

Sources of Error in a Survey on Sexual Behavior

*Roger Tourangeau¹, Kenneth Rasinski¹, Jared B. Jobe², Tom W. Smith¹
and William F. Pratt²*

This article describes the results of a methodological study on survey reports about sexual behaviors and other sensitive topics. More than 1,000 women took part in the experiment; the study also included a comparison sample of 100 men. Most of the respondents were selected from an area probability sample; in addition, some of the women were selected from the rosters of two abortion clinics. 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 investigated five experimental factors; the one with the most consistent effect on the results was the method of administering the questions. Self administration significantly increased the number of sexual partners, sexually transmitted diseases, and the level of condom use reported by the women compared to administration by an interviewer. Self administration also increased the number of sexual partners reported by men. Computer assistance occasionally interacted with the site of the interview (in the respondent's home or outside the home) to effect reporting. The other two experimental variables – the version of the questionnaire and the data collection staff – had few discernible effects.

Key words: Computer administration; CASI; interviewer effects; mode effects; sensitive questions.

1. Introduction

Since the onset of the AIDS epidemic, the demand for data on such sensitive topics as sexual behaviors and illicit drug use has increased sharply. Surveys have traditionally included questions about private, embarrassing, or illegal behaviors, and even surveys

¹ National Opinion Research Center, 1350 Connecticut Ave., NW, Washington, DC 20036, U.S.A.

² National Institute on Aging, Behavioral and Social Research Program, 7201 Wisconsin Avenue, Bethesda, MD 20892-9205, U.S.A.

Acknowledgments: Roger Tourangeau is now Senior Methodologist at the Gallup Organization. Kenneth Rasinski is Senior Survey Methodologist at NORC. Jared B. Jobe is now Chief, Adult Psychological Development Branch, at the National Institute on Aging. Tom Smith is Director of the General Social Survey at NORC. William F. Pratt was with the Division of Vital Statistics at NCHS and is now retired. The research reported here was supported by Contract 200-91-7099 from NCHS to NORC. Funds for this project were provided by the Office of the Associate Director HIV/AIDS, Centers for Disease Control and Prevention, the Center for Population Control, National Institute of Child Health and Human Development, and NCHS. Additional funds were provided by an NSF grant (SES 9122488) to Tom Smith and Roger Tourangeau.

We thank Christine Bachrach and Norman Bradburn for their valuable suggestions during the design of this study; Lisa Lee, Susan Heine, and Cheryl Gilbert, who played indispensable roles in the collection of the data presented here; Gwen Merker and Ellen Schwartzbach, who trained the interviewers and oversaw the collection of data in the experiment; our field managers, Ezella Pickett, Debra Garrison, and Sue Martinson, who directly supervised the interviewers; Ron Dorsey and Geoff Walker, who were responsible for the development of the data collection systems; Alison Baldwin and Steven Pedlow, who assisted with the data analysis; and Joan Law, who served as project manager. We also thank William Nicholls II and two anonymous reviewers for their valuable suggestions for improving the article.

on otherwise innocuous topics have often included individual items on sensitive issues, such as household income. But in recent years survey questionnaires have ventured into ever more sensitive territory, such as homosexual behaviors or specific sexual activities. Although the need for such data is clear, the accuracy of the information collected in surveys on these topics remains very much at issue.

Several findings provide an empirical basis for concerns about the accuracy of self-report data on sensitive topics. For example, since its inception, the National Survey of Family Growth (NSFG) has asked women about abortions. From these data, it is possible to estimate the total number of abortions performed in specific years. Comparisons of these estimates with figures derived from abortion clinics suggest that fewer than half of all the abortions are reported in the survey (Jones and Forrest 1992). There are also grounds for scepticism about the findings from surveys on sexual behavior. Within a closed population, equal numbers of opposite-sex sexual partners should be reported by men and women, because the same pairings are being reported by respondents of both sexes. As Smith (1992) has documented, however, across a range of surveys in several countries, men report more opposite-sex sexual partners than women do, a difference that persists even when differences in the population sizes are taken into account. A 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, Gibson, Chitwood, and Coates 1990, p. 339). The findings on abortion reporting and on the relative numbers of sexual partners reported by men and women suggests that the validity is far from perfect.

These findings also illustrate two of the main methods used to assess the accuracy of data on sensitive topics – comparison of survey reports with external records (in the case of abortion reports, at an aggregate level) and internal consistency checks. Many methodological studies on sensitive topics have not measured accuracy directly; instead, they have simply assumed that the sensitive behavior in question is underreported and that methods that increase the level of reporting have improved accuracy.

1.1. Collecting sensitive data in surveys

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. Presumably, the vast majority of respondents who agree to be interviewed want to cooperate by providing accurate information; at the same time, however, respondents doubtless seek to avoid embarrassment or, when the behavior in question is prohibited by law, legal repercussions that might result from honest reporting. Much of the methodological research designed to improve answers to sensitive questions has concentrated on techniques that reduce the perceived threat of embarrassment or other consequences by increasing the privacy of the data collection process. Recent methodological studies have also begun to examine the effect of computer-assisted data collection techniques on reporting of sensitive behaviors.

Increasing the privacy of data collection is the approach most widely believed to improve the accuracy of the answers to sensitive questions. 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 collection is not completely private because at least the interviewer can link the answers to the identity of the respondent who provided them; in addition, when interviewers administer the questions and record the answers, it is possible for other household members to overhear what the respondent is reporting. Surveys that employ SAQs (in which respondents record their answers without the mediation of an interviewer) help overcome these threats to confidentiality.

There is considerable evidence that self-administered questionnaires produce higher levels of reporting of sensitive behaviors than do face to face interviews (Bradburn 1983). The advantages of SAQs have been demonstrated for a number of sensitive topics, including sexual behavior (Boekeloo, Schiavo, Rabin, Conlon, Jordan, and Mundt 1994), illicit drug use (Aquilino 1994; Aquilino and LoSciuto 1990; Schober, Caces, Pergamit, and Branden 1992; and Turner, Lessler, and DeVore 1992), alcohol consumption (Aquilino and LoSciuto 1990; Hochstim 1967), and abortion reporting (London and Williams 1990; Mosher and Duffer 1994; Mott 1985).

The most consistent findings have shown that more respondents admit to illicit drug use under self administration than under other methods of data collection. In an early comparison, Aquilino and LoSciuto (1990) found substantially higher reporting of both drinking and drug use with self-administered questions than with questions administered by an interviewer over the telephone. Three subsequent comparisons between face to face interviewing and self-administered questionnaires also found greater reporting of cocaine and marijuana use when the questions were self-administered than when they were administered by interviewers in a face to face setting (Aquilino 1994; Schober et al. 1992; Turner et al. 1992). Generally, the differences by mode are most marked with recent drug use (that is, use of drugs in the month preceding the survey) and with "harder" drugs (such as cocaine). The results on reported alcohol consumption generally parallel those for illicit drugs (see, for example, Hochstim 1967).

Self administration also appears to reduce survey respondents' reluctance to admit that they have had an abortion (London and Williams 1990; Mosher and Duffer 1994; Mott 1985), and self administration also apparently increases respondents' candor in reporting on other sexual topics. Boekeloo and his colleagues report that respondents were more likely to admit to unprotected sexual intercourse and a history of sexually transmitted diseases when the questions were self-administered than when they were administered by an interviewer (Boekeloo et al. 1994).

