Validating Self-Reports of Socially Desirable Behaviors

Jane E. Burris, Timothy P. Johnson, and Diane P. O'Rourke Survey Research Laboratory, University of Illinois at Chicago

KEY WORDS: Social Desirability, Marlowe-Crowne, Self-Reports, Cancer Screening

1. INTRODUCTION

Although measures of socially desirable response behavior have been available for several decades, few studies have attempted to determine their ability to discriminate between the accurate and inaccurate reporting of socially desirable behaviors. Health promotion and disease prevention behaviors, such as cancer screening exams, are seen by most respondents and interviewers as socially desirable, similar to voting, giving to charities, and attending religious services (Cahalan 1968; Presser and Stinson 1998). There has been much research on the underreporting of socially undesirable behaviors, such as drug and alcohol use (Aquilino 1994; Warnecke et al. 1997; Johnson and Fendrich 2002; Kim and Hill 2002) and on self-reports of sexual behaviors (Meston, Heiman, Tropnell, and Paulhus 1998). However, to our knowledge, no research has examined the association between social desirability and the reporting of health promotion and disease prevention activities, which not only are generally viewed as socially desirable behaviors, but are often overreported. This paper explores the value of the Marlowe-Crowne Social Desirability Scale short version (Strahan and Gerbasi 1972) in predicting the accuracy of self-reports of such socially desirable behaviors.

Social desirability may be viewed as a personality construct or as a feature of a particular attitude, opinion, or situation (DeMaio 1984). In his view of social desirable responding, Paulhus (1984, 1991) considers its multi-dimensional nature and links it to personality constructs such as optimism, confidence, and self-esteem. Paulhus's conception of social desirable responding includes two dimensions: selfdeception and impression management. Self-deception tends to be a reflection of positive optimism, confidence, and self-esteem, while impression management is believed to be situational (Meston et al. 1998).

Crowne and Marlowe (1964) viewed social desirability as primarily a personality trait in which the respondent typically seeks to present himself or herself in a more favorable light out of the desire or need for social approval. Much research has examined the association between the Marlowe-Crowne scale and sociodemographic variables, such

as gender, age, race, social economic status, and education (Fisher 1967; Clancy and Gove 1973; Gove and Greeken 1977; Bradburn and Sudman 1979; DeMaio 1984; Ross and Mirowsky 1984). Minorities have been shown to score higher on the Marlowe-Crowne scale than do Whites, and women typically have higher scores than men. More educated respondents are reported to score lower on the Marlowe-Crowne than those with less education (Bradburn and Sudman 1979; DeMaio 1984). The Marlowe-Crowne scale also is associated with age: the older the respondent, the more likely he or she is to score high on the scale. Some researchers believe that older respondents give more socially desirable responses because of the era in which they were raised or because of feelings of powerlessness (Bradburn and Sudman 1979; Ross and Mirowsky 1984).

The computer has been promoted as a possible tool for reducing the likelihood that respondents will overreport socially desirable behaviors because computer self-administration is likened to self-administered paper and pencil questionnaires, which eliminate the influence of the presence of an interviewer (Lautenschlager and Flaherty 1990; Booth-Kewley, Edwards, and Rosenfeld 1992; Turner et al. 1998). However, the results of the impact of selfadministered computerized questionnaires on social desirability are not conclusive (Tourangeau and Smith 1996). Because of the frequent use of social desirability scales such as the Marlowe-Crowne measures in surveys investigating self-reports of health behaviors, it is important to explore the impact of this scale on the accuracy of reporting.

1.1 Study Background

The overreporting of participation in health promotion and disease prevention activities is well documented and often leads to overestimates of the effectiveness of such programs (Sudman et al. 1994). The primary purpose of this study was to evaluate four approaches to collecting autobiographical data that may help to reduce the overreporting of such health behaviors as getting Pap smear tests and mammograms. The four experimental methods evaluated include (1) the mode in which data are collected (computer-assisted telephone interview [CATI] vs. audio computer-assisted self-interview [ACASI]), (2) asking about the intention to obtain a procedure before asking whether the procedure has been performed, (3) asking about barriers that might prevent a behavior from occurring before asking whether the behavior occurred, and (4) asking about exceptions to the regularity of behavior and allowing respondents to consider those times when unforeseen events may have prevented them from having a regularly scheduled procedure. To validate self-reports of physical examinations, Pap smear tests, and mammograms, respondents were asked for their permission to have their medical records abstracted.

