Design and Results of the Women's Health Study

Roger Tourangeau, Jared B. Jobe, William F. Pratt, and Kenneth Rasinski

ABSTRACT

The Women's Health Study was a methodological experiment carried out in Chicago. More than 1,000 women took part; a comparison sample of 100 men was also included. The sample was selected from two sources. Most of the women and all of the men were selected from an area probability sample that had been screened to identify women in the eligible age range; the rest of the women were selected from rosters at cooperating abortion clinics and were known to have had an abortion. Questionnaires based on the one used in the National Survey of Family Growth were administered to the sample; the questionnaire included items on abortion, sexual behavior, and illicit drug use. The experiment examined five variables: whether the questionnaire began with a series of medical questions or with questions on pregnancy; whether the interview was conducted by a nurse or field interviewer; whether the interview was done at the respondent's home or outside the home; whether the interviewer or respondent administered the questions; and whether the data were collected on paper or via computer. Of the five experimental factors, the one with the most consistent effect was the method of administering the questions. Self-administration significantly increased the reported number of sexual partners, sexually transmitted diseases, and the level of condom use compared to administration by an interviewer. Computer assistance occasionally interacted with the site of the interview to effect reporting. The other two experimental variables—the version of the questionnaire and the data collection staff—had few discernible effects. None of the variables affected reported drug use over the lifetime.

INTRODUCTION

This study investigates some sources of error in surveys that collect information on sensitive topics, topics that involve illegal or embarrassing activities. More specifically, the study tested procedures to improve the accuracy of data collected in the National Survey of

Family Growth (NSFG). Since its beginning in 1971, NSFG has obtained detailed information on fertility and reproductive health. In each of its four cycles conducted to date, the survey has explored a broad range of sensitive questions that concern topics such as contraceptive practices, pregnancy histories (including fetal and infant deaths), unplanned and unwanted pregnancies, sexually transmitted diseases (STDs), and infertility (Judkins et al. 1991).

From the outset, there have been concerns about the sensitive nature of NSFG questions. For example, the first two cycles of NSFG (carried out in 1973 and 1976) generally excluded women who had never been married because it was believed that many unmarried women would not answer questions about pregnancy and contraceptive practice truthfully. Despite these concerns, the sample was expanded in the 1982 NSFG to represent all women regardless of marital status. Changes to the content of the NSFG questionnaires have also increased the sensitivity of the interview over time. For example, the 1988 NSFG added questions on risk factors for acquired immunodeficiency syndrome (AIDS). Despite the increasingly intimate information being sought in NSFG, response rates have remained high. Around 80 percent of the cases selected for NSFG complete the interview and only about one-third of the non-respondents are outright refusals (Judkins et al. 1991). Moreover, nonresponse to individual questions has generally been less than 1 percent. Of course, the fact that respondents answer the questions offers no assurance that their answers are truthful (Jones and Forrest 1992).

Sensitive Questions in Surveys

From the point of view of survey methodology, this study concerns a very general problem—how to collect data on topics that most people are likely to regard as private. Many surveys include questions about private or potentially embarrassing matters, asking respondents about their annual income, their employment status, and so on. Since the outbreak of the AIDS epidemic, the need for data on such sensitive topics as sexual behaviors and illicit drug use has dramatically increased. But though the need for such data is clear, it is not clear whether the data collected are accurate.

Findings from surveys on sexual behavior illustrate the problems in collecting sensitive data in surveys. Within a closed population (that is, a population with no sexual contacts with outsiders), equal numbers of opposite-sex sexual partners should be reported by men and

women; the total number should be the same because the same sexual pairings are being reported by respondents of both sexes. As Smith (1992) has demonstrated, however, men consistently report more opposite-sex sexual partners than do women, a difference that persists even when differences in the population sizes are taken into account. The most plausible account of the discrepancy is that men overstate their partners and that women overlook theirs. A recent review of the methodological problems in AIDS research described the situation this way: "Most sex research is based on self-reported sexual behavior of unknown validity" (Catania et al. 1990, p. 339). Much the same judgment would apply to research on illicit drug use, abortion, and other sensitive topics.

Improving Reporting on Sensitive Topics. A hypothesis guiding much of the survey literature on reports about sensitive topics is that a major source of error is more or less deliberate misreporting. Questions about sensitive topics create conflicts for the respondents, who generally want to cooperate by giving correct answers, but who also want to avoid embarrassment or, when the behavior in question is illegal, legal repercussions. Much of the methodological research designed to improve answers to sensitive questions has concentrated on techniques that reduce the perceived threat of the questions by increasing the privacy of data collection. More recently, methodological studies have begun to examine the effect of computer-assisted data-collection techniques on reporting of sensitive behaviors.

