

ORIGINAL ARTICLE

Reliability of Adolescents' Self-reported Sexual Behavior: A Comparison of Two Diary Methodologies

ALEXANDRA M. MINNIS, M.P.H., AND NANCY S. PADIAN, Ph.D.

Purpose: To evaluate techniques for measuring high-risk sexual behaviors by comparing the reliability and acceptability of two daily sexual behavior diary modes: a written calendar and an automated telephone interview.

Methods: This randomized controlled study included 105 sexually active female adolescents aged 15-19 years recruited from among teens seeking reproductive health care services at a family planning clinic in the San Francisco Bay Area. Participants completed a standardized sexual behavior questionnaire each day for 4 weeks. Contraceptive use by method type was recorded. Reporting differences between the two diary modes were assessed using generalized estimating equations, concordance of diary and retrospective interview responses was evaluated using kappa statistics, and contingency table analysis and Poisson regression models were constructed to examine mode acceptability.

Results: Respondents randomized to the telephone diary cohort reported less frequent use of barrier contraceptive methods, specifically less spermicide use (odds ratio 0.27, 95% confidence interval 0.08, 0.95), and decreasing male condom use over time, whereas reports of male condom use increased for written diary respondents ($p = .007$). Participant characteristics associated with diary acceptability, defined as the frequency of diary completion, were assessed and teens classified as higher risk provided fewer diary reports ($p < .01$). Regardless of mode completed, 65% of respondents believed the tele-

phone diary would be preferable to the written diary for most teens.

Conclusions: The automated telephone diary offered an acceptable, even preferred, methodologic alternative to the written diary calendar and elicited more accurate reporting of selected contraceptive behavior. © Society for Adolescent Medicine, 2001

KEY WORDS:

Adolescents
Contraceptive behavior
Data collection
Measurement
Questionnaires
Reproducibility of results
Self-report
Sexual behavior

In 1996 there were 15.3 million new cases of sexually transmitted infections (STIs) in the United States (1,2). Among U.S. women, deaths associated with STIs increased by 31% from 1985 to 1992, a rise that reflects the growing incidence of heterosexually transmitted human immunodeficiency virus (HIV) infection (3). With the emergence of HIV infection and the high incidence of STIs, epidemiologic research has increasingly focused on assessing associated high-risk sexual behaviors. Describing the prevalence of sexual activities and their relationships with subsequent STIs requires accurate measurement techniques. Information on sexual behavior, however, is limited primarily to self-report and in general cannot be validated through observation or through direct correlation with biologic end points (4,5). Reliability, which functions as a surrogate

From the Center for Reproductive Health Research and Policy, Department of Obstetrics, Gynecology and Reproductive Sciences, University of California, San Francisco; and Department of Epidemiology, School of Public Health, University of California, Berkeley, California.

Address correspondence to: Alexandra M. Minnis, M.P.H., Center for Reproductive Health Research and Policy, University of California San Francisco, 74 New Montgomery, Suite 600, San Francisco, CA 94105-3444. E-mail: aminnis@uclink4.berkeley.edu.

Manuscript accepted September 21, 2000.

measure of validity, can be considered in instances where the same behavior is measured in multiple ways. Reliability studies evaluate the consistency of data obtained for a particular individual through several modes of data collection, or, alternatively, in studies that include participants' sexual partners, through concordance ratings for couples' reports (6–10). Although examining reliability does not offer definitive conclusions about over- and underreported behaviors, a greater frequency of sensitive reports are generally assumed to be more valid (11).

The mode of data collection (e.g., in-person interview, self-administered questionnaire [SAQ], and audio computer-assisted self-interviewing [ACASI]) has been shown to affect the degree to which sensitive behaviors are reported. Studies suggest that respondents are more likely to report high-risk sexual behaviors in more anonymous settings provided by SAQs and ACASI surveys than during in-person interviews (12). One national study found that individuals who completed an automated telephone interview reported higher levels of sensitive behaviors than those randomized to an interviewer-administered telephone interview (13). Yet, even among data collection modes in which the interviewer is absent, reporting of sensitive behavior has been found to vary. For example, comparison of written SAQs and ACASI using a randomized controlled design in the National Survey of Adolescent Males revealed that participants assigned the ACASI mode reported significantly higher levels of sensitive information, including having sex with someone who injects drugs (odds ratio [OR] 17.1, $p < .05$) and male-to-male sexual contact (OR 4.2, $p < .01$) (11).