1.2 Questionnaire Design

The questionnaire was designed to elicit information that would be useful in determining the accuracy of self-reports. The questionnaire contained four major parts: (1) a section on barriers or problems that may prevent cancer screening, (2) a section on the respondent's plan or intention to have a procedure performed some time in the future, (3) a section examining the regularity of behavior, and (4) a section that collected the dates of the respondent's self-reports of physical examinations and cancer screening procedures that were performed over a three-year period. The date section began with a global question asking whether a respondent had a physical examination or a gynecological exam in the last three years. The items within the date section recorded the month and year of the latest exam and worked backwards until reaching three years prior to the month of the interview. Within the context of each year, respondents were asked the date of their latest physical or gynecological exam, Pap smear test, and mammogram.

In addition to the four primary sections, the questionnaire included items on insurance coverage, medical facilities frequented by respondents, memory, the Marlowe-Crowne short scale of social desirability, and demographics. Furthermore, in addition to the above substantive questions, the ACASI questionnaire contained a tutorial, a set of computer use questions for the respondent, and debriefing questions for the interviewer. A voice recording was made of each question and response category in the ACASI questionnaire. For flexibility of administration, respondents were presented with the opportunity to listen to or turn off the sound.

2. METHODS

2.1 Sample Design

A random-digit-dial sample of 12,923 telephone numbers was used to screen households in Champaign-Urbana, Illinois, for eligible respondents. The CATI sample accounted for two-thirds of the phone numbers dialed (8,699 cases). The remainder of the sample was allocated to the ACASI condition (4,224 cases). Women age 50 and older who had lived and received health care in the Champaign-Urbana area for three years prior to the date of their interview were eligible to participate in the study. Households with African-American women were oversampled.

Of those women completing interviews, 82.8% consented to have their medical records abstracted (n=832). Seventy-five percent of those granting consent (n=621) actually returned consent forms to our offices, and for 90% of those (n=560), we were able to obtain complete data from every medical facility for which they had returned a consent form. However, for the other 10% (n=61), only partial or no data were collected. This paper focuses only on the women for whom we were able to obtain complete or partial medical record data (n=588).

2.2 Study Procedures

Interviewers trained in the use of CATI software screened Champaign-Urbana households for eligible respondents. During the screening process, interviewers explained the study, answered questions, informed respondents of the \$10.00 gratuity, and gained cooperation. If an eligible respondent was identified and agreed to participate in the CATI condition, the interview occurred immediately following the screening. Following a completed telephone interview, respondents received a letter thanking them for their participation, a \$10.00 gratuity, two copies of the consent form used for medical records abstraction (one for the respondent, the other for research use), a postage-paid return envelope, and an example of the data abstraction form.

In the ACASI condition, screened households were re-contacted by trained face-to-face interviewers. The questionnaire was completed in the homes of the respondents and at their convenience. Interviewers arrived with a laptop computer, a pair of headphones, and relevant project materials. Prior to the respondent's self-administration of the questionnaire, the interviewer led the respondent through the ACASI tutorial and answered any questions she had regarding the interview, the audio, and/or the use of the computer. Consent forms for medical record abstraction were signed on the day of the interview.

All interviews were conducted between October 2001 and April 2002 in English by female interviewers. Telephone interviews averaged 25 minutes in length, while the average ACASI interview lasted 35 minutes. Record abstraction occurred at the end of the study for all respondents. The only information extracted from respondents' medical records were the

dates of their physical or gynecological exams, Pap smear tests, and mammograms. Area medical facilities were provided with signed consent forms and blank abstraction forms. Records' personnel from each of the medical facilities abstracted all data. The Institutional Review Board (IRB) at the University of Illinois at Chicago reviewed and approved this research, as did the IRBs at the two regional medical facilities from which the majority of the records were abstracted.

2.3 Measures

In this study, we focused on self-reports of physical exams and cancer screening examinations, such as Pap smear tests and mammograms. The selfreport indicator was based on reports of these procedures during a three-year period prior to the date of interview. Self-reports were compared to objective medical records, and concordance was defined as self-report of procedure validated by record confirmation. The Marlowe-Crowne ten-item short version (Straham and Gerbasi 1972) was used to assess respondents' social desirable responding behavior. This scale contains ten true-false statements. The five socially desirable true items in this scale include (1) "I never hesitate to go out of my way to help someone in trouble," (2) "I have never intensely disliked anyone," (3) "When I don't know something, I don't at all mind admitting it," (4) "I am always courteous, even to people who are disagreeable," and (5) "I would never think of letting someone else be punished for my wrong-doings." The five socially desirable false items include (1) "I sometimes feel resentful when I don't get my way," (2) "There have been times when I felt like rebelling against people in authority even though I knew they were right," (3) "I can remember 'playing sick' to get out of something," (4) "There have been times when I was quite jealous of the good fortune of others," and (5) "I am sometimes irritated by people who ask favors of me."

Historically, higher scores on the Marlowe-Crowne scale are considered representative of respondents desirous of presenting themselves in a more favorable light. The alpha reliability coefficient for the short version of the Marlowe-Crowne scale for this sample was 0.65. In addition to the above measures, sociodemographic variables, including age, race, and education, were examined.