Increasing the privacy of data collection is widely believed to improve the accuracy of the answers. One of the most practical methods for increasing perceived privacy is to use self-administered questionnaires (SAQs) rather than face-to-face interviews to collect the data. In most surveys, the data are not entirely confidential because at least the interviewer is aware of the respondent's answers; further, when interviewers administer the questions and record the answers, it is possible for other household members to overhear what the respondent is saying. Surveys that employ SAQs (in which respondents record their answers without the mediation of an interviewer) overcome these threats to confidentiality. SAQs generally obtain higher levels of reporting of sensitive behaviors than do face-to-face interviews, with telephone interviews falling somewhere in between the other two modes in levels of reporting (see Bradburn 1983 for a review). The advantages of SAQs have been demonstrated for a number of sensitive topics, including sexual behavior (Boekeloo et al. 1994), illicit drug use (Aquilino and

LoSciuto 1990; Schober et al. 1992; Turner et al. 1992), alcohol consumption (Aquilino and LoSciuto 1990; Hochstim 1967), and abortion reporting (London and Williams 1990; Mosher and Duffer 1994; Mott 1985).

Several studies, for example, have shown that self-administration increases reporting of illicit drug use, alcohol consumption, or both. Aquilino and LoSciuto (1990) compared drug use data collected by interviewers over the telephone with data collected by self-administration as part of a personal interview. They found substantially higher reporting of both drinking and drug use with the self-administered questions. Two subsequent comparisons between face-to-face interviewing and SAQs also found greater reporting of cocaine and marijuana use in the self-administered condition (Schober et al. 1992; Turner et al. 1992). Finally, an early comparison of face-to-face data collection with data collection by mail and telephone revealed lower levels of reported alcohol consumption in face-to-face interviews (Hochstim 1967).

Self-administration also appears to reduce survey respondents' reluctance to admit that they have had an abortion. Mott (1985) reports evidence that self-administration greatly increased the number of abortions reported, and similar results have been obtained in studies of abortion reporting by London and Williams (1990; Mosher and Duffer 1994). Boekeloo and colleagues demonstrate that self-administration also increases reporting on other sexual topics; respondents were more likely to admit to unprotected sexual intercourse and a history of STDs in an SAQ than in a face-to-face interview (Boekeloo et al. 1994).

For all of their advantages, however, SAQs also have their limitations. Self-administration with a paper-and-pencil questionnaire requires that the respondents be able to read. In addition, the routing instructions have to be kept simple; the elaborate skip patterns used in many interviewer-administered surveys may be impossible to duplicate in an SAQ.

Another method that increases the apparent confidentiality of survey responses is the randomized response technique (Warner 1965). In this technique, a random device rather than the interviewer determines what question the respondent answers (e.g., the respondent spins a dial to determine which of two questions to answer); in this way, the interviewer cannot know for sure what the respondent's answer means. The randomized response technique method has been

shown to increase the proportion of women reporting that they have had abortions (Abernathy et al. 1970; I-Cheng et al. 1972; Shimizu and Bonham 1978). However, the procedure is difficult to use in a large survey and greatly complicates the analysis of the results.

OTHER VARIABLES AFFECTING LEVELS OF REPORTING

Answers to threatening questions also appear to be affected by the format and wording of the questions (Bradburn 1983), although the results for these variables are not so well documented as those for self-administration. Whether the items use an open or closed response format appears to have an effect on reporting of sensitive behaviors. For example, Bradburn and colleagues (1979) found that, compared to closed questions, open questions produced increases in reporting that ranged from 14 percent for frequency of sexual intercourse during the past month to 108 percent for frequency of masturbation; over a number of sensitive items, the average increase in the level of reporting was 52 percent for the open as compared to the closed versions of the questions. However, the format of the questions did not affect whether respondents report engaging in the behaviors at all.

Longer questions may also yield fuller reporting. Reports about the frequency of behavior and amount of consumption are subject to memory errors even when there are no motivational obstacles to truthful reporting (see Jobe et al. 1993 for a review). Particularly if the behavior is frequent and episodes are not highly differentiated (as with frequent use of illicit drugs), respondents may not remember how many times they have engaged in the behavior during a particular reference period. By giving respondents more cues and more time to search their memories, longer questions can produce more complete reporting (Marquis and Cannell 1971). In their study of sensitive behaviors, Bradburn and coworkers (1979) found that longer questions produced consistently higher levels of reported behavior.

A promising new technique that may increase perceived privacy and produce more accurate data on sensitive behavior is the computer-assisted self-administered interview (CASI). A study by Waterton and Duffy (1984) found that a computer-administered questionnaire produced greater reports of alcohol use than a conventional face-to-face interview. This study confounds the effects of computer assistance and those of self-administration, as do several other studies on CASI (Locke et al. 1992; Lucas et al. 1977; Robinson and West

1992). There is, however, some evidence that computer assistance by itself can enhance the reporting of sensitive behaviors. In a test of computer-assisted personal inter- viewing (CAPI), Baker and Bradburn (1991) found that CAPI respondents were more likely than respondents to a paper-and-pencil interview to report having used birth control methods in the past month.

Variables for This Study

The Women's Health Study was sponsored by the National Center for Health Statistics, with additional support from the National Science Foundation. Its purpose was to test alternative methods for collecting sensitive information; the results were used in planning for Cycle V of the NSFG. Along with the experiment described here, the Women's Health Study included cognitive interviews and focus group discussions. The experiment was based on the assumption that survey reports about abortion and other sensitive topics might be improved through several means. The study examined three of these strategies for investigation. The first strategy—increasing the privacy of the data-collection process—was already well established in the survey methods literature. This strategy was chosen over some of the other possibilities in the literature because past results suggested that privacy was the single most powerful variable affecting reporting on sensitive topics and that manipulating this variable would have the largest impact on the survey estimates.

