been abused in a relationship, $F(4,324) = 2.48$, $p < .05$; and having been an abuser in a relationship, $F(4,324) = 4.93$, $p < .01$. On the latter two abuse items, the effect was due to an apparent overestimation by the full RRT which was significantly greater than the estimates provided by the partial RRT for on both the abused and abuser items, *t*s (127) = 2.22 and 2.24, respectively, *p*s < .05. In contrast, on the impaired driving item, partial and full RRTs provided similar estimates (.30 and .29, respectively) and the effect was apparently associated with the under-reporting on the part of respondents in the face-to-face survey condition. Consideration of the differences between partial and full RRT response estimates on the remaining sensitive survey items, there were no statistically significant differences, *t*s (127) $\leq$ 1.10.

Inexplicably, there were significant differences observed on two of the non-sensitive items concerned with whether they typically ate breakfast (yes/no), $F(4,324) = 3.25$, $p < .05$, and how often they ate breakfast in a typical week, $F(4,324) = 5.02$, $p < .01$. For both, the effect appeared to be due to under-reporting again in the face-to-face condition; the estimates provided by the partial and full RRT did not differ significantly on either item, *t*s (127) < 1. Consideration of response estimates between partial and full RRT among the remaining non-sensitive items did not indicate any significant differences, *t*s (127) < 1.

Additional individual analyses were also conducted to compare responses between the anonymity-assured and bogus pipeline conditions. Although there were sometimes slight differences in the responses between these two conditions, none of them were of any significant magnitude, *t*s (135) < 1. Considering the effectiveness of the bogus pipeline technique in eliciting more accurate, truthful responses on sensitive questions (e.g., Aquinis et al., 1993), it would appear that the group-assessment in the present study assuring the anonymity of individual respondents was capable of eliciting equally truthful responses. This is consistent with previous research indicating a bogus pipeline technique may not be necessary when respondent confidentiality and anonymity can indeed be assured (Hill, Dill, & Davenport, 1988; Murray & Perry, 1987). In a direct-questioning situation, the bogus pipeline would be more effective in eliciting truthful responses, however, its practical application in such survey research is obviously limited. As such, a RRT would be advantageous in terms of "portability" and also in terms of selective use with only those items predetermined to be sensitive as nothing is gained with its use on non-sensitive items.

However, compared to direct questioning, sometimes the use of the RRT may produce no significant difference in responses even when questioned about illicit activities such as drug use (e.g., Weissman, Steer, & Lipton, 1986; Zdep, Rhodes, Schwarz, & Kilkenny, 1979). Although Weissman et al. reported that three-fourths of the respondents favored the use of a RRT, they also noted that people who were not sensitive about their activities would have simply preferred direct questioning. In the present study, despite a perceived difference between sensitive and non-sensitive items, there may have been no apparent concern responding to sensitive items as no one refused to answer any question in the face-to-face condition. In contrast, Fidler and Kleinknecht (1977) reported people refusing to respond to sensitive items when questioned directly and that the RRT was able to overcome this resistance and obtain responses to those sensitive, potentially stigmatizing, questions. It may be that young people have become less self-conscious about reporting on their activities, whether sensitive or not. Then, again, there may be differential responding depending on the nature of the sensitive item itself, whether the question reflects moral or social values (e.g., multiple sexual partners) or has legal implications (e.g., abusiveness in a relationship), and an interaction with the sex of the respondent (Gupta, Thornton, Shabbir, & Singhal, 2003).

Each of the methods considered here – excessive assurance of anonymity, bogus pipeline, or randomized response techniques – is concerned with circumventing the impression management concerns underlying SDB, in particular, the overly favorable presentation of one's self to others. However, none of these techniques address that aspect of SDB reflecting self-deceptive concerns reflecting concerns for maintaining or enhancing personal feelings of esteem, control, or well-being and points to the importance of a two-factor approach in assessing SDB (Meston et al., 1988). Even SDB scales typically do not assess the self-deceptive component and therefore cannot be used to statistically control for this type of positive response bias (Paulhus, 1991). As such, it is not yet possible to determine whether the assurance of anonymity, a bogus pipeline strategy, or randomized response techniques are capable of overcoming self-deceptive response bias but would be an area of further research.