The reliability of self-reported sexual behavior data has most often been estimated by comparing retrospective interviews (in-person or SAQ) to diaries. Diaries generally use a SAQ calendar format to collect daily behavioral data. This provides an opportunity to improve data reliability both by affording the respondent anonymity and by increasing the temporal proximity of data collection to the events being measured (14–16). Thus, diaries are assumed to reduce biases associated with interviewer-administered questionnaires and with recall of sexual behaviors (17,18).

Several recent studies have compared the frequency and timing of events reported through sexual behavior diaries with data reported retrospectively for the same period through questionnaires conducted as either SAQs or in-person interviews. Conclusions regarding reporting differences between diaries and interviews remain inconsistent across

studies (14,19,20). Whereas studies that compare data obtained through diaries and retrospective interviews have revealed differences in the reporting of sensitive behaviors by measurement mode, few published studies have varied the diary methodology itself to evaluate and refine its reliability. The accuracy and validity of written diary data may depend on the frequency with which the research participant completes the diary calendar. In prospective studies, participants are instructed to complete the diary daily, and to return it either weekly by mail or months later at their next interview visit. Thus, participants may complete multiple missed days or weeks before a study visit and thereby reduce the validity of data presumed to be most accurate. Without accurate data about when the diary is really completed, relying on written diaries to validate interview data may be inappropriate.

This study employed an automated interview technology designed to improve the accuracy of diary data both by increasing the anonymity and confidentiality involved in completing the diary and by validating reporting frequency. Our primary hypothesis was that among this population of female adolescents, the automated telephone interview would confer a greater sense of privacy than the written calendar, and result in increased reporting of sensitive sexual behaviors. Our second hypothesis was that the diary data, regardless of mode, would elicit greater reporting of sensitive behaviors compared with the retrospective interview.

Methods

Study Population

Female participants ranged in age from 15 to 19 years and were recruited from among clients seeking reproductive health care services at a Planned Parenthood teen clinic in San Rafael, California, and through community outreach during May through December 1998. This study originated from a larger prospective cohort study of contraceptive use that investigated the acceptability of barrier contraceptive methods among sexually active young women and their male partners. Eligibility (determined through an interview administered either over the telephone or in person) included having had sex at least three times in the previous 3 months, not being pregnant or trying to get pregnant during the next year, and residing in the San Francisco Bay Area for the study month.

Study Design and Data Collection

We conducted a randomized controlled trial in which participants were randomized into either a written calendar or automated telephone interview diary cohort to record data on sexual behavior and contraceptive method use each day for 1 month. Study participation involved completion of the parent study baseline interview (including data on sociodemographic characteristics, sexual behavior, drug use, contraceptive method use, and medical history), a daily diary (written or telephone) for 4 weeks, and a follow-up interview 1 month after baseline (referred to below as "retrospective interview"). Participants were randomized to the telephone or written diary based on a study identification number assigned in numeric order at baseline. All individuals with an even study number were placed in the telephone diary protocol and those with odd study numbers were assigned the written diary. Participants received an educational presentation on a range of contraceptive methods, with an emphasis on barrier methods, including male and female condoms and vaginal spermicides. After the presentation, barrier methods were offered to all participants.

Interviewers explained the diary protocol to participants in a standardized presentation. Teens completed either the written or telephone diary by answering three questions each day: whether they had had vaginal sex in the past 24 hours, and if so, how many times; what birth control method(s) they used, if any; and whether they had used any drugs or alcohol in the previous 24 hours. The telephone diary cohort was asked to call a 24-hours toll-free number each day and respond using the telephone key pad and by voice to a prerecorded, automated interview (i.e., "Press 1 if you had vaginal sex," and "Name any methods of birth control you used when you had sex"). Because mailing a diary each day would have imposed a high burden on study participants, the written diary cohort was instructed to complete one column of a weekly calendar each day and to return it in a preaddressed, stamped envelope by mail weekly. Thus, the design allowed more precise validation of reporting frequency for telephone respondents. Participants in both groups were also encouraged, through an open-ended question, to use the diary to ask reproductive health questions of the research staff and to request additional contraceptive methods.