In examining the privacy variable, the experiment assessed the impact of self-administration rather than the major alternative, the randomized response technique, because of the practical and statistical difficulties associated with that procedure. However, in addition to the use of self-administered questions, the study tested the impact of moving the interview outside the respondent's home (and away from other family members). Relatively few studies have recorded whether face-to-face interviews involving sensitive topics were conducted in private or with other household members present or able to overhear the respondent's answers. As a result, the effects of the privacy of the setting in which the interview is carried out are unclear. It was hypothesized that moving the interview to a neutral site away from other family members might increase the respondent's sense of privacy and thus improve reporting.

The second approach examined by the study was that of placing the interview as a whole, and especially the questions on abortion, in a medical context. A medical context for the interview might reinforce

the need for accurate data for health planning purposes; in addition, it seemed possible that respondents might be more accustomed to providing candid answers in the setting of a medical interview than in the survey setting. Attempts to foster a medical context were done in two ways: First, interviewers who were themselves medical practitioners were used—that is, nurses and nursing assistants interviewed some of the respondents. Earlier work investigating this approach for collecting sensitive data is scarce and this project tests its effectiveness. Second, the interview began with a long series of questions about medical conditions and procedures.

A final strategy investigated in this project was the use of computer-assisted data collection. Early evaluations suggest that computer assistance may enhance either the apparent privacy of data collection or the perceived objectivity and importance of the study; either way, it was hypothesized that computerization of the data-collection process might increase respondents' willingness to report truthfully.

Methods

This large-scale field experiment was conducted in the city of Chicago. More than 1,000 women were interviewed, along with a small comparison sample of 100 men. The sample was selected from two sources. All of the men and most of the women respondents were selected from an area probability sample that had been screened to identify persons in the eligible age range (ages 15 through 35); the rest of the women were selected from rosters at cooperating health clinics and were known to have had abortions.

Questionnaires based on the one used in the NSFG were administered to the sample; the questionnaires included items on abortion, sexual behavior, and illicit drug use and took about an hour to complete (Rieger et al. 1991). The experiment examined five variables: whether the questionnaire began with a series of medical questions or with questions on pregnancy, whether the interview was conducted by a nurse or a regular field interviewer, whether the interview was done at the respondent's home or at a site outside the home, whether the interviewer or respondent administered the questions, and whether the data were collected via computer or on paper. The analysis examined a number of outcome variables, including the response rates under the various experimental conditions, the level and accuracy of abortion reporting, and the level of reporting on other sensitive topics (such as the number of sexual partners). The focus here is on reports about

sexual behavior (see Jobe et al., in press, for findings on the other topics).

Sample

Area Probability Sample. The area probability portion of the sample was a stratified, multistage sample of dwellings in the city of Chicago, selected using standard methods. In the first stage of selection, a sample of 85 area segments was drawn; each segment consisted of a single block or group of adjoining blocks, defined using data from the 1990 Census. After all of the blocks in the city of Chicago had been sorted by geographic area, a systematic sample of 85 of them was selected. Selection probabilities for each segment were proportional to the 1990 census count of the number of housing units it contained. This method of sample selection assured that each area in the city of Chicago would receive proportionate representation in the sample. Each segment included at least 40 housing units (according to the census data); blocks that did not meet this size standard were linked to adjacent blocks until the combined unit included 40 or more housing units.

In the 85 sample segments, a subsample of dwellings was designated to receive a short screening interview to identify persons eligible for the main experiment. In total, 6,325 occupied dwellings were selected for screening. Screening interviews were completed at 4,659 of these, for a response rate of 73.7 percent. Much of the nonresponse occurred in a few high-rise buildings where the interviewers were unable to gain entry. Since assignment to an experimental condition came after screening, the experimental groups remain comparable despite the relatively low response rate to the screening interview; however, the generalizability of the results may be limited. The screening interview gathered information on the race, gender, age, and Hispanic background of persons living at the dwelling. The screeners yielded information about 10,998 persons, of whom 3,141 were within the eligible age range (i.e., 15 to 35 years old at the beginning of the field period for the experiment).

Clinic Sample. Two Chicago health clinics agreed to cooperate in the study by providing the names of women who had had abortions during the preceding year or so. The time period was defined so that no one would be selected who had had an abortion during the 3-month period before the beginning of data collection for the experiment. (Because the field period for the experiment was delayed, this window of eligibility in fact ended more than 9 months before the experiment

began.) The clinic sample was also restricted to women who lived in the city of Chicago; the eligible age range, however, was expanded slightly relative to that for the area probability sample to include women between the ages of 15 through 40. The two clinics provided a total of 1,088 names. A systematic sample of 732 of these women was selected for the experiment (using procedures described below).

To protect the confidentiality of the women selected from the clinic sample, the first author carried out the selection of both the clinic and area probability samples and was the only person aware of the sample from which the individual cases had been selected. In addition, the interviewers carried out an after-the-fact permission form procedure in which they asked women who completed the interview to sign a release form giving the researchers access to their medical records at their sources of gynecological care. Women from the clinic sample who refused to sign the permission form were dropped from the analysis and their data were eliminated from the data files. A total of 48 members of the clinic sample were dropped for this reason.