In conclusion, despite statistically non-significant differences, there were some rather consistent trends observed in the present data that warrant consideration. As would have been expected, face-to-face questioning generally resulted in a more socially desirable response on qualitative and quantitative questions when compared to either anonymous or bogus pipeline conditions, especially on the more personally sensitive items. Previous research also has demonstrated a RRT to provide for different responses to sensitive

questions compared to a direct assessment and has assumed that the difference is in the direction of greater veracity (e.g., Fidler & Kleinknecht, 1977; Himmelfarb & Lickteig, 1982). The present study has been the only one to provide for conditions in which the accuracy or truthfulness of responses may be determined for purposes of comparison for corrected estimates produced by RRTs. On the basis of the present data, it would appear that the full RRT is less accurate in estimating responses (consistently overestimating proportions and frequencies) compared to a partial RRT. Of particular interest to the present study, comparison of a partial and full RRT indicated that the partial RRT more consistently provided for an adjusted estimate closer to the responses obtained in the anonymous and bogus pipeline conditions than those provided by a full RRT. Moreover, partial RRT provided closer estimates with consistently less error of estimation that full RRT. These results are consistent with those from previous simulation research indicating the advantage of the partial RRT over the full RRT (Gupta et al., 2002). In addition, differences between partial and full RRTs were not evident when considering less sensitive items as similar estimates were provided by both and is consistent with prior research (Fidler & Kleinknecht, 1977).

References

Aquinis, H., Pierce, C. A., & Quigley, B. M. (1993). Conditions under which a bogus pipeline procedure enhances the validity of self-reported cigarette smoking: A meta-analytic review. *Journal of Applied Social Psychology, 23*, 352-373.

Aquinis, H., Pierce, C. A., & Quigley, B. M. (1995). Enhancing the validity of self-reported alcohol and marijuana consumption using a bogus pipeline procedure: A meta-analytic review. *Basic and Applied Social Psychology, 16*, 515-527.

Carr, J. W., Marascuilo, L. A., & Busk, P. (1982). Optimal randomized response models and methods for hypothesis testing. *Journal of Educational Statistics, 7*, 295-310.

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathy. *Journal of Consulting Psychology, 24*, 349-354.

Edwards, A. L. (1957). *The social desirability variable in personality assessment and research*. New York: Dryden Press.

Eichorn, B. H., & Hayre, L. S. (1983). Scrambled randomized response methods for obtaining sensitive quanititative data. *Journal of Statistical Planning and Inference, 7*, 307-316.

Fidler, D. S., & Kleinknecht, R. E. (1977). Randomized response versus direct questioning: Two data-collection methods for sensitive information. *Psychological Bulletin, 84*, 1045-1049.

Fisher, R. J. (1993). Socially desirable responding and the validity of indirect questioning.*Journal of Consumer Research, 20*, 303-315.

Fisher, R. J. (2000). The future of social-desirability bias research. *Psychology and Marketing, 17*, 73-77.

Fisher, R. J., & Katz, J. E. (2000). Social-desirability bias and the validity of self-reported values. *Psychology and Marketing, 17*, 105-120.

Furnham, A. (1986). Response bias, social desirability, and dissimulation. *Personality and Individual Differences, 7*, 385-400.

Gupta, S. N., Gupta, B. C. and Singh, S. (2002). Estimation of sensitivity level of personal interview survey questions. *Journal of Statistical Planning and Inference, 100*, 239-247.

Gupta, S. N., & Thornton, B. (2002). Circumventing social desirability response bias in personal interview surveys. *American Journal of Management and Mathematical Sciences, 22*, 369-383.

Gupta, S. N., Thornton, B., Shabbir, J., & Singhal, S. (2003, October). Randomized response techniques and stratified sampling designs: Methods for dealing with social desirability and gender response bias. Paper presented at the International Conference on Statistics, Combinatorics, and Related Areas, Portland, ME.

Hill, P. C., Dill, C. A., & Davenport, E. C. A reexamination of the bogus pipeline. *Educational and Psychological Measurement, 48*, 587-601.

Himmelfarb, S., & Lickteig, C. (1982). Social desirability and the randomized response technique. *Journal of Personality and Social Psychology, 43*, 710-717.

Horvitz, D. G., Greenberg, B. G., & Abernathy, J. R. (1975). Recent developments in randomized response designs. In J. N. Srivastava (Ed.), *A survey of statistical design and linear models*. NY: American Elsevier.

Jones, E. E., & Sigall, H. (1971). The bogus pipeline: A new paradigm for measuring affect and attitude. *Psychological Bulletin, 76*, 349-364.

Kerkvliet, J. (1994). Estimating a logit model with randomized data: The case of cocaine use. *Australian Journal of Statistics, 36*, 9-20.

Lamb, C. W., & Stem, D. E. (1978). An empirical validation of the randomized response technique. *Journal of Marketing Research, 15*, 616-621.

Maher, B. A. (1978). A reader's, writer's, and reviewer's guide to assessing research reports in clinical psychology. *Journal of Consulting and Clinical Psychology, 46*, 835-838.

Mangat, N. S. and Singh, R. (1990). An alternative randomized response procedure. *Biometrika, 77*, 439-442.

Meehl, P. E., & Hathaway, S. R. (1946). The k factor as a suppressor variable in the MMPI. *Journal of Applied Psychology, 30*, 525-564.

Mensch, B. S., & Kandel, D. B. (1988). Underreporting of substance use in a national longitudinal youth cohort. *Public Opinion Quarterly, 52*, 100-124.