All study participants received \$35 reimbursement for interviews and for completing the diary

(\$20 at baseline and \$15 at follow-up). Because in designing the study we believed the daily telephone call introduced an increased burden on participants compared with the requirements of the written diary, additional opportunities to earn small incentives for completing the diary daily were integrated into the telephone protocol. As an incentive to complete the diary, participants received one ticket into a monthly gift certificate drawing for each diary calendar they returned or each telephone call they made. In addition, small prizes were given to respondents who phoned the diary on either of two randomly selected incentive days each week or returned their written diary calendar for a randomly selected week each month. All study participants provided written informed consent before their participation in the research. The study was approved by the Committee for Human Research at the University of California San Francisco.

At the completion of the 4-week diary period, all participants were phoned by study staff and were asked to complete a brief retrospective interview about their sexual behavior, contraceptive method use, and reproductive health over the previous month. To insure confidentiality, if research staff were not able to speak directly with participants, a code name and message agreed upon with the participant at baseline were used. In addition, an appointment was scheduled for participants to return to the clinic to complete an interview that evaluated factors related to diary acceptability and examined changes in sexual and contraceptive use behaviors associated with keeping the diary. Any incentives earned were disbursed at this visit.

Statistical Analysis

To determine the effectiveness of our randomization, we evaluated bivariate associations between diary mode and a range of sociodemographic factors, sexual and drug use behaviors, contraceptive method use, and reproductive health outcomes.

We used generalized estimating equations (GEEs) to analyze differences in reporting between the automated telephone and written calendar diaries for all sexual behavior, contraceptive method use, and alcohol/drug use outcomes. The GEE approach estimates linear models that account for the correlation among the repeated measurements collected for each respondent (21). Using either a binomial or a Poisson distribution to correspond with the dependent variable, the coefficients of the models (expressed as ORs or incidence rate ratios) describe marginal probabil-

Table 1. Distribution of Baseline Sociodemographic Factors, Sexual Behaviors, and Reproductive Health Outcomes: A Comparison Across Study Groups

	Diary Type				Test Statistic	p Value
	Written (n = 52)		Telephone (n = 53)			
	n	(%)	n	(%)		
Sociodemographic factors						
Age (mean y)	17.5		17.4		0.3 <i>t</i>	.26
Race/ethnicity						
White	30	(57.7)	34	(64.1)		
Black	5	(9.6)	2	(3.7)		
Latina	4	(7.7)	4	(7.6)		
Asian	7	(13.5)	4	(7.6)		
Other/multiracial	6	(11.5)	9	(17.0)	2.9 χ^2	.57
Proportion in school	42	(80.8)	47	(88.7)	1.3 χ^2	.26
Sexual behaviors						
Age at onset of sexual intercourse (mean y)	15.0		14.8		1.1 <i>t</i>	.24
Coital debut <15 y old	18	(35.3)	22	(41.5)	.4 χ^2	.52
>1 partner past 6 mo	23	(44.2)	27	(50.9)	.5 χ^2	.49
Contraceptive method use (previous 6 months)*						
Condoms	49	(94.2)	51	(96.2)	.2 χ^2	.63
Spermicides	6	(11.5)	5	(9.4)	.1 χ^2	.73
Birth control pills	17	(32.7)	27	(50.9)	3.6 χ^2	.06
Withdrawal	0	(0)	2	(3.8)		.5 [†]
No methods	17	(32.7)	23	(43.4)	1.3 χ^2	.26
Emergency contraception (ever used) [‡]	9	(20.5)	11	(25.6)	.3 χ^2	.57
Reproductive health						
Ever pregnant	9	(17.3)	8	(15.4)	.1 χ^2	.79
Ever abortion	8	(15.4)	5	(9.4)	.9 χ^2	.36
Ever STI	7	(13.5)	6	(11.3)	.1 χ^2	.74

* Contraceptive methods used sum to >100% because of dual method use and because respondents changed methods used during this period.

[†] Seventeen participants not asked because they indicated that they had never heard of emergency contraception.

[‡] Fisher's exact test two-tailed.

STI = sexually transmitted infections.

ities for the likelihood of reporting each outcome as predicted by the diary mode and relevant covariates. Individual models were constructed for each outcome, which included having had vaginal sex, the total daily frequency of vaginal sex, use of contraceptives, use of particular methods (male condoms, female condoms, spermicides, hormonal methods, and withdrawal), and use of alcohol/drugs. Each observation included in the analysis represented 1 day of reported behavior. We evaluated trends in reporting over the month for each of these outcomes by assessing time since the start of the diary period using a continuous variable that measured time in days. Furthermore, we defined an interaction term between the time variables and diary mode and examined a modification of the relationship between time since the start of the diary period and each outcome by diary mode.