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 which question the respondent answers (e.g., the respondents spin a dial to determine which of two questions they are 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, Greenberg, and Horvitz 1970; I-Cheng, Chow, and Rider 1972; Shimizu and Bonham 1978). However, the procedure is difficult to use in a large survey and 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, although the results

for these variables are not so well documented as those for self administration (Bradburn 1983). Question format refers to whether the item includes prespecified response categories (closed format) or the respondent simply generates his or her own answer (open format). Whether the items use an open or closed response format can have a substantial effect on the reporting of sensitive behaviors. For example, Bradburn, Sudman, and their colleagues (1979) found that, compared to closed questions, open questions produced increases in reporting that ranged from 14 per cent for frequency of sexual intercourse during the past month to 108 per cent for frequency of masturbation. Across a number of items about sensitive behaviors, the average increase in the level of reporting was 52 per cent for the open as compared to the closed versions of the questions.

Longer questions may also yield fuller reporting. Reports about the frequency of behavior or the amount of alcohol consumption are subject to memory errors even when there are no motivational obstacles to truthful reporting (see Jobe, Tourangeau, and Smith 1993, for a review). Particularly if the behaviors are frequent and specific episodes 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 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 his 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, or CASI. CASI and newer computer-assisted methods of data collection appear to combine the advantages of self administration with those of computerization. A study by Waterton and Duffy (1984) found that a computer-administered questionnaire produced greater reports of alcohol consumption than a conventional face to face interview. Their study does not permit separate estimates of the effects of computer assistance and self administration, a problem shared with several other early studies on CASI (Locke et al. 1992; Lucas et al. 1977; Robinson and West 1992). However, two recent studies suggest that audio-CASI, in which the computer program presents the questions to respondents both visually on the screen of the laptop and aurally via earphones, may produce gains relative to other methods of self administration (O'Reilly, Hubbard, Lessler, Biemer, and Turner 1994; Tourangeau and Smith 1996). These findings suggest that either auditory presentation or computerization may improve reporting (at least when the questions are self-administered).

There is also some evidence from comparisons of computer-assisted personal interviewing (CAPI) with paper-and-pencil interviews that computer assistance by itself can enhance the reporting of sensitive behaviors. Baker, Bradburn, and Johnson (1995) 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.

1.2. Variables for this study

The major aim of this study was to identify methods that could be used to improve the accuracy of data collected in the National Survey of Family Growth (NSFG). Since its beginning in 1971, the NSFG has obtained detailed information on fertility and reproductive health.

In each of its five cycles, the NSFG has explored a broad range of sensitive questions, concerning topics such as contraceptive practices, pregnancy histories (including fetal and infant deaths), unplanned and unwanted pregnancies, sexually transmitted diseases, and infertility. Despite the increasingly intimate information being sought in the NSFG, response rates have remained high. Around 80 per cent of the cases selected for the NSFG complete the interview and only about one-third of the nonrespondents are outright refusals (Rieger, Judkins, and Sperry 1991).

The study examined three strategies for improving the accuracy of self-reports. The first strategy – increasing the privacy of the data collection process – was already well-established in the survey methods literature. We nonetheless chose this strategy because we believed that privacy was the single most powerful variable affecting reporting on sensitive topics and we believed that it would have the largest effect on the survey estimates. More specifically, we decided to test the improvements produced by self administration; we ruled out 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, we sought to test the effect 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 (see Aquilino 1993, for an exception). As a result, the effects of the privacy of the setting in which the interview is carried out are unclear. We 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 we examined was that of placing the interview as a whole and, especially the questions on abortion, in a medical context. The ultimate purpose of collecting data on topics like sexual behavior and risky behaviors is to facilitate health planning; a medical context for the interview might remind respondents of this underlying aim. In addition, we believed that respondents might be more accustomed to providing candid answers in the setting of a medical interview than in the less familiar setting of a survey interview. We attempted to foster a medical context in two ways. First, we used nurses and nursing assistants as interviewers for some of the respondents; these interviewers identified themselves to the respondents as nurses at the outset of the interviews. We know of little prior work investigating this approach for collecting sensitive data and sought to test its effectiveness in this study. Second, for some of the respondents, we attempted to foster a medical context by beginning the interview with a series of questions about medical conditions and procedures.

The final strategy we investigated 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, we hypothesized that computerization of the data collection process might increase respondent's willingness to report truthfully.

2. Methods

We conducted a large-scale field experiment 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, where they had 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. The experiment varied five factors: 1) whether the questionnaire began with a series of questions on a range of medical procedures and conditions or with a series of questions about pregnancy; 2) whether the interview was conducted by a nurse or by a regular NORC interviewer; 3) whether the interview was done at the respondent's home or at a site outside the home; 4) whether the interviewer asked the questions or the questions were self-administered; and 5) whether the data were collected via computer or on paper. The analyses examined a number of outcome variables, including the unit and item nonresponse 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 mainly on reports about sexual behavior. (A more detailed discussion of abortion reporting is given in Jobe, Pratt, Tourangeau, Baldwin, and Rasinski 1997.)

2.1. 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, 85 area segments were drawn. Each segment consisted of a single block or group of adjoining blocks, defined using data from the 1990 Census. The segments were selected systematically, with selection probabilities for each segment proportional to the 1990 Census count of its housing units. In the second stage of selection, a subsample of dwellings in the sample segments were designated for the screening interview. A total of 6,325 occupied dwellings were selected for screening and screening interviews were completed at 4,659 of them, for a response rate of 73.7 per cent. The screening data included information on 3,141 persons 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. The time frame was originally defined so that no one would be selected who had had an abortion during the three-month period prior to the beginning of data collection for the experiment. However, because the field period for the experiment was delayed, this window of eligibility in fact ended more than nine months before the experiment began. Like the area probability sample, the clinic sample was restricted to residents of the city of Chicago; the eligible age range, however, was expanded slightly to include women between the ages of 15 through 40. The two clinics provided a total of 1,088 names.

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 only he was aware of the sample from which the individual cases had been selected. In

addition, we used an after-the-fact permission form procedure in which women who completed the interview were asked to sign a release form giving us access to their medical records. 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. (These procedures were approved by the Institutional Review Boards at NCHS and at one of the clinics.) A total of 54 members of the clinic sample were dropped because of problems in obtaining signed permission forms.

Selection of cases for the experiment. The area probability and clinic samples included more than 4,200 persons eligible for the experiment. We selected a subsample of 2,266 of these and randomly assigned them to a treatment cell. Within the area probability sample, the selection of persons was carried out within six strata defined by sex, age (15 through 19 vs 20 and over), and minority group membership (African-American and Hispanic households vs all others). Within a household, no more than one eligible person was selected for the main study; households with eligible members in multiple strata were randomly assigned to a single stratum. Women in the clinic sample were also stratified by age category and minority status and a random sample was selected from each group. In total, 732 women were selected from the clinic sample for the experiment. Table 1 shows the number of cases selected for the experiment by source and stratum.

2.2. Data collection

A few cases turned out to be ineligible: Some of the names provided by the clinics were duplicates; some persons selected from the area probability sample fell outside the desired age range; 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. Data collection took place over a five-month period, beginning May 24, 1993 and ending on September 26.

After the sample was fielded, it became necessary to subsample males as a cost-saving measure; ultimately, only 100 men were interviewed. The overall response rate for women was 55.2 per cent. Although this rate is much lower than the response rates attained in most high quality national studies, our sample was restricted to a single large city and the final response rate is comparable to the rates national studies typically achieve within major metropolitan areas. For example, in the Cycle IV of the NSFG, the response rate in the ten largest metropolitan areas was below 60 per cent (Rieger, Judkins, and Sperry

Table 1. Initial sample sizes

Stratum	Source		Total
	Area	Clinic	
Younger minority women (15–19 years old)	237	82	319
Older minority women (20 or older)	549	398	947
Other younger women (15–19 years old)	52	22	74
Other older women (20 or older)	372	230	602
Younger men (18–19 years old)	29	–	29
Older men (20 or older)	325	–	325
Total	1,564	732	2,296

1991). Of the 354 completed clinic cases, the data for 54 were dropped – 48 women refused to sign permission forms and forms for six others were lost.