Selection of Cases for the Experiment. Between the area probability and clinic samples, a total of more than 4,200 persons was available for the experiment. A subsample of 2,266 were randomly assigned to a treatment cell. Within the area probability sample, the selection of persons for the experiment required several steps. In the first step, each household with eligible members was placed in one of six strata that were defined by gender, age (15 through 19 versus 20 and over), and minority group membership. Households with members in more than one eligible group were randomly assigned to a single stratum. Because it was impractical to interview more than one person from the same household, only one eligible household member was retained for the main study. Then, after each household had been assigned to a single stratum, a systematic sample was selected; the use of a systematic procedure assured that the members of the sample were drawn from all of the area segments. Altogether, 1,564 cases were selected for the experiment from the area probability sample.

The selection process for the clinic cases was considerably simpler than that for the area probability cases. Once each woman on the clinic lists had been classified by age category and minority status, a systematic sample was selected from each group. In total, 732 women were selected from the clinic lists. Table 1 shows the number of cases selected for the experiment by source and stratum.

TABLE 1. *Initial sample sizes.*

Stratum	Source		Total
	Area	Clinic	
Younger minority women	237	82	319
Older minority women	549	398	947
Other younger women	52	22	74
Other older women	372	230	602
Younger men	29	--	29
Older men	325	--	325
Total	1,564	732	2,296

Response Rates. A few names provided by the clinics turned out to be duplicates; in addition, the screening data regarding a person's age were sometimes in error and some members of the sample had moved outside of Chicago before the field period began. After these losses, 1,914 women and 350 men remained eligible for the study. After the sample was fielded, it became necessary to subsample males as a cost-saving measure; ultimately, only 100 men were interviewed. Table 2 shows the response rates for the study; the overall response rate for women was 55.2 percent. More than two-thirds of the nonrespondents were cases who were never contacted (primarily because they could not be located during the field period); of the women who were contacted, about 85 percent completed an interview.

Of the 354 completed clinic cases, 48 refused to sign permission forms and the permission forms for another 6 women were lost; data for these cases were dropped, leaving 300 clinic cases for the analysis.

Although the response rates for the study seem relatively low, they are comparable to the rates obtained by NSFG in large metropolitan areas; in Cycle IV of NSFG, the response rate for the 10 largest metropolitan areas was below 60 percent (Rieger et al. 1991). It is difficult to predict the overall impact of nonresponse on the obtained levels of reporting. It seems likely that persons who are very active sexually are less willing to take part in surveys than persons with few partners; similarly, women who have had an abortion may be more reluctant to take part than other women. As table 2 indicates, the response rate was lower for the clinic sample than for the area sample. As a result of such differences in

TABLE 2. *Response rates*.

Group	Number eligible	Number complete	Response rate
Males	350	100	--
Females	1,914	1,059	55.2%
Black	1,053	652	61.7%
White	569	270	47.5%
Hispanic	191	88	46.1%
Other	101	49	48.5%
Area	1,191	705	59.1%
Clinic	723	354	48.8%

response rates, the figures presented here probably underestimate levels of sexual activity. Underreporting would, in addition, increase this bias.

Experimental Design

Five variables were manipulated in this experiment in a completely crossed design. Two of the variables, interviewing staff and version of the questionnaire, were attempts to enhance the medical context of the interview; it was thought that respondents might be more willing to discuss sensitive topics in a survey if the context reinforced the health-related purposes of the study and if medical practitioners administered the questions.

Accordingly, interviewing staff was varied, comparing nurses and nursing assistants with regular field interviewers. The hypothesis was that nurses would elicit more reports of sensitive behaviors than would regular field interviewers. The two versions of the questionnaire included the same items but varied the order in which two sets of abortion questions appeared. In one version, a series of pregnancy history questions came first; in the other, a set of questions about medical procedures was first. In the pregnancy first version, the topic of abortion was initially raised during a series of questions about the respondent's pregnancy history; in the medical conditions first version, abortion was first mentioned in connection with a series of medical procedures affecting reproduction. The study tested the hypotheses that more abortions would be reported by respondents receiving the medical procedures questions first and that the

combination of the two sets of abortion questions would yield more reported abortions than either set of questions alone.

The experiment varied the mode of data collection, comparing paper-and- pencil to computer-assisted interviews, and the method of administration, comparing interviewer-administered to self-administered interviews. Crossing the mode of data collection and method of administration resulted in four groups: interviewer-administered paper-and-pencil interviews (PAPI); computer-assisted personal interviews (CAPI);
paper-and-pencil self-administered questionnaires (SAQ); and computer-assisted self-administered questionnaires (CASI). It was hypothesized that respondents in both self-administered conditions (those completing the SAQ or CASI questionnaires) would report higher levels of sensitive behaviors.

The site of data collection was varied so that interviews were conducted either in the respondent's home or at a neutral site. Levels of reporting were expected to be higher in the neutral site interviews, where other members of the household could not overhear the answers. A variety of sites were used for the interviews conducted outside the home, with the offices of the National Opinion Research Center and neighborhood restaurants being the most frequent.