2.4 Analysis

After assessing the concordance rate for each procedure, a bivariate analysis was conducted on all primary variables. Logistic regressions also were

employed to explore the associations between the Marlowe-Crowne and procedure concordance after controlling for demographic variables. Although we were interested in the differences between White and African-American respondents, for the purposes of this analysis, minority respondents were grouped into an "Other" category. The races/ethnicities included in this category are African American, Native American, Hispanic/Latina, and Asian American.

2.5 Results

The mean Marlowe-Crowne (MC) score for this sample of respondents was 6.83, with a standard deviation of 2.31 and a mode of 7. As expected, the MC score was highly associated with demographic variables. The mean MC score for White respondents was 6.74, which was significantly lower than the mean score of 7.73 for respondents in the "Other" category (one-way ANOVA, F-test=6.714, df=1, p < .01). Marlowe-Crowne scores also were highly associated with and inversely related to education. As the educational level of the respondent increased, the average MC score significantly decreased (p<.001). Age also covaried with the MC scale, with younger respondents (those age 50 to 59) scoring, on average, lower than other respondents. With an average score of 8.00, respondents age 80 and over scored highest on the MC scale.

As mentioned earlier, concordance was defined as a self-report of an exam validated by objective record confirmation. The concordance rate for each procedure was reasonably high for all three procedures. The highest level of concordance was found for the mammogram procedure. Of the women answering a question regarding having had a mammogram in the preceding three years, we were able to validate via their medical records the self-reports of 87.7% of them. The concordance rate was lowest for the Pap smear test (79.1%). For women reporting receipt or not of a physical exam during the past three years, we were able to validate the reports of 83.2% of the sample. No significant differences in concordance rates were found across the procedures.

An examination of the mean MC score for women for whom we were able to find a records match revealed no significant differences among the reporting of procedures. The average MC score approximated 6.8 for all three procedures. The range for procedure concordance was 6.76 for mammograms to 6.83 for Pap smear tests, with exam concordance being 6.79. The average MC score for women overreporting a procedure was highest for those reporting having a mammogram, (MC = 7.17) and lowest for those reporting having a Pap smear test (MC = 6.85). For all procedures, the mean MC score was higher for women who overreported behaviors than for those who gave accurate reports. However, no significant differences were found between the two groups.

MC scores were examined for their predictive value on the accuracy of self-reports for each procedure after controlling for sociodemographic variables. Logistic regression models that explored the predictive value of the MC measure show that it is not associated with the accuracy of self-reports of any of these procedures. However, for the physical exam procedure, sociodemographic variables, such as age and education, were found to be highly predictive of the accuracy of self-reports.

We also considered the possibility of mode effects. The mean MC score for respondents in the CATI condition was 7.06. In the ACASI condition. the mean MC score was 6.33. The difference between the two conditions was significant (one-way ANOVA, F-test=14.245, df=1, p<.001). With the exception of the two logistic regression models for exam concordance, none of the other models examined by mode were found to be associated with the accuracy of self-reports. However, in both conditions, in the two exam models, the significant predictors of self-reporting behavior was not the MC score but rather sociodemographic variables, such as age and education. The MC score was found to be a significant predictor only for the mammogram procedure within the ACASI condition (p < .05). However, the model itself was not predictive.

3. DISCUSSION

With the exception of the mammogram procedure within the ACASI condition, we found no evidence that the MC scale was predictive of the accuracy of self-reports of cancer screening behaviors for this sample of women. The mode effects found in this study are not surprising and are similar to those reported in other studies, which have examined socially desirable response behavior in telephone interviews versus interviews that are conducted faceto-face (Holbrook et al. 2003).

The fact that the MC scale covaried with sociodemographic characteristics was expected. Although we controlled for sociodemographic variables, we are well aware that this sample of respondents may not be representative of the average female respondent. That being the case, further exploration may be warranted. As a group, these women were highly educated, most were insured, and their self-reports of health behaviors lacked much variability.

In earlier research conducted by one of the authors of this paper, the MC scale was examined for its ability to discriminate between respondents who reported socially undesirable behavior (Johnson and Fendrich 2002). However, little evidence was found to support a relationship between the MC scores and self-report accuracy. Consequently, this paper investigated the question of whether the MC scale would prove to be more predictive of self-reports of socially desirable behaviors. However, no such evidence was found, further calling into question the use of the MC scale as a potential indicator of social desirability bias.

4. **REFERENCES**

Aquilino, W. S., (1994), "Interview Mode Effects in Surveys of Drug and Alcohol Use," *Public Opinion Quarterly*, 58, 210–240.