2.3. *Experimental design*

We manipulated five variables in this experiment, in a completely crossed design. The first two experimental factors – *interviewing staff* and *version of the questionnaire* – were attempts to create a medical context for the interview. We tested the hypothesis that respondents would answer the questions more accurately if nurses or nursing assistants rather than regular field interviewers conducted the interviews and if the questionnaire began with items that made the health-related purposes of the study more salient. Accordingly, we varied the interviewing staff, comparing the data collected by 17 nurses and nursing assistants with those collected by 23 regular NORC field interviewers. The nurse interviewers received the same training as new field interviewers – a day of general orientation plus two days of study-specific training. (Experienced field interviewers attended only the two-day study-specific training.) We also compared two versions of the questionnaire. The two versions included identical items, but varied the order in which two sets of abortion questions appeared. In one version, the questionnaire first asked a series of questions about the respondent's pregnancies; in the other, it first asked a set of questions about a list of medical conditions and procedures. In the version that began with the pregnancy history questions, the topic of abortion was initially introduced in a follow-up item asking about the outcome of each reported pregnancy; in the version that began with the items on medical conditions, the topic was first introduced in connection with a number of medical procedures affecting reproduction. The procedures listed included several methods for inducing an abortion.

The next two experimental factors involved the mode of data collection. The experiment varied whether the interview was *computer-assisted* and whether the questions were *self-administered*. These two factors were crossed, yielding four data collection 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). To our knowledge, this is the first study that manipulated computerization and self administration independently. The design allows us to test the hypotheses that respondents in both self-administered conditions (those completing the SAQ or CASI questionnaires) would report higher levels of sensitive behaviors and that computerization would also yield increases in reporting.

Finally, the design varied the *site* of data collection, comparing interviews conducted in the respondent's home with those done at a site outside the home. We tested the hypothesis that levels of reporting would be higher in the site interviews, where other members of the household could not overhear the answers. Many sites were used for the interviews conducted outside the home, with NORC offices and neighborhood restaurants being used the most frequently. In an attempt to keep the response rates roughly equal across the two site conditions, we offered sample members assigned to an interview outside their homes a 40 USD incentive. The final experimental variable is, thus, actually a composite that reflects both the level of payment offered (0 USD versus 40 USD) and the site of the interview (at the respondent's home or outside the home). Although we will sometimes

refer to the "site" variable, this is merely a shorthand way of describing a comparison involving differences in both the site of interview and the level of incentive offered.

Each interviewer was to collect data under conditions representing every combination of the experimental factors (except the interviewing staff variable); thus, each nurse and each regular field interviewer was to collect data using both computer and paper questionnaires, in the respondent's home and outside the home, with both versions of the questionnaire, and administering the questions to some respondents and overseeing self administration with others.

2.4. *Questionnaires*

At the beginning of each interview, the respondent was asked to note three or four important personal events on a calendar to help date the events reported later in the questionnaire (cf., Loftus and Marburger 1983). Both versions of the questionnaire proper began with a series of demographic questions. These were followed either by the medical procedures questions or by the questions on pregnancy history; both sets of questions included the key items on abortion. The pregnancy history questions asked the respondent to list all of her pregnancies in order and to report the outcome of each one (i.e., live birth, stillbirth, ectopic pregnancy, miscarriage, or abortion). The medical procedures questions asked the respondent whether she had undergone any of a number of gynecological procedures. Six of the procedures listed were methods for inducing an abortion. The two versions of the questionnaire differed only in the order of these two sets of questions. In one version, the medical condition questions came before the pregnancy history items; in the other version the pregnancy history questions preceded the medical condition questions.

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; other items asked about the number of sex partners during the previous year, the previous five years, and in total. The questionnaires also contained items on whether respondents had had a sexually transmitted disease. 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 several items on illicit drug use. The initial drug question asked whether the respondent had ever used any illegal drug, and follow-up questions asked about their use of a number of specific drugs.

3. Results

Our discussion of the results examines the effect of the experimental variables on both nonresponse and response errors. More specifically, we examined both unit and item nonresponse rates, the average number of sexual partners reported, and reports about other sensitive behaviors as a function of the experimental variables. In a final set of analyses, we also examined interviewer effects on responses to key questionnaire items. Because so few men completed the interview, we mainly discuss the results for the women; however, a few analyses compare reports by men and women.

The hypotheses tested were that

- Self administration and conducting the interviews outside the respondent's home would increase the proportion of women reporting sensitive behaviors (including abortions and sexually transmitted diseases) and would increase the average number of sexual partners reported by women;
- The use of nurses as interviewers and the version of the questionnaire that began with the list of medical conditions would increase the proportion of women reporting abortions and raise the level of reporting of sensitive behaviors more generally;
- The apparent legitimacy and objectivity conferred by computerization would increase the level of reporting of sensitive behaviors.

We also examined the effects of the experimental variables on both unit and item nonresponse. We were concerned that the combination of computerization and self administration would affect overall response rates, because some respondents might be reluctant to interact with the computer directly (Couper and Rowe 1996). Accordingly, we tested the hypothesis that

- Computerization and self administration would interact such that cases assigned to computer-assisted self-administered interviews would have the lowest response rate among the four data collection groups.

In addition, we tested two hypotheses regarding item nonresponse

- Self administration would increase the proportion of data that were missing relative to administration of the items by interviewers;
- Computerization would reduce the rate of missing data relative that of paper-and-pencil questionnaires.

Finally, we tested two hypotheses regarding interviewer effects

- Both computerization and self administration of the questions would reduce the effect of the interviewers on the data obtained.

The data from the experiment were geographically clustered, and the analyses reported here take this clustering into account. The analyses treat the sample as if it included 38 strata, each consisting of respondents from two or three nearby area segments. In addition, the two abortion clinics were treated as two clusters from a final stratum. Unless we clearly state otherwise, we used SUDAAN to estimate standard errors and carry out significance tests. SUDAAN uses Taylor Series methods to produce approximate standard errors (Shah, Barnwell, Hunt, and LaVange 1993).

3.1. Nonresponse

Although our main interest was in the effects of the experimental variables on answers to the questions, we also examined their effects on unit and item nonresponse. The analysis of nonresponse is important for several reasons. First, as Catania and his colleagues have argued, the prevalence of AIDS risk behaviors may be underestimated because those persons who are most at risk (such as injection drug users) may also be the least willing to take part in the survey at all or to answer the key questions on risky behaviors (Catania et al.

1990). More generally, nonresponse bias may contribute substantially to errors in surveys on sexual behaviors or other sensitive topics. Thus, if any of the experimental variables produced marked gains in response rates, this would represent an important practical finding. A second reason for analyzing nonresponse is to determine whether it produced differences between the experimental groups. Because random assignment was used, the different experimental groups begin as comparable subsamples; nonresponse may, however, introduce differences in the composition of the groups. Any resulting differences in the characteristics of the respondents in the different groups might account for any differences in levels of reporting. A final reason for examining nonresponse is to determine the effect of different modes of data collection on the level of missing data within completed questionnaires. The analyses tested the hypotheses that computerization would reduce the level of item nonresponse and that self administration would increase it.