To compare responses obtained from retrospective interviews completed at the end of the diary period and from the diaries themselves (both for written and telephone separately, and collapsed by

diary mode), we assessed concordance beyond chance using weighted kappa statistics. Kappa statistics characterize interrater reliability for ordered, categorical variables, with a maximum value of 1.0 indicating perfect agreement (>0.75 represents excellent agreement), a value of zero connoting no agreement beyond chance (<0.4 represents poor agreement), and a negative value suggesting worse than chance agreement (22). We aggregated several outcome variables reported across the month through diaries to compare reported frequency of vaginal sex, use of any contraceptive methods, use of each contraceptive method (ever/never during the month for male condoms, female condoms, spermicides, hormonal methods, and withdrawal), and frequency of use across the two data collection modes. When calculated separately by diary mode, concordance between diary and retrospective interview data did not vary by mode. Therefore, we calculated kappa statistics comparing diary data (both modes combined) with retrospective interview data.

Assuming that higher completion rates corre-

Table 2. Differences in Reporting of Sensitive Behavior: Telephone Compared With Written Diary

Outcome Reported for Previous 24-h Period	%	Unadjusted OR/IRR ^a (Telephone/Written)	95% CI
Sexual behavior			
Vaginal sex (<i>N</i> = 1999)	26.0	1.21	.81, 1.89
Frequency of sex (<i>N</i> = 455) ^b	1.6 (mean times/day)	1.07	.89, 1.34
Contraceptive method use (<i>N</i> = 459)^b			
Any method used	81.4	2.21*	.91, 5.35
Female condom	3.5	.46	.15, 1.72
Male condom	47.9	.87	.46, 1.80
Spermicide	12.4	.27**	.08, .95
Hormonal	44.0	2.26*	.93, 5.04
Withdrawal	23.8	0.86	.35, 2.20
Alcohol/drug use (<i>N</i> = 1699)	25.5	1.09	.61, 1.94

^a Written diary cohort is baseline. ORs and IRR therefore express probability of reporting each outcome for telephone compared with written diary respondents.

^b Frequency and percentage of days when vaginal sex reported. Total *N* varies owing to missing data.

* $p < .1$; ** $p < .05$.

sponded to greater diary acceptability, we explored the acceptability of each diary mode by identifying predictors of diary completion. Predictors of acceptability were defined by baseline risk characteristics and factors related to an individual's desire for her study participation to remain confidential. We left the outcome as a count of total days of diary data reported and used Poisson regression to assess associations with baseline sexual risk factors and contraceptive practices (early age at onset of sexual intercourse (<15 years old), histories of pregnancy and abortion, multiple sexual partners in the previous 6 months, and use of contraception during sex in the previous 6 months). In a separate model we considered confidentiality-related factors (currently lives with parents, confidential contact from research staff requested, and whether the participant had her own telephone and/or pager). In addition, we assessed participants' acceptability of their respective diary mode directly. Specifically, we inquired about motivations for completing the diary, preference between the two diary modes (regardless of the mode to which they were randomized), and confidentiality concerns related to study participation.

We performed all statistical analyses using STATA 5.0 (College Station, Texas, 1990) and SAS System 6.12 (Research Triangle Park, North Carolina, 2000).

Results

Participant Characteristics

Of 139 adolescents, 118 were eligible (21 ineligible primarily because not sexually active currently) and 107 agreed to participate in the study. Females who

declined participation did not differ from those who did enroll with regard to demographic characteristics or current sexual risk behaviors. Fifty-four participants were randomized to the written calendar cohort and 53 were assigned the automated telephone diary. Two participants did not complete the baseline interview, and thus the final sample included 105 respondents (89% of eligible subjects), with 52 written diary and 53 telephone diary participants.

The two study groups did not differ with respect to any baseline characteristics (Table 1). Participants were, on average, slightly more than 17 years old. At least 80% in each group were in school, and approximately 60% of each cohort were white. Over one-third of participants had first had vaginal intercourse before age 15 years, and approximately half reported having more than one sexual partner in the previous 6 months. Ninety-five percent of respondents reported using condoms in the previous 6 months, although use was inconsistent: Nearly 40% also reported using no contraception at some point during this period. A substantial number of participants reported that they had ever been pregnant (16%), had an abortion (13%), or had an STI (13%).