Meston, C. M., Heiman, J. R., Trapnell, P. D., & Paulhus, D. L. (1988). Social desirable responding and sexuality self-reports. *Journal of Sex Research, 35*, 148-157.

Mick, D. G. (1996). Are studies of dard side variables confounded by socially desirable responding? *Journal of Consumer Behavior, 23*, 106-119.

Murray, D. M., & Perry, C. L. (1987). The measurement of substance use among adolescents: When is the "bogus pipeline" method needed? *Addictive Behaviors, 12*, 225-233.

Nederhof, A. J. (1985). Methods of coping with social desirability bias: A review. *European Journal of Social Psychology, 15*, 263-280.

Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (pp. 17-59). San Diego: Academic Press.

Paulhus, D. L. (1992). The balanced inventory of socially desirable responding (Reference manual. BISDR Version 6). Vancouver, BC: University of British Columbia.

Scheers, N. J. (1992). A review of randomized response techniques. *Meaurement and Evaluation in Counseling and Development, 25*, 27-41.

Scheers, N. J., & Daton, C. M. (1987). Improved estimation of academic cheating behavior using the randomized response technique. *Research in Higher Education, 26*, 61-69.

Roese, N. J., & Jamieson, D. W. (1993). Twenty years of bogus pipeline research: A critical review and meta-analysis. *Psychological Bulletin, 114*, 363-375.

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.

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*, *8*, 225-233.

Zdep, S. M., Rhodes, I. N., Schwarz, R. M., & Kilkenny, M. J. (1979). The validity of the randomized response technique. *Public Opinion Quarterly*, *43*, 544-549.

Zerbe, W. J., & Paulhus, D. L. (1987). Socially desirable responding in organizational behavior: A reconception. *Journal of Management Review*, *12*, 250-264.

APPENDIX 1

Randomization Models for Qualitative Variable

Full Randomization

Warner (1965):

π = P (Yes) in the population

p = P (Yes) in the deck

p_y = P(Reported Yes)

$= \pi p + (1-\pi)(1-p)$

$$\hat{\pi}_w = \frac{\hat{p}_y - (1-p)}{2p-1}, \qquad p \neq .5$$

$$\text{var}(\hat{\pi}_w) = \frac{\pi(1-\pi)}{n} + \frac{p(1-p)}{n(2p-1)^2}$$

Partial Randomization

Mangat and Singh (1990). Similar to Warner (1965) except that here a predetermined pro portion (T) of subjects are asked to tell the truth and the rest are asked to report a randomized reponse.

$$p_y = T\pi + (1-T)\{p\pi + (1-p)(1-\pi)\}$$

$$\hat{\pi}_{ms} = \frac{\hat{p}_y - (1-T)(1-p)}{(2p-1) + 2T(1-p)}$$

$$\text{var}(\hat{\pi}_{ms}) = \frac{\pi(1-\pi)}{n} + \frac{(1-t)(1-p)\{1-(1-T)(1-p)\}}{n\{2p-1+2T(2-p)\}^2}$$

$$= \text{var}(\hat{\pi}_w) \text{ if } T = 0$$

$$\text{Note } \text{var}(\hat{\pi}_{ms}) \leq \text{var}(\hat{\pi}_w) \text{ if } T > \frac{1-.2p}{1-p}$$

APPENDIX 2

Randomization Models for Quantitative Variable

Full Randomization

Warner (1971)

X = sensitive variable with mean μ_x and variance σ_x^2

S = scrambling variable with known mean μ_s and known variance σ_s^2, independent of X

Y = Reported response

$= X + S$ (Additive Model)

$$\hat{\mu}_{xf} = \bar{y} - \mu_s$$

$$\text{var}(\hat{\mu}_{xf}) = \frac{\sigma_y^2}{n} = \frac{\sigma_x^2}{n} + \frac{\sigma_s^2}{n}$$

Partial Randomization

A known propotion (T) of subjects report true response and the rest of the subjects randomize.

$$Y = \begin{cases} X & \text{with prob. } T \\ X + S & \text{with prob. } (1-T) \end{cases}$$

$$E(Y) = \mu_x \cdot T + \{\mu_x + \mu_s\}(1-T)$$

$$= \mu_x + (1-T)\mu_s$$

$$\hat{\mu}_{xp} = \bar{y} - (1-T)\mu_s$$

$$\text{var}(\hat{\mu}_{xp}) = \frac{\sigma_y^2}{n}$$

$$= \frac{\sigma_x^2}{n} + \frac{(1-T)\sigma_s^2 + \mu_s^2 T(1-T)}{n}$$

Note $\text{var}(\hat{\mu}_{xp}) = \text{var}(\hat{\mu}_{xf})$ if $T = 0$

$= \frac{\sigma_x^2}{n}$ if $T = 1$

Received 12/01/2003; Accepted 03/02/2004