Unit nonresponse. How did the experimental variables affect the levels of overall nonresponse? Table 2 displays the response, contact, and cooperation rates for each level of the five experimental variables. The response rates represent the proportion of eligible sample members who completed an interview. The contact and cooperation rates distinguish two classes of problems that can lead to nonresponse – failure to locate sample cases and failure to persuade them to take part in the study. The contact rate refers to the proportion of sample cases who were located and contacted during the field period; the cooperation rate refers to the proportion of those contacted who agreed to take part in the study. As is apparent from Table 2, failures to contact sample members accounted for the bulk of the nonresponse; more than three-fourths on the nonrespondents were never contacted during the field period. In some cases, the failure to contact a member of the sample may, in fact, have reflected covert resistance by that person.

We fit logit models to the response rate data, examining the effects of the experimental variables, including their interactions. A simple additive model incorporating only the five

Table 2. Response, contact, and cooperation rates by experimental group

Group	Response rate (<i>n</i>) (per cent)	Contact rate (<i>n</i>) (per cent)	Cooperation rate (<i>n</i>) (per cent)
Staff			
Field interviewer	60.8 (970)	71.3 (970)	85.3 (692)
Nurse	49.5 (944)	58.9 (944)	84.0 (556)
Site/incentive			
Respondent's home/0 USD	50.5 (965)	62.8 (965)	80.4 (606)
Outside the home/40 USD	60.1 (949)	67.7 (949)	88.8 (642)
Mode of data collection			
Paper	57.0 (951)	66.0 (951)	86.3 (628)
Computer	53.5 (963)	64.4 (963)	83.1 (620)
Method of administration			
Self administration	55.1 (958)	65.7 (958)	83.9 (629)
Interviewer administration	55.3 (956)	64.7 (956)	85.5 (619)
Version			
Medical items first	55.7 (961)	65.1 (961)	85.6 (626)
Pregnancy items first	54.7 (953)	65.3 (953)	83.8 (622)

main effects was rejected. The simplest model that provided an adequate fit to the data included both the interaction of the site/incentive variable with the staff variable and a three-way interaction involving staff, computer assistance, and self administration; for the overall model, $G^2 = 23.1$ ($df = 22, p > .30$). The site and staff variables also had significant relations to response rates in simple bivariate analyses. The interaction between the site and staff variables is easy to summarize. The site of the interview/incentive variable had much greater effect on the response rates of the nurses (who completed 57.4 per cent of their site cases but only 41.7 per cent of those slated for in-home interviews) than on those of the regular field interviewers (who completed 62.7 per cent of their site cases and 59.0 per cent of their in-home cases). Additional analyses indicated that the interaction between the site/incentive variable and staff reflected differences in the rate at which sample members were contacted rather than differences in the cooperation rates. The nurses contacted significantly more of their site cases (64.5 per cent) – to whom they could offer a 40 USD payment – than of their in-home cases (53.5 per cent); the regular field interviewers, by contrast, contacted similar proportions of site and in-home cases (70.8 per cent vs 71.9 per cent).³

The three-way interaction between the staff variable, computer assistance, and self administration is more complex; the relevant results are displayed in Table 3. As the table demonstrates, the hypothesis that CASI would produce the lowest response rates received some support among the cases assigned to the regular field interviewers but not for those assigned to the nurses. When the data for the two types of field staff are combined, there were no clear differences in response rates by the four methods of data collection in the experiment. Within each group of interviewers, however, differences by mode of data collection were apparent. For the regular field interviewers, the overall response rates were lowest for sample cases assigned to CASI data collection (54.8 per cent response rate). For the nurses, the pattern across the four methods of data collection was quite different, with the *highest* response rates achieved among the CASI cases (53.1 per cent). Additional analyses indicate that the differences in response rates apparent in Table 3 were the product of differences in the levels of cooperation by staff and mode of data collection.⁴

Effect of nonresponse. We examined the data for the women respondents to determine whether nonresponse introduced any differences between the groups that might affect the interpretation of differences in levels of reporting. We compared the four data collection groups (CAPI, PAPI, CASI, and SAQ) on the distribution of respondents by stratum, race, age (under 20, 20–29, and 30 and older), source (clinic vs area probability sample), and marital status. None of the background differences across data collection groups was significant, even by a liberal criterion ($p < .20$). Similarly, the site/incentive variable (which had significantly affected the overall response rate) did not affect the composition of the

³ In a logit analysis of the contact rates, the only significant effects to emerge were the main effect of the staff variable and the staff by site interaction. Overall, the nurses contacted only 58.9 per cent of their assigned cases versus 71.3 per cent for the field interviewers.

⁴ Logit analyses of the cooperation rates revealed only two effects. The first was the three-way interaction apparent in the final column of Table 3 (with the regular field interviewers achieving their lowest rates of cooperation under CASI and PAPI data collection, and the nurses achieving their highest rates under these two modes). The other was a main effect for the site variable (with a cooperation rate of 88.8 per cent among those assigned to be interviewed outside the home versus 80.4 per cent among those assigned to be interviewed at home; see Table 2). The 40 USD incentive offered to cases slated for an interview at a site outside their homes more than compensated for the greater inconvenience of these interviews.

Table 3. Response and cooperation rates by staff and mode of data collection

Group	Response rate (n) (per cent)	Cooperation rate (n) (per cent)
All cases		
PAPI	57.8 (472)	86.9 (314)
SAQ	56.2 (479)	85.7 (314)
CAPI	52.9 (484)	83.9 (305)
CASI	54.1 (479)	82.2 (315)
Field interviewers		
PAPI	62.4 (242)	85.3 (177)
SAQ	67.5 (234)	90.3 (175)
CAPI	61.0 (241)	88.0 (167)
CASI	54.8 (239)	79.4 (165)
Nurses		
PAPI	53.0 (230)	89.1 (137)
SAQ	45.3 (245)	79.9 (139)
CAPI	44.9 (243)	79.0 (138)
CASI	53.1 (239)	85.2 (149)

samples; for example, roughly the same proportion of in-home respondents and neutral site respondents were married.

We did find that the two types of interviewers completed interviews with somewhat different types of respondents. The nurses interviewed a higher proportion of younger, non-minority women (the third stratum listed in Table 1) than the regular field interviewers ($\chi^2_3 = 14.3$, $p < .03$). This was the only instance in which an experimental variable was related to a difference in the characteristics of the respondents.⁵

Item nonresponse. The various forms of unit nonresponse seem to be relatively unaffected by the method of data collection, with neither self administration nor computerization having much effect on response rates. Of the experimental factors, the only one with a consistent effect on response rates was the site variable, a finding that probably reflects the effect of the 40 USD incentive given to respondents interviewed outside their homes.

The results change considerably when we examine item nonresponse. We analyzed responses to 43 of the questionnaire items administered to the women respondents. These items covered a range of topics – demographics, sexually transmitted diseases, pregnancy history, medical procedures for inducing abortion, illicit drug use, and sexual behavior. Because of the skip patterns in the questionnaire, few respondents were actually supposed

⁵ We also examined a final form of nonresponse – refusal to sign the permission forms giving access to medical records. Overall, 87.0 per cent of the women who completed the interview also signed permission forms. The permission form refusal rates varied significantly by only one of the five experimental factors, the site of the interview/level of incentive variable. Of the women interviewed at a site outside their homes, 90.9 per cent signed a permission form; of those interviewed at home, only 82.5 per cent signed the form ($\chi^2_1 = 16.2$; $p < .001$). Logit analyses indicate that none of the other main effects or interactions significantly affected the rate at which respondents signed permission forms. The incentive offered to those interviewed outside their homes may have increased their willingness to sign the form. Or, having taken the extra trouble to go to a site outside the home, respondents may have been more inclined to comply with the interviewer's final request. We examined the data from the area probability cases (who were retained in the sample even if they refused to sign permission forms) to determine whether permission form nonresponse affected the composition of the sample. Across a number of dimensions (including the proportion of women who reported having had an abortion), those who signed the forms did not differ significantly from those who refused to sign them.