Comparing Telephone and Written Diary Reports

Of the 2940 possible observations, 1999 were obtained, and having had vaginal sex was reported by 26% of those. The probability of daily use varied significantly between the written and telephone diary cohorts only for spermicide use (Table 2). Telephone diary participants were four times less likely

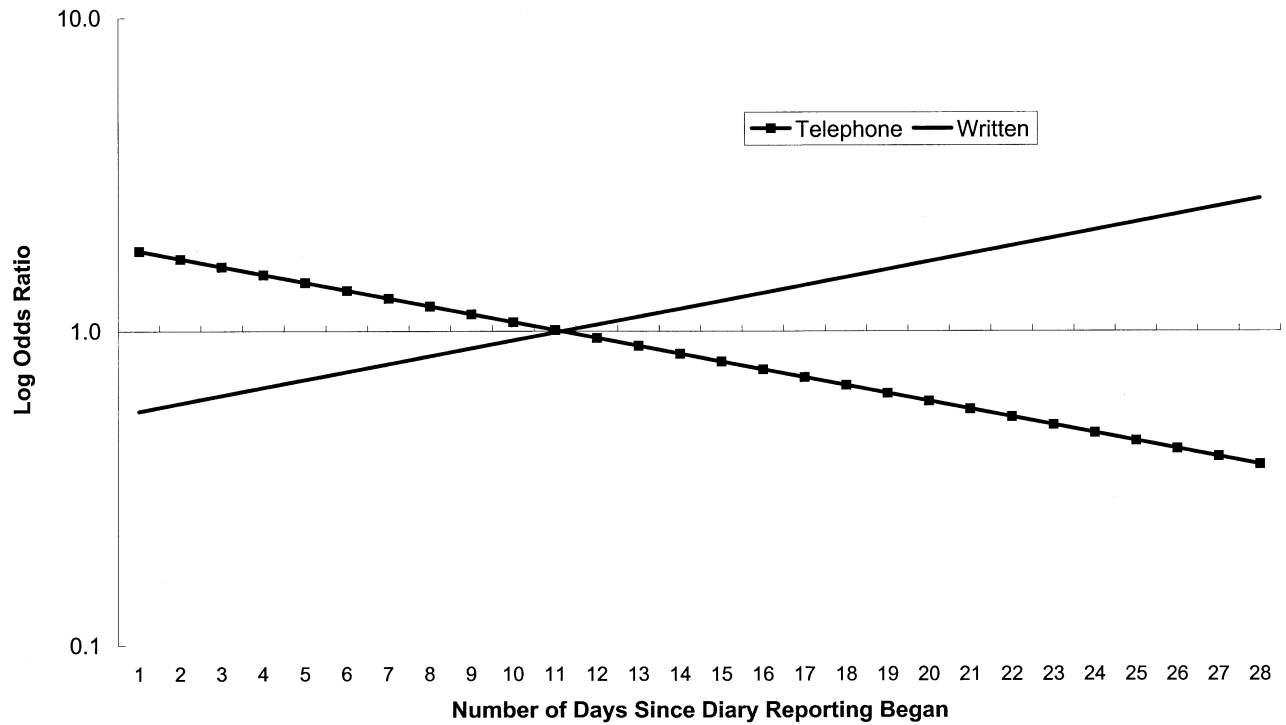


Figure 1. Log Odds of Reporting Male Condom Use Over Time by Diary Mode (Telephone vs. Written).

than written diary respondents to report spermicide use (OR 0.27, 95% confidence interval [CI] 0.08, 0.95). No statistically significant differences were observed for other outcomes: sexual intercourse frequency; use of condoms, female condoms, or withdrawal; or use of alcohol or drugs, although telephone diary respondents were two times more likely to report use of any method of contraception and use of hormonal contraception.

We were also able to examine the probability of reporting specific contraceptive use behaviors over the study duration. We assessed the relationship between time since start of diary and the reporting of contraceptive method use by diary mode. Only for male condom use did the probability of reported use vary across the 28 days, and the direction of this relationship differed by diary mode (Figure 1). Telephone diary participants were initially nearly two times more likely than written diary participants to report condom use; however, their probability of reporting condom use decreased over the month, whereas it increased for written diary participants ($p = .007$).