Instruments

At the beginning of each interview, the respondent was asked to note three or four important personal events on a calendar to help date events later in the questionnaire. Both versions of the questionnaire began by asking demographic questions. These were followed by the medical procedures and pregnancy history questions in counterbalanced order; both of these series of questions included items on abortion. The pregnancy history questions were the questions usually used on NSFG, and asked the respondent to list all her pregnancies in order and to report certain data about each pregnancy, including its outcome (i.e., live birth, stillbirth, ectopic pregnancy, miscarriage, or abortion). The medical procedures questions were developed for this experiment and asked whether the respondent had had any of a number of medical procedures affecting reproductive health. Six of the procedures were methods for inducing an abortion: dilation and curettage (D and C) to end a pregnancy; dilation and evacuation (D and E) or suction curettage to end a pregnancy; injection of saline solution or prostaglandin to end a pregnancy;

hysterectomy to end a pregnancy; hysterectomy during a pregnancy; and abortion, type unknown.

For the remaining topics, the two versions of the questionnaire were identical. Both versions contained numerous questions about the respondent's sexual behavior. Items asked when and with whom the respondents first had sexual intercourse, and whether it was voluntary; other items asked about the number of sex partners during the previous year, the previous 5 years, and in total. The questionnaires also contained items on whether respondents had ever had an STD. In the section of questions on medical conditions, respondents were asked whether they had had chlamydia, gonorrhea, genital warts, genital herpes, or syphilis. Finally, there were items asking the respondents about their use of condoms in the last year and the last 30 days.

The questionnaires also included a series of items on illicit drug use. The initial drug question asked whether the respondent had ever used any illegal drug, and followup questions asked about their use of marijuana, amphetamines, barbiturates, tranquilizers, psychedelics, cocaine, crack, and heroin. Another series of questions, for users of injectable drugs, asked how they cleaned their needles and related drug paraphernalia, and how often they shared them with other users.

RESULTS

The discussion of the results focuses on sexual behaviors. More specifically, the research examined the average number of sexual partners reported as a function of the gender of the respondent and of the experi-mental variables; responses were then examined on the other sexual topics in the questionnaire, including STDs and condom use. The results bearing on abortion reporting are discussed elsewhere (Jobe et al., in press). Because so few men completed the interview, reports here are mainly the results for women. Because the emphasis is on comparisons between the different experimental groups, the results reported here are unweighted.

Reported Sexual Partners

The data on the number of reported sex partners are counts and, as is common with such data, the distribution of the responses is highly skewed. To compensate for this departure from normality, 0.5 was added to the values and a logarithmic transformation carried out

before the analyses of variance were performed. For ease of interpretation, untransformed values are reported in the group means. For respondents who had been sexually active for only 1 year, or for only 5 years, the number of sexual partners for that period was used as the value for questions about longer time spans.

Experimental Effects. For all three time periods, women who completed SAQs reported more sexual partners than women who responded to questions administered by an interviewer. There were significant effects for the method of administration for reported partners during the past year, the past 5 years, and the respondent's lifetime. For the past year, the women who answered self-administered questions reported a mean of 1.72 sexual partners versus 1.44 for those who answered questions administered by an interviewer ($F(1,39) = 9.30$, $p < 0.01$). For the 5-year period, women who completed SAQs reported a mean of 3.87 sexual partners versus 2.82 for those who answered interviewer-administered questions ($F(1,39) = 5.74$, $p < 0.05$). For the lifetime item, women who completed SAQs reported a mean of 6.51 sexual partners versus 5.43 for those who answered questions administered by an interviewer ($F(1,39) = 9.54$, $p < 0.01$). No other main effects were significant.

Computerization seemed to interact with the site of the interview to affect the number of sexual partners reported. During home interviews, more sexual partners were reported by women interviewed using computer- assisted questionnaires than by those responding to conventional, paper- and-pencil questionnaires; for women interviewed outside the home, more sexual partners were reported on the pencil-and-paper questionnaires. Table 3 displays the relevant means. For the previous year, women interviewed at home reported fewer sexual partners on the paper-and-pencil questionnaires than on the computer-assisted ones (1.36 versus 1.84), whereas the women interviewed outside the home reported more partners on the paper-and-pencil than on the computer-assisted questionnaires (1.68 versus 1.43; $F(1,39) = 7.72$, $p < 0.01$). Similarly, for the lifetime partners question, women interviewed at home reported fewer partners on the paper-and-pencil than on the computer-assisted questionnaires (5.06 versus 7.48), whereas those interviewed outside their homes showed the opposite pattern, reporting more partners on the paper-and-pencil than on the computer-assisted questionnaires (6.26 versus 5.08; $(F1,39) = 5.89$, $p < 0.05$). The pattern is in the same direction but not significant for the 5-year partners item. Overall levels of reporting are consistently higher using computer-assisted questionnaires, although not significantly so. Bringing computers into