Table 4. Proportion of questions answered by method of data collection

Group	Mean proportion (n)
All cases	
PAPI	97.8 (261)
SAQ	94.6 (256)
CAPI	99.2 (244)
CASI	97.9 (244)
Self-administered	96.2 (500)
Interviewer-administered	98.5 (505)
Computer-assisted	98.6 (488)
Paper-and-pencil	96.2 (517)
Nurses	
Self-administered	97.8 (191)
Interviewer-administered	98.0 (219)
Field interviewers	
Self-administered	95.3 (309)
Interviewer-administered	98.8 (236)

to answer all 43 of these questions; the average women respondent should have completed only 30.2 of the items. Across the experimental treatments, the women answered almost all of the items that they were supposed to – an average of 29.4 items, or about 97.4 per cent of the total.

For each of the women respondents, we calculated the proportion of the items that the respondent was supposed to answer which the respondent actually did answer; for this analysis, “don’t know” and “refused” responses were treated as missing, along with items that were inadvertently skipped. We then examined the averages of these proportions across respondents in each experimental group. At a descriptive level, the clearest findings were that both computer assistance and self administration appeared to affect the completeness of the data (see Table 4); computer assistance increased the proportion of the questions that were answered relative to paper questionnaires (98.5 per cent vs 96.2 per cent), and self administration decreased the proportion relative to administration of the questions by an interviewer (96.2 per cent vs 98.5 per cent). In an analysis that fit a simple additive model involving the five experimental factors, both of these main effects were significant. For the effect of computer assistance, $F(1, 43) = 39.9$, and for the effect of self administration, $F(1, 43) = 31.9$.

However, when we fit fully a saturated five-way model (which incorporated all the interactions among the experimental variables), both of these main effects were qualified. The effect of self administration depended on the type of interviewer – $F(1, 43) = 5.52$ for the interaction ($p < .01$). As can be seen from Table 4, the difference between self administration and interviewer administration is more apparent for the regular field interviewers than for the nurses. There is also a significant three-way interaction among computer assistance, staff, and the version of the question; $F(1, 43) = 5.01$. The means (which are not shown) indicate that, although computer assistance always improved the completeness of the data, the exact magnitude of the improvement depended on both the type of interviewer and version of the questionnaire.

The patterns apparent in Table 4 were also apparent for specific subsets of the items. When we analyzed each group of items separately, we found significant effects for computer assistance on the missing data rates for the five demographic items, the six items on sexually transmitted diseases, the six items on medical procedures used to perform abortions, the nine items on illicit drug use, and the eight items on sexual behavior. (The difference was in the same direction but not significant for the pregnancy history questions.) Similarly, the effect of self administration was significant for four of the six subgroups of items we examined. These results indicate that the effect of computer assistance and self administration on the completeness of the data did not depend on the topic of the questions; similar effects were found for relatively nonsensitive items (such as the demographic questions) and for highly sensitive ones (such as the items on sexual behavior and drug use).

3.2. Reports on sexual behaviors

Average number of sexual partners. A central variable in models of AIDS transmission is the number of sexual partners acquired over time. As we noted earlier, there are several reasons for doubting the accuracy of reports about the number of sexual partners, including the consistent difference in the average number of opposite-sex sex partners reported by men and women. Most researchers believe that this disparity results from misreporting by members of both sexes, with women underreporting their sex partners and men overreporting theirs. Tourangeau and Smith (1996) report further evidence of error in reports about sex partners; they show that respondents reporting more than ten partners tend to report their totals in round numbers.

We examined answers to the questions about the number of opposite-sex sexual partners during the past year, the past five years, and over the lifetime. Because respondents who reported that they had been sexually active only for the past year were skipped out of the five-year and lifetime partner questions, we treated their answers to the one-year item as representing their five-year and lifetime totals as well. Similarly, for respondents who reported they had been active for less than five years, we assumed that the number of partners they reported in response to the five-year item also represented their lifetime total.

For all three time periods, we examined the average number of sex partners reported as a function of the five experimental factors.⁶ For each time period, women who completed self-administered questionnaires reported more sexual partners than women who responded to questions administered by an interviewer. There were significant effects for self administration on reported partners during the past year, the past five years, and the respondent's lifetime (see Figure 1). 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 ($F(1, 39) = 8.64, p < .01$). For the five-year period, women who completed self-administered questionnaires reported a mean of 3.87 sexual partners versus 2.82 for those who answered interviewer-administered

⁶ The data on the number of reported sex partners are counts, and the distribution of the responses for all three time periods is highly skewed. To compensate for this departure from normality, we added .5 to the values and then carried out a logarithmic transformation prior to performing analyses of variance. For ease of interpretation, we report untransformed values in presenting the group means.

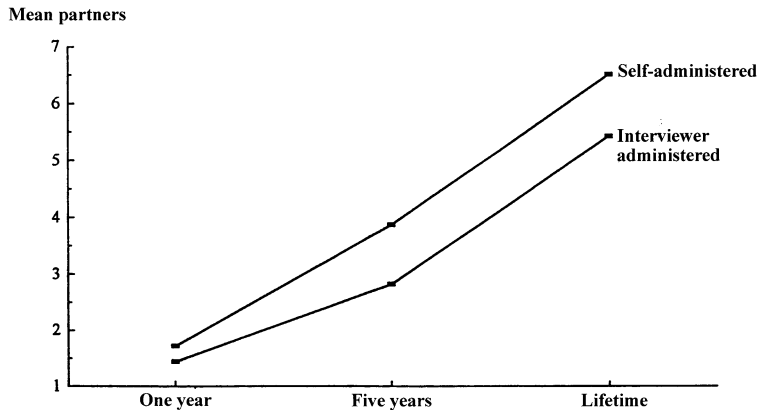


Fig. 1. Average reported sexual partners, by method of administration and time period

questions ($F(1, 39) = 5.42, p < .05$). For the lifetime item, women who completed self-administered questionnaires reported a mean of 6.51 sexual partners versus 5.43 for those who answered questions administered by an interviewer ($F(1, 39) = 9.08, p < .01$). No other main effects were significant.

Computerization interacted with the site of the interview/level of incentive to affect the number of sexual partners reported. During home interviews (with no payment), more sexual partners were reported by women interviewed using computer-assisted questionnaires than by those responding to paper-and-pencil questionnaires; for women interviewed outside the home (with a 40 USD payment), more sexual partners were reported on the pencil-and-paper questionnaires. Table 5 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 vs 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 vs 1.43; $F(1, 39) = 7.78, p < .01$). 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 vs 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 vs 5.08; $F(1, 39) = 6.16, p < .05$). The pattern is in the same direction but not significant for the five-year partners item. Overall levels of reporting are consistently higher using computer-assisted questionnaires, although not significantly so. Bringing computers into the respondents' homes may have fostered a sense of the importance or objectivity of the survey, promoting more complete reporting of sexual partners. Outside the home, especially in public places, the computer may make respondents feel conspicuous, inhibiting reporting.