Reliability of Diary and Interview Data

In addition to comparing data collected through the two diary methods, we evaluated reliability by com-

paring diary data (both modes combined) to retrospective interview data collected through a telephone interview at the conclusion of the 1-month diary period. Reports of vaginal intercourse frequency, use of contraception (any method), contraceptive use by method, and frequency of use among users of a specific method were compared using weighted kappa statistics (Table 3). Concordance of reported vaginal intercourse frequency was moder-

Table 3. Concordance Between Aggregated Diary and Retrospective Interview Data

Behavior	κ Statistic	95% CI
Vaginal sex frequency	.57	.44, .70
Contraceptive methods: ever used		
Any method	.33	.13, .52
Female condom	.53	.31, .76
Male condom	.34	.15, .54
Spermicide	.57	.38, .75
Hormonal	.71	.57, .86
Withdrawal	.55	.36, .74
Contraceptive methods: frequency of use ^a		
Female condom	1.0	1.0, 1.0
Male condom	.52	.35, .70
Withdrawal	.52	.36, .69

^a Spermicides and hormonal methods excluded because frequency of use categories not consistent across data collection modes.

ate. The reliability of reported use of contraception between diaries and retrospective interviews, however, was poor. In investigating concordance between reports of specific method use, concordance for use of female condoms, spermicides, and withdrawal was classified as moderate, whereas for male condoms concordance was poor. For spermicides and both female and male condoms, discordance was driven by greater reported use of these methods in the retrospective interview than in the diary. Conversely, more reports of withdrawal were obtained through diary data than through retrospective interview. Reports of hormonal contraceptive use from the two data collection modes showed excellent concordance. Concordance in frequency of condom use (male and female) and withdrawal was moderate to excellent.

Acceptability of Diary Mode

Overall, 52% of participants completed 21 or more days of the 28-day diary period, with over one-third of those completing the diary every day. Whereas written diary participants constituted 67% of the individuals above the median completion rate, they also accounted for nearly all of the noncompleters (80%). In addition to the diary mode itself, we examined participant characteristics that might influence compliance. We explored baseline sexual risk factors, contraceptive use behaviors, and factors related to the participant's living situation and desire for confidentiality. Completion frequency was significantly related to having had multiple partners in the previous 6 months ($p = .03$) and to whether the participant had her own telephone line ($p = .05$), with lower levels of completion among teens who had multiple partners and among those who did not have their own telephone line.

In multivariate Poisson regression, several baseline sexual risk behaviors, contraceptive use practices, and confidentiality-related factors were significantly associated with diary completion frequency. Two distinct models, one that included sexual risk and contraceptive behaviors and another that addressed confidentiality-related factors, were constructed. Each was adjusted for diary mode. In our final sexual behavior model, early age of onset of sexual intercourse, having ever been pregnant, having had an abortion, and reporting multiple sexual partners in the previous 6 months were associated with a reduced completion rate ($p < .001$) (Table 4). High-risk contraceptive practices were not significantly associated with completion frequency, al-

Table 4. Multivariate Poisson Regression Models for Factors Associated With Diary Response Frequency

Predictors	Adjusted	
	IRR	95% CI
Model 1: baseline sexual risk behaviors		
Young age of coital debut (<15 y old)	.87*	.79, .97
Ever pregnant	.17**	.09, .31
Ever abortion	5.62**	3.06, 10.31
Multiple partners previous 6 months	.96**	.94, .99
Ever used no birth control previous 6 months	.98	.89, 1.09
No male condom used during last vaginal sex	.99	.90, 1.10
Diary mode (baseline = written)	.80**	.72, .88
Model 2: confidentiality-related factors		
Lives with parents	.72**	.63, .83
Confidential messages at home requested	.85*	.76, .95
Teen has own telephone	1.04	.93, 1.15
Teen has own pager	.95	.86, 1.05
Diary mode (baseline = written)	.69**	.62, .76

* $p < .01$; ** $p < .001$.

though the directions of the associations suggested that they were associated with decreased diary completion rates. In the confidentiality-related factors analyses, living with parent(s) and having requested confidential messages from study staff were both significantly associated with a decreased completion rate. Access to a personal telephone or to a pager was not related to response patterns.