TABLE 3. *Average number of reported sexual partners by mode and site.*

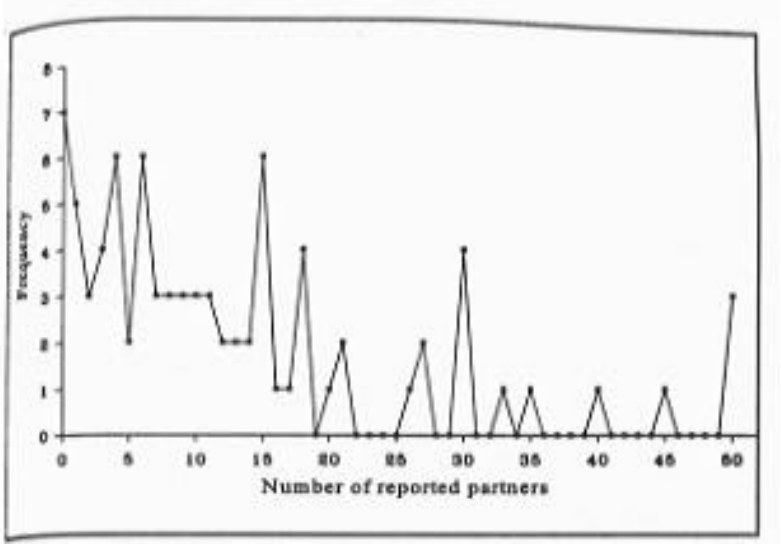
	At home		Outside the home	
	Paper	Computer	Paper	Computer
1 year	1.36	1.84	1.68	1.43
5 years	2.81	4.51	3.33	2.74
Lifetime	5.06	7.48	6.26	5.08

NOTE: Means based on untransformed counts.

the respondents' homes may have fostered a sense of the importance or objectivity of the survey, promoting fuller reporting of sexual partners. Outside the home, especially in public places, the computer may make respondents feel conspicuous, inhibiting reporting.

Males versus Females. As has been observed in earlier surveys on sexual behavior, the men reported more opposite-sex sexual partners than the women did. This was true for the past year (4.19 for the men versus 1.58 for the women), the past 5 years (12.47 versus 3.34), and lifetime (23.96 versus 5.97); all 3 differences are highly significant (F values all greater than 10; p values all less than 0.001). In the analyses that include the data for men, the main effect of self-administration remains significant and that variable does not interact with sex. (The sex of the respondent did occasionally enter into higher order interactions with the experimental variables, but none of these interactions was readily interpretable.)

Rounding of Values. Morris (1993) has argued that the discrepancy between men and women in the reported number of sexual partners largely reflects differences within the subgroup of respondents with a relatively large number of partners to report; within this subgroup, the differences between men and women may reflect differences in rounding behavior (with the women rounding their answers down and the men rounding theirs up). Figure 1 shows the distribution of the number of lifetime sexual partners reported by the men in the current authors' study. (The results for the women, which are not shown, are quite similar.) The preponderance of reported values that are exact multiples of 5 strongly suggests that respondents of both sexes are reporting their answers in round numbers. More than 57.2 percent of the 145



respondents who reported 11 or more sexual partners gave an answer that is a multiple of 5.

Other Sexual Topics

Women who answered self-administered questions reported more STDs than those answering questions administered by an interviewer (22.0 percent versus 17.0 percent). This effect of the method of administration was only marginally significant ($0^2_1 = 2.93$, $p < 0.10$). No other main effects or interactions were significant. Results were the same for both the logistic regression models and chi-square tests.

The ratio between two items concerning condom use and sexual intercourse in the last 30 days was analyzed; the ratio represented the percentage of time the respondent used a condom in the past month. An analysis of variance was performed to examine this variable. Women who reported that they had not had sexual intercourse in the last 30 days were dropped from this analysis; data from 641 women were included in the analysis. Significantly more condom use was reported with self-administered questionnaires (average reported use 47 percent of the time) than with interviewer-administered questions (35 percent). The main effect for the method of administration variable was significant ($F(1,39) = 8.18$, $p < 0.001$). Apparently, many respondents still regard the use of condoms as embarrassing behavior. No other significant effects were found on the condom use variable.

DISCUSSION

Effects of Self-Administration

The variable with the most consistent impact on the level of reporting was the method of administering the questions. Women who completed SAQs reported more sexual partners, more STDs, and greater use of condoms than those who responded to questions read by an interviewer. These findings are summarized in table 4, which displays the ratio between the levels of reporting under the self-administered and interviewer-administered conditions. As the table shows, the levels of reporting are substantially higher—from 19 to 37 percent higher—when the questions are self-administered. The effects of self-administration are similar for men. The lack of effects for the site of the interview suggests that respondents may be more concerned about the reactions of the interviewer than about the threat of other family members' overhearing.

TABLE 4. *Reported sexual behavior.*

	Method of administration		
	Self-administered	Administered by interviewer	Ratio
Sexual partners			
Past year, women	1.72	1.44	1.19
Past year, men	4.52	3.88	1.16
Past 5 years, women	3.87	2.62	1.37
Past 5 years, men	14.72	10.43	1.41
Lifetime, women	6.51	5.43	1.20
Lifetime, men	22.76	25.00	0.91
Condom use (women)			
Past 30 days	46.7%	35.3%	1.32
Past year	23.8%	17.9%	1.33
STDs (women)	22.0%	17.0%	1.29

The findings on the impact of self-administration are quite consistent with the results of earlier comparisons of SAQs with face-to-face interviews carried out by field interviewers. The largest studies comparing the two methods of data collection are those reported by Schober and colleagues (1992) and by Turner and colleagues (1992). Both showed that self-administration resulted in higher levels of reported use of illicit drugs. In both studies, the effect of self-administration was restricted to recent as

opposed to lifetime drug use; unfortunately, the questionnaire asked only about lifetime use and no effects were found for the experimental variables on drug reporting. The impact of self-administration was not entirely uniform across topics. For example, no significant effect was found on abortion reporting (Jobe et al., in press), even though several earlier studies (such as London and Williams 1990; Mott 1985) found that self-administered questions increased abortion reporting. It is not clear why these earlier findings could not be replicated.