Males vs females. Given the small sample size and low response rates for the men, any comparisons between the men and women must be regarded as exploratory. Thus, we present only a few findings involving the male respondents. As has been observed in earlier surveys on sexual behavior, men reported more opposite-sex sexual partners on average than women did. This was true for the past year (4.19 for the men versus 1.58 for the women), the past five years (12.47 vs 3.34), and lifetime (23.96 vs 5.97); all three

Table 5. Mean number of sexual partners reported by computer assistance and site

	At home		Outside the home	
	Paper	Computer	Paper	Computer
One year	1.36 (225)	1.84 (226)	1.68 (278)	1.43 (254)
Five years	2.81 (225)	4.53 (228)	3.33 (279)	2.74 (253)
Lifetime	5.06 (216)	7.48 (224)	6.26 (267)	5.08 (249)

Note: Means are based on untransformed counts. Sample sizes (given in parentheses) vary by row due to item nonresponse.

differences are highly significant (*F* values all greater than ten; *p* values all less than .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. However, the men did report (non-significantly) fewer lifetime sex partners when the questions were self-administered (a mean of 22.8) than when they were administered by an interviewer (25.0). This difference in the effect of self administration on the reports of men and women is even clearer in a study by Tourangeau and Smith (1996) that had a larger, more representative sample of men; that study found that self administration increased the average number of sexual partners reported by women and decreased the average number reported by men for the prior year, the prior five years, and over the lifetime.

Rounding of values. Morris (1993) has suggested that problems in reports about sex partners may be confined to the subgroup of respondents who report a large number of partners. Within this subgroup, most respondents report numbers of partners that are round values – that is, exact multiples of five. Figure 2 shows the distribution of the number of lifetime sexual partners reported by the men in our study. The results for the women, which are not shown, display similar heaping at round values. Of the 194 respondents of both sexes who reported ten or more sexual partners, 62.4 per cent gave an answer that was an exact multiple of five.

We carried out logit analyses on the proportion of the respondents reporting more than ten lifetime sexual partners who reported a value that was an exact multiple of five. An

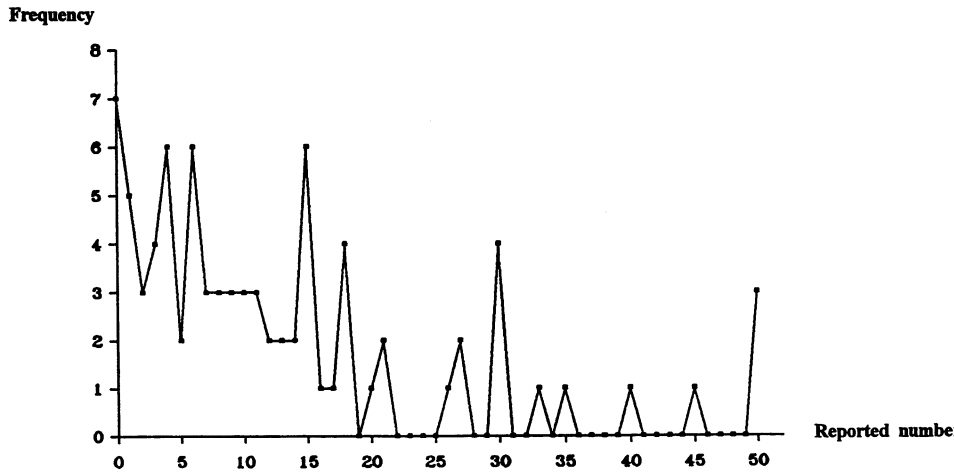


Fig. 2. Distribution of reported lifetime sexual partners: male respondents

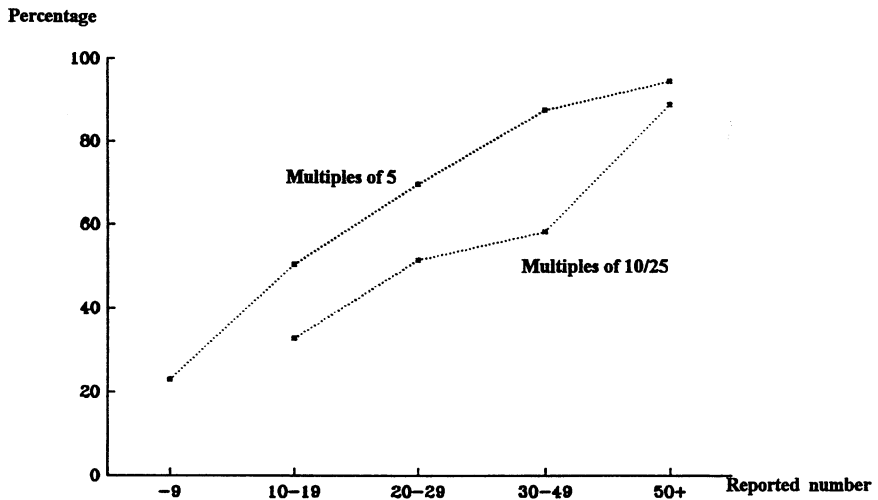


Fig. 3. Percentage of round values by reported number of lifetime sexual partners: male and female respondents

additive model that included the five experimental variables and the sex of the respondent revealed only one noteworthy effect – a higher proportion of respondents reported round values when data collection was computer-assisted than when it was done via paper-and-pencil (70.2 per cent vs 56.4 per cent; $z = 2.62$). Men and women did not differ markedly in their apparent propensity to use round values: 56.3 per cent of the 48 men and 64.4 per cent of the 146 women who reported ten or more lifetime sexual partners used a multiple of five to report their lifetime totals. The main determinant of the use of round values seemed to be the sheer number of partners reported. The larger the number, the more likely it was to be a multiple of five (see Figure 3).

Figure 3 also suggests that as the number of partners reported becomes larger, respondents switch to a less exact rounding rule; their answers appear to be rounded to multiples of ten or 25 rather than to other multiples of five. Among respondents reporting 50 or more partners, 83.3 per cent report values that are exact multiples of ten or 25.

3.4. Abortion and other sensitive topics

Besides the number of sexual partners, the questionnaire collected data on other topics related to sexual behavior, including sexually transmitted diseases, condom use, and abortion. In addition, the questionnaire included items on illicit drug use.

Sexually transmitted diseases and condom use. Women who answered self-administered questions reported more sexually transmitted diseases (STDs) than those answering questions administered by an interviewer (22.0 per cent vs 17.0 per cent). This effect of self administration was only marginally significant ($\chi^2_1 = 2.93$, $p < .10$). No other main effects or interactions were significant.

We analyzed the ratio between two items concerning condom use and sexual intercourse in the last 30 days; the ratio represented the proportion of time the respondent used a condom in the past month. We performed an analysis of variance on this index. We dropped women who reported that they had not had sexual intercourse in the last

30 days from this analysis, leaving a total of 641 women. Significantly more condom use was reported with self-administered questionnaires (average reported use 47 per cent of the time) than with interviewer-administered questions (35 per cent). The main effect for self administration was significant ($F(1,39) = 7.89, p < .001$). We found no other significant effects on the condom use variable.⁷

Abortion reporting. Surprisingly, none of the experimental variables had an effect on the level of abortion reporting, either with the traditional NSFG items or the new medical procedure items. The results were similar whether the analysis examined all of the women in the sample or just those who were known to have had an abortion. About 74 per cent of the women selected from clinic records admitted that they had had an abortion during their lifetime; however, only 51.7 per cent of the clinic sample reported an abortion during the specific period during which, according to the clinic records, they had had an abortion. This latter figure is quite consistent with the findings reported by Jones and Forrest (1992) in their comparison of the NSFG estimates with reports from abortion providers. (For a more detailed exploration of the results on abortion reporting, see Jobe et al. 1997).