Booth-Kewley, S., Edwards, J. E., and Rosenfeld, P. (1992), "Impression Management, Social Desirability, and Computer Administration and Attitude Questionnaires: Does the Computer Make a Difference?" *Journal of Applied Psychology*, 77, 562–566.

Bradburn, N. M., and Sudman, S. (1979), *Improving Interview Method and Questionnaire Design*, San Francisco: Jossey-Bass Publishers.

Cahalan, D. (1968), "Correlates of Respondent Accuracy in the Denver Validity Survey," *Public Opinion Quarterly*, 32, 607–621.

Clancy, K., and Gove, W. (1973), "Sex Differences in Mental Illness: An Analysis of Response Bias in Self-Reports," *American Journal of Sociology*, 80, 205–216.

Crowne, D. P., and Marlowe, D. (1964), *The Approval Motive*, New York: John Wiley and Sons.

DeMaio, T. J. (1984), "Social Desirability and Survey Measurement: A Review," in *Surveying Subjective Phenomena* (Vol. 2), eds. C. G. Turner and E. Martia, New York: Russell Sage Foundation, pp. 257–281.

Fisher, G. (1967), "The Performance of Male Prisoners on the Marlowe-Crowne Social Desirability Scale: II. Differences as a Function of Race and Crime," *Journal of Clinical Psychology*, 23, 473–475.

Gove, W. R., and Greeken, M. R. (1977), "Response Bias in Surveys of Mental Health: An Empirical Investigation," *American Journal of Sociology*, 82, 1289–1317. Holbrook, A. L., Green, M. C. and Krosnick, J. A. (2003). "Telephone Versus Face to Face Interviewing of National Probability Samples with Long Questionnaires." *Public Opinion Quarterly*, 67, 79–125.

Johnson, T., and Fendrich, M. (2002), "A Validation of the Crowne-Marlowe Social Desirability Scale," Paper Presented at the American Association for Public Opinion Research, St Petersburg, FL.

Kim, M. T., and Hill, M. N. (2002), "Validity of Self-Reports of Illicit Drug Use in Young Hypertensive Urban African American Males," *Addictive Behaviors*, 28, 795–802.

Lautenschlager, G. J., and Flaherty, V. L. (1990), "Computer Administration of Questions: More Desirable or More Social Desirability?" *Journal of Applied Psychology*, 75, 310–314.

Meston, C. M., Heiman, J. R., Trapnell, P. D., and Paulhus, D. L. (1998), "Social Desirable Responding and Sexuality Self-Reports," *The Journal of Sex Research*, 35, 148–157.

Paulhus, D. L. (1984), "Two-Component Social Desirable Responding and Sexuality Self-Reports," *The Journal of Sex Research*, 35, 148–157.

Paulhus, D. L. (1991), "Measurement and Control of Response Bias," in *Measures of Personality and Social Psychological Attitudes*, eds. J. P. Robinson, P. R. Shaver, and L. S. Wrightsman, San Diego: Academic Press, pp 17–59.

Presser, S., and Stinson, L. (1998), "Data Collection and Social Desirability Bias in Self-Reported Religious Attendance," *American Sociological Review*, 63, 137–145. Ross, C. E., and Mirowsky, J. (1984), "Socially-Desirable Response and Acquiescence in a Cross-Cultural Survey of Mental Health," *Journal of Health and Social Behavior*, 25, 189–197.

Strahan, R., and Gerbasi, K.C. (1972), "Short, Homogeneous Versions of the Marlowe-Crowne Social Desirability Scale," *Journal of Clinical Psychology*, 28, 191–193.

Sudman, S., Warnecke, R., Johnson, T. O'Rourke, D., Davis, A. M. and Jobe, J. B. (1979). "Understanding the Cognitive Processes used by Women Reporting Cancer Prevention Examinations and Tests." *Journal of Official Statistics*, 13(3), 305–315.

Sudman, S., Warnecke, R. W., Johnson, T. O'Rourke, D., and Davis, A. M. (1994), "Cognitive Aspects of Reporting Cancer Prevention Examinations and Tests," National Center for Health Statistics, *Vital Health Statistics*, 6(7).

Tourangeau, R., and Smith, T. W. (1996), "Asking Sensitive Questions: The Impact of Data Collection Mode, Question Format, and Question Context," *Public Opinion Quarterly*, 60, 275–304.

Turner, C. F., Ku, L., Rogers, S. M., Lindberg, L. D., Pleck, J. H., and Sonenstein, F. L. (1998), "Adolescent Sexual Behavior, Drug Use, and Violence: Increased Reporting with Computer Survey Technology," *Science*, 280, 867–873.

Warnecke, R. B., Johnson, T. P., Chavez, N., Sudman, S., O'Rourke, D. P., Lacey, L., and Horm, J. (1997), "Improving Question Wording in Surveys of Culturally Diverse Populations." *Annual Epidemiological Psychology*, 7, 334–342.