Effects of Site and Medical Context

In contrast to the clear results for self-administration, few effects were observed for the site of the interview. In addition, no effects were observed for either of the attempts to induce a medical context for the questions.

Several studies have attempted to observe the impact of the presence of other family members on reports of sensitive behaviors. For example, in two studies on illicit drug use reporting, interviewers noted whether other family members were present during the interview (Schober et al. 1992; Turner et al. 1992); neither study found an effect of this variable on reported drug use. Mosher and Duffer (1994), on the other hand, report an effect for the site of the interview on abortion reporting. It may be that the effects of this variable are hard to observe consistently. As has already been suggested, respondents may be worried less about the reactions of other household members than about those of the interviewer. In addition, respondents may live alone, or with others (e.g., infants) whose presence is not a cause for concern. Such circumstances will reduce the impact of the site of the interview and make it difficult to demonstrate the effect of this variable.

Neither the version of the questionnaire nor the type of interviewer collecting the data had any discernible effects on reporting. These variables may have made little impression on the respondents. The nurses did not wear distinctive uniforms and, although they introduced themselves as nurses, this fact probably did not remain very salient to the respondents as the interview progressed. It is also quite possible that respondents see nurses and other medical personnel as authority figures and are no more willing to make embarrassing revelations to them than to ordinary survey interviewers. Several studies demonstrate that respondents admit more sensitive behaviors in an SAQ than they do in interviews conducted by medical personnel (e.g., Boekeloo et al. 1994; Locke et al. 1992); these results suggest that respondents withhold sensitive information from medical personnel just as they do with field interviewers.

Effects of Computerization

Computerization by itself had no consistent effects on levels of reporting among the respondents. Instead, the effects of computer assistance seemed to vary somewhat by the topic of the question and the site of the interview (see table 3). In reports on sexual partners, computer assistance seemed to increase the number of partners reported when data collection took place in the home, but it reduced the number reported when data collection took place outside the home. No compelling explanation for this mode by site interaction suggests itself.

Past investigations of computerized interviewing have tended to emphasize its effects on item nonresponse, timeliness, and cost rather than on the answers that are obtained. Only a few studies have reported effects of computer-assisted data collection on levels of reporting. The experiment comparing CAPI with conventional paper-and-pencil data collection on the National Longitudinal Study of Labor Market Behavior/Youth Cohort found that more respondents reported using birth control under CAPI than under paper-and-pencil interviewing (Baker and Bradburn 1991). Several other studies have shown effects on reporting for computer-assisted self-administration, but in these studies, it is impossible to disentangle the effects of computerization from those of self-administration (e.g., Waterton and Duffy 1984). It appears that computerization by itself has little effect on the answers respondents give, a conclusion consistent with much of the previous literature on computer-assisted telephone interviewing (Groves and Mathiowetz 1987). Nevertheless, computerization may have subtle effects on the respondent, effects that can vary depending on the circumstances of the interview.

REFERENCES

Abernathy, J.; Greenberg, B.; and Horvitz, D. Estimates of induced abortion in urban North Carolina. *Demography* 7:19-29, 1970.

Aquilino, W., and LoSciuto, L. Effect of interview mode on self-reported drug use. *Public Opin Q* 54:362-395, 1990.

Baker, R.P., and Bradburn, N.M. "CAPI: Impacts on Data Quality and Survey Costs." Paper presented at the Public Health Conference on Records and Statistics, Washington, DC, 1991.

Boekeloo, B.; Schiavo, L.; Rabin, D.; Conlon, R.; Jordan, C.; and Mundt, D. Self-reports of HIV risk factors at a sexually transmitted disease clinic: Audio vs written questionnaires. *Am J Public Health* 84:754-760, 1994.

Bradburn, N.M. Response effects. In: Rossi, P.; Wright, J.; and Anderson, A., eds. *Handbook of Survey Research*. New York: Academic Press, 1983.

Bradburn, N.; Sudman, S.; and Associates. *Improving Interview Method and Questionnaire Design: Response Effects to Threatening Questions in Survey Research*. San Francisco: Jossey-Bass, 1979.

Catania, J.; Gibson, D.; Chitwood, D.; and Coates, T. Methodological problems in AIDS behavioral research: Influences on measurement error and participation bias in studies of sexual behavior. *Psychol Bull* 108:339-362, 1990.

Groves, R., and Mathiowetz, N. A comparison of CATI and non-CATI questionnaires. In: Thornberry, O., ed. *An Experimental Comparison of Telephone and Health Interview Surveys*. National Center for Health Statistics, Vital and Health Statistics, Series 2, No. 106. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1987.