Illicit drug use. A substantial minority of the sample – 42.7 per cent – admitted that they had used illicit drugs. Of those reporting drug use, the vast majority had used marijuana (97.2 per cent of those admitting any drug use) or cocaine (31.9 per cent of those admitting drug use). Logit models revealed no significant main effects of the experimental variables on rates of reported illicit drug use, but did indicate one significant interaction – between the staff variable and self administration. Whether a nurse or field interviewer was assigned to the respondent made no difference in reporting rates when the questions were self-administered, but when interviewers asked the questions, the regular field interviewers elicited higher rates of reported drug use than did the nurses (44.6 per cent to 35.9 per cent; $\chi^2_1 = 3.79, p = .05$). This effect was no longer significant when the clustering of the sample was taken into account.

3.5. Interviewer effects

One of the potential benefits of self administration is the reduction of the effect of the interviewers on the answers; when interviewers merely hand the respondent a questionnaire to be completed (or set up a laptop for self administration), they may have less effect on the responses than when they read the questions and record the answers. Similarly, computer-assisted methods of data collection may, by reducing variability in interviewer behavior, also reduce the interviewers' effect on the responses.

We sought to examine these hypotheses in some exploratory analyses. We carried out one-way analyses of variance for nineteen variables from the study. All nineteen involved sensitive items, including questions on sexual partners, sexually transmitted diseases, specific sexual practices (such as oral sex), abortion, and illicit drug use. The single factor in these analyses was the interviewer. Separate analyses were carried out for respondents interviewed in each of the four modes of data collection. Because the interviewer assign-

⁷ It is not clear whether using a condom is generally regarded as socially desirable or not. On the one hand, using a condom does represent a form of safe sex. On the other, condoms may be used more often in new or short-term sexual relationships than in long-term, monogamous ones; as a result, admitting to condom use may be embarrassing for some respondents.

Table 6. Estimates of ρ by method of data collection and variable

	PAPI	SAQ	CAPI	CASI
	$\hat{\rho}$	$\hat{\rho}$	$\hat{\rho}$	$\hat{\rho}$
Median across 19 variables	.072	.000	.016	.000
One-year sex partners	.158	.036	.000	.000
Five-year sex partners	.072	.000	.000	.000
Lifetime sex partners	.178	.000	.021	.000
Used illicit drugs	.108	.003	.002	.000
Used marijuana	.100	.000	.013	.000
Used cocaine	.100	.044	.035	.000

Note: PAPI refers to paper-and-pencil interviews, SAQ to self-administered questionnaires, CAPI to computer-assisted personal interviews, and CASI to computer-assisted self-administered interviews.

ments were geographically clustered, including sample cases drawn from several area segments and from the clinic sample, the estimates from the ANOVAs represent both interviewer effects and the effects of the overall area in which the interviewer’s cases were concentrated. These ANOVAs assume simple random sampling. The findings are, thus, suggestive at best.

A clear pattern emerged from these analyses. Using the estimated variance components from the ANOVAs, we estimated values of ρ , the intraclass correlation, for each variable and mode of data collection. The intraclass correlation represents the expected correlation between any two observations from the same interviewer. (When the estimate of the between-interviewer variance component was negative, we set $\hat{\rho}$ to zero.) Table 6 displays the median value of the ρ estimates across the nineteen variables, as well as the results for the three sex-partner items and three items on illicit drug use. The data strongly suggest that both computer assistance and self administration reduce variation across interviewers; for most variables, there is no evidence for interviewer effects when the questions are self-administered – the estimate for ρ was typically zero.

4. Discussion

Effects of self administration. The variable with the most consistent effect on the reporting of sensitive information was the method of administering the questions. Women who completed self-administered questionnaires reported more sexual partners, more sexually transmitted diseases, and greater use of condoms than those who responded to questions read by an interviewer. These findings are summarized in Table 7, 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 generally higher – as much as 41 per cent higher – when the questions are self-administered than when they are administered by an interviewer. The effects of self administration are similar for both sexes, with one notable exception – men reported fewer lifetime sexual partners, on average, in self-administered than in interviewer-administered questionnaires. This exception is consistent with the hypothesis that men exaggerate the number of their sexual partners under conditions of interviewer administration. The effect of self administration combined with the absence of effects on reporting for the site of the interview suggests that

Table 7. Reported sexual behavior

	Method of administration		Ratio
	Self administration	Interviewer administration	
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

respondents may be more concerned about the reactions of the interviewer than about the threat of other family members overhearing their answers.

Self administration had one additional positive effect on the quality of the data – it appeared to reduce the effect of the interviewers on the answers to the questions. Although there were few discernible differences between the nurses and field interviewers, analyses across all the interviewers revealed that self-administered questions lowered the interviewers’ contribution to overall variance. This finding is consistent with earlier results reported by the Census Bureau (U.S. Bureau of the Census 1979).

Self administration did nevertheless have one clear drawback – it increased the rate of missing data by a little more than 2 per cent (see Table 4). Even with a computer-assisted questionnaire, respondents skipped more questions when they read the questions and entered the answers than when the interviewers did.

Computerization. We found that 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 5). In reports on sexual partners, computer assistance appeared to increase the number of partners reported when data collection took place in the home, but to reduce the number reported when data collection took place outside the home. We have no compelling explanation for this mode by site interaction.

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, Bradburn, and Johnson 1995). Several other studies have shown effects on reporting for computer-assisted self administration, but in these studies, it is impossible to separate the effects of computerization from those of self administration (e.g., Waterton and Duffy 1984). We suspect that by itself computerization

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). We were, however, able to document two clear benefits from computerization. First, the computer-assisted interviews obtained substantive responses for an additional 2 per cent of the questions on average. (By contrast, computer assistance had no effect on unit nonresponse.) Second, computer assistance appeared to reduce variation in the results across interviewers (Table 6).

Effects of site/incentive and medical context. In contrast to the relatively clear results for self administration, we observed few effects for the site of the interview/level of payment variable. Similarly, we observed no effects for either of our attempts to induce a medical context for the questions.

Several studies have attempted to observe the effect 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; Lessler 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. The effects of this variable may depend on other aspects of the situation. For example, when other family members are already aware of the respondent's sensitive behaviors, their presence may make it more difficult for the respondent to deny these behaviors in the interview than when the respondent is alone. In addition, respondents may generally be worried less about the reactions of other household members than about those of the interviewer. Finally, some respondents live alone or with others (such as infants) whose presence is not a cause for concern. Each of these factors will reduce the effect of the site of the interview and make it difficult to demonstrate the effects of this variable. As a practical matter, the use of restaurants and other convenient locations in our study may have reduced the effect of moving the interview outside the respondent's home; the in-home and outside-the-home interviews may not have differed as much in privacy as we would have liked.

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 a self-administered questionnaire than they do in interviews conducted by medical personnel (see, for example, 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.

Errors in reports about sexual behavior. Unfortunately, even when the questions are self-administered the evidence indicates that there is still considerable error in reports about sexual topics. We found the usual underreporting of abortions; about half of the women in the clinic sample did not admit they had had an abortion in the year prior to the survey and the level of misreporting was not markedly affected by self administration. In addition, self administration did not eliminate the gap between the number of sexual

partners reported by men and women or the prevalence of round values in the numbers reported by members of both sexes. We suspect that there are several sources – both cognitive and motivational – contributing to these reporting problems; it will, no doubt, take a combination of methods to reduce reporting errors about sexual topics further.

5. References

- Abernathy, J., Greenberg, B., and Horvitz, D. (1970). Estimates of Induced Abortion in Urban North Carolina. *Demography*, 7, 19–29.
- Aquilino, W. (1993). Effects of Spouse Presence During the Interview on Survey Responses Concerning Marriage. *Public Opinion Quarterly*, 57, 358–376.
- Aquilino, W. (1994). Interview Mode Effects in Surveys of Drug and Alcohol Use. *Public Opinion Quarterly*, 58, 210–240.
- Aquilino, W. and LoSciuto, L. (1990). Effect of Interview Mode on Self-reported Drug Use. *Public Opinion Quarterly*, 54, 362–395.
- Baker, R.P., Bradburn, N.M., and Johnson, R. (1995). Computer-assisted Personal Interviewing: An Experimental Evaluation of Data Quality and Survey Costs. *Journal of Official Statistics*, 11, 415–431.
- Boekeloo, B., Schiavo, L., Rabin, D., Conlon, R., Jordan, C., and Mundt, D. (1994). Self-reports of HIV Risk Factors at a Sexually Transmitted Disease Clinic: Audio vs Written Questionnaires. *American Journal of Public Health*, 84, 754–760.
- Bradburn, N.M. (1983). Response Effects. In *Handbook of Survey Research*, P. Rossi, J. Wright, and A. Anderson (eds.). New York: Academic Press, 289–328.
- Bradburn, N., Sudman, S., and Associates. (1979). *Improving Interview Method and Questionnaire Design: Response Effects to Threatening Questions in Survey Research*. San Francisco: Jossey Bass.
- Catania, J., Gibson, D., Chitwood, D., and Coates, T. (1990). Methodological Problems in AIDS Behavioral Research: Influences on Measurement Error and Participation Bias in Studies of Sexual Behavior. *Psychological Bulletin*, 108, 339–362.
- Couper, M.P. and Rowe, B. (1996). Computer-Assisted Self-Interviews. *Public Opinion Quarterly*, 60, 89–105.
- Groves, R. and Mathiowetz, N. (1987). A Comparison of CATI and Non-CATI Questionnaires. In *An Experimental Comparison of Telephone and Health Interview Surveys*, O. Thornberry (ed.). Vital and Health Statistics, Series 2, 106, 33–39.
- Hochstim, J. (1967). A Critical Comparison of Three Strategies of Collecting Data from Households. *Journal of the American Statistical Association*, 62, 976–989.
- I-Cheng, C., Chow, L.P., and Rider, R.V. (1972). *The Randomized Response Techniques as Used in the Taiwan Outcome of Pregnancy Study*. New York: A Publication of the Populations Council 33, No. 11: 265–269.
- Jobe, J.B., Pratt, W.F., Tourangeau, R., Baldwin, A., and Rasinski, K. (1997). Effects of Interview Mode on Sensitive Questions in a Fertility Survey. In *Survey Measurement and Process Quality*, L. Lyberg et al. (eds.), New York: Wiley, 311–330.
- Jobe, J.B., Tourangeau, R., and Smith, A.F. (1993). Contributions of Survey Research to the Understanding of Memory. *Applied Cognitive Psychology*, 7, 567–584.
- Jones, E. and Forrest, J. (1992). Underreporting of Abortions in Surveys of U.S. Women: 1976 to 1988. *Demography*, 29, 113–126.

- 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. (1992). Computer-based Interview for Screening Blood-donors for Risk of HIV Transmission. *Journal of the American Medical Association*, 29, 1301–1305.
- Loftus, E.F. and Marburger, W. (1983). Since the Eruption of Mt. St. Helens, Has Anyone Beaten You Up? Improving the Accuracy of Retrospective Reports with Landmark Events. *Memory and Cognition*, 11, 114–120.
- London, K. and Williams, L. (1990). 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.
- Lucas, R.W., Mullen, P.J., Luna, C.B.X., and McInroy, D.C. (1977). Psychiatrist and Computer Interrogators of Patients with Alcohol-Related Illnesses: A Comparison. *British Journal of Psychiatry*, 131, 160–167.
- Marquis, K. and Cannell, C. (1971). Effect of Some Experimental Techniques on Reporting in the Health Interview. Vital and Health Statistics, National Center for Health Statistics, DHEW Publication No. 1000, Series 2, No. 41. Washington, D.C.: U.S. Government Printing Office.
- Morris, M. (1993). Telling Tails Explain the Discrepancy in Sexual Partner Reports. *Nature*, 365, 437–440.
- Mosher, W.D. and Duffer, A.P., Jr. (1994). Experiments in Survey Data Collection: The National Survey of Family Growth Pretest. Paper presented at the meeting of the Population Association of America, Miami, May.
- Mott, F. (1985). Evaluation of Fertility Data and Preliminary Analytic Results from the 1983 Survey of the National Longitudinal Surveys of Work Experience of Youth. A report to the National Institute of Child Health and Human Development by the Center for Human Resources Research, January.
- O'Reilly, J., Hubbard, M., Lessler, J., Biemer, P., and Turner, C. (1994). Audio and Video Computer Assisted Self-Interviewing: Preliminary Tests of New Technology for Data Collection. *Journal of Official Statistics*, 10, 197–214.
- Rieger, S., Judkins, D., and Sperry, S. (1991). National Survey of Family Growth: Cycle IV CATI Phase. Final Report, Hyattsville, MD: National Center for Health Statistics.
- Robinson, R. and West, R. (1992). A Comparison of Computer and Questionnaire Methods of History-Taking in a Genito-Urinary Clinic. *Psychology and Health*, 6, 77–84.
- Schober, S., Caces, M.F., Pergamit, M., and Branden, L. (1992). Effects of Mode of Administration on Reporting of Drug Use in the National Longitudinal Survey. In *Survey Measurement of Drug Use: Methodological Studies*, C.F. Turner, J.T. Lessler, and J.C. Gfroerer (eds.), Rockville, MD: National Institute on Drug Abuse, 267–276.
- Shah, B., Barnwell, B., Hunt, N., and LaVange, L. (1993). SUDAAN User's Manual, Release 6.34. Research Triangle Park, NC: Research Triangle Institute.
- Shimizu, I. and Bonham, G. (1978). Randomized Response Technique in a National Survey. *Journal of the American Statistical Association*, 73, 35–39.
- Smith, T. (1992). Discrepancies Between Men and Women in Reporting Number of Sexual Partners: A Summary from Four Countries. *Social Biology*, 39, 203–211.
- Tourangeau, R. and Smith, T. (1996). Asking Sensitive Questions: The Impact of Data

- Collection Mode, Question Format, and Question Context. *Public Opinion Quarterly*, 60, 275–304.
- Turner, C.F., Lessler, J.T., and Devore, J. W. (1992). Effects of Mode of Administration and Wording on Reporting of Drug Use. In *Survey Measurement of Drug Use: Methodological Studies*, C.F. Turner, J.T. Lessler, and J.C. Gfroerer (eds.), Rockville, MD: National Institute on Drug Abuse, 177–220.
- U.S. Bureau of the Census (1979). Enumerator Variance in the 1970 Census. PHC(E)-13, U.S. Department of Commerce.
- Warner, S. (1965). Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias. *Journal of the American Statistical Association*, 60, 63–69.
- Waterton, J. and Duffy, J. (1984). A Comparison of Computer Interviewing Techniques and Traditional Methods for the Collection of Self-Report Alcohol Consumption Data in a Field Survey. *International Statistical Review*, 52, 173–182.

Received March 1996

Revised January 1997