Hochstim, J. A critical comparison of three strategies of collecting data from households. *J Am Stat Assoc* 62:976-989, 1967.

I-Cheng, C.; Chow, L.P.; and Rider, R.V. The randomized response techniques as used in the Taiwan outcome of pregnancy study. Publication 33, No.11:265-269. New York: Population Council, 1972.

Jobe, J.B.; Pratt, W.F.; Tourangeau, R.; Baldwin, A.; and Rasinski, K. Effects of interview mode on sensitive questions in a fertility survey. In: Lyberg, L., et al., eds. *Survey Measurement and Process Quality*. New York: Wiley, in press.

Jobe, J.B.; Tourangeau, R.; and Smith, A.F. Contributions of survey research to the understanding of memory. *Appl Cog Psychol* 7:567-584, 1993.

Jones, E., and Forrest, J. Underreporting of abortions in surveys of U.S. women: 1976 to 1988. *Demography* 29:113-126, 1992.

Judkins, D.; Mosher, W.; and Botman, S. *National Survey of Family Growth: Design, Estimation, and Inference.* National Center for Health Statistics, Vital and Health Statistics, Series 2, Vol. 109. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1991.

Locke, S.E.; Kowaloff, H.B.; Hoff, R.G.; Safran, C.; Popovsky, M.A.; Cotton, D.J.; Finkelstein, D.M.; Page, P.L.; and Slack, W.V. Computer-based interview for screening blood-donors for risk of HIV transmission. *JAMA* 29:1301-1305, 1992.

London, K., and Williams, L. "A Comparison of Abortion Underreporting in an In-person Interview and Self-administered Questionnaire." Paper presented at the Annual Meeting of the Population Association of America, Toronto, Canada, May 1990.

Lucas, R.W.; Mullen, P.J.; Luna, C.B.X.; and McInroy, D.C. Psychiatrist and computer interrogators of patients with alcohol-related illnesses: A comparison. *Br J Psychiatry* 131:160-167, 1977.

Marquis, K., and Cannell, C. *Effect of Some Experimental Techniques on Reporting in the Health Interview*. National Center for Health Statistics, Vital and Health Statistics, DHEW Pub. No. 1000, Series 2, No. 41. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1971.

Morris, M. Telling tails explain the discrepancy in sexual partner reports. *Nature* 365:437-440, 1993.

Mosher, W.D, and Duffer, A.P., Jr. "Experiments in Survey Data Collection: The National Survey of Family Growth Pretest." Paper presented at the meeting of the Population Association of America, Miami, FL, May 1994.

Mott, F. "Evaluation of Fertility Data and Preliminary Analytic Results from the 1983 Survey of the National Longitudinal Surveys of Work Experience of Youth." Columbus, OH: National Institute of Child Health and Development by the Center for Human Resources Research, January 1985.

Rieger, S.; Judkins, D.; and Sperry, S. *National Survey of Family Growth: Cycle IV CATI Phase. Final Report*. Hyattsville, MD: National Center for Health Statistics, 1991.

Robinson, R., and West, R. A comparison of computer and questionnaire methods of history-taking in a genito-urinary clinic. *Psychol Health* 6:77-84, 1992.

Schober, S.; FeCaces, M.; Pergamit, M.; and Branden, L. Effects of mode of administration on reporting of drug use in the National Longitudinal Survey. In: Turner, C.; Lessler, J.; and Gfroerer, J., eds. *Survey Measurement of Drug Use: Methodological Studies*. Rockville, MD: National Institute on Drug Abuse, 1992. pp. 267-276.

Shimizu, I., and Bonham, G. Randomized response technique in a national survey. *J Am Stat Assoc* 73:35-39, 1978.

Smith, T. Discrepancies between men and women in reporting number of sexual partners: A summary from four countries. *Soc Biol* 39:203-211, 1992.

Turner, C.; Lessler, J.; and Devore, J. Effects of mode of administration and wording on reporting of drug use. In: Turner, C.; Lessler, J.; and Gfroerer, J., eds. *Survey Measurement of Drug Use: Methodological Studies.* Rockville, MD: National Institute on Drug Abuse, 1992. pp. 177-220.

Warner, S. Randomized response: A survey technique for eliminating evasive answer bias. *J Am Stat Assoc* 60:63-69, 1965.

Waterton, J., and Duffy, J. A comparison of computer interviewing techniques and traditional methods for the collection of self-report alcohol consumption data in a field survey. *Int Stat Rev* 52:173-182, 1984.

AUTHORS

Roger Tourangeau, Ph.D.
Senior Scientist
National Opinion Research Center
1350 Connecticut Avenue, NW
Washington, DC 20036

Jared B. Jobe, Ph.D.
Chief
Cognitive Functioning and Aging
National Institute on Aging
7201 Wisconsin Avenue
MSC 9205
Bethesda, MD 20892-9205

William F. Pratt, Ph.D.
Division of Vital Statistics
National Center for Health Statistics (Retired)

Kenneth Rasinski, Ph.D.
Senior Survey Methodologist
National Opinion Research Center
1155 East 60th Street
Chicago, IL 60637

Click here to go to page 366