The most important reason for diary completion cited by both groups in the evaluation interview was the opportunity to earn the incentives disbursed for randomly selected days on which diary data were reported (telephone cohort, 54%; written cohort, 68%). Over half of respondents also mentioned their interest in contributing to research and their consequent obligation to the study. The most important reason for not responding cited by both groups was that they forgot (telephone cohort, 76%; written cohort, 67%). When asked to rate their preference between the telephone and written diary modes, a greater proportion of the telephone respondents (78%) compared with the written cohort (53%) preferred the diary method to which they had been assigned ($p < .01$). The three most frequently cited reasons for preferring the telephone diary were that calling on the telephone is more convenient than writing, that mailing the written diary is difficult, and that the telephone diary is "novel" and "fun." Of those who believed one diary method would be preferable to another, 84% of telephone and 45% of written diary respondents believed that the telephone diary method would be better for most teens.

The importance of protecting confidentiality was underscored by written diary participants. The most commonly cited location for storing the diary calendars was a "secret" bedroom drawer (39%). In fact, only 46% of all participants told their parents they were enrolled in the study, whereas 97% told their friends and 76% told their sexual partner(s).

Discussion

The comparisons of diary data obtained through the written and telephone modes suggest that, apart from spermicide use, the overall likelihood of reporting specific contraceptive method and alcohol/drug use did not vary significantly by collection mode. Use of any contraceptive method during vaginal sex, however, was reported, but not significantly, somewhat more frequently among the telephone diary cohort. This difference appeared to be driven by the higher reported use of hormonal contraceptives, which constituted a large proportion of protected sexual acts. As participants were randomly assigned to the two diary modes, these unadjusted odds and incidence rate ratios are likely to be unbiased measures of the effect of diary mode on each behavioral outcome. Although the small sample size renders the results tentative, the probability of a Type 1 error is reduced by having included each day of reported sexual intercourse, adjusted for the correlation between the repeated observations for each participant, as an individual observation in the models.

In addition, reporting of male condom use varied by diary mode over time. The probability of reporting use of male condoms decreased over the month for the automated telephone diary cohort, whereas it increased over time for the written diary cohort. Based on findings from national studies indicating that automated interviews administered through ACASI yield a greater prevalence of sensitive data compared with self-administered questionnaires (11), we expected to detect a difference in prevalence of reported behaviors between these two diary modes. The two differences we did detect in reported barrier contraceptive method use (spermicides and male condoms over time) may reflect participants' perception that using condoms and/or spermicides during vaginal sex is socially desirable. The decreased probability of reporting male condom use over time for the telephone cohort that paralleled an increased probability of reporting condom use over time for the written cohort suggests that the telephone diary may reduce the tendency to overreport socially desirable behavior.

The observed trend in reporting of male condom use may also point to differential data collection mode-effects on practicing particular behaviors. Through the written recording of behavior on a 1-week calendar, participants' awareness of their contraceptive use may be heightened, which itself may engender an increase in condom use. Alternatively, the act of recording may alter reporting of such behavior. The automated telephone interview, conversely, does not provide participants with a tangible summary of their behavior, and did not appear to have the same effect on respondents' reported condom use. Thus, the observed reporting differences could reflect a recall bias introduced by the presumed delayed reporting through the written diary compared with the "real time" (date and time of the telephone call is known precisely by the researcher) telephone diary reports.

We found substantial discordance between the diary data (regardless of mode) and interview data collected retrospectively for the diary period. Most studies that have evaluated concordance between diary and comparable retrospective interviews have concluded that agreement is fairly strong for most behaviors and, therefore, that retrospective interviews with recall periods of several months may sufficiently approximate real-time data collected through diaries (15,16,18-20). In contrast, our assessments of agreement for each contraceptive method individually suggest that the retrospective data may be subject to substantial recall bias, or that participants are more likely to report socially desirable behavior through the interviewer-administered questionnaire. Use of male and female condoms and spermicides was more likely to have been reported during the interview than through the diary. Finally, a higher reported frequency of use of withdrawal through the diary compared with the interview may reflect greater candidness in reporting of sexual behavior through this mode.

Our study results are limited by several factors related to study design. Our conclusions regarding differences in acceptability and in reporting of sensitive behaviors between the written calendar and automated telephone diaries rely on data obtained from simple randomization of participants into one of two study arms as opposed to a crossover design in which respondents switched arms at the midpoint of follow-up. A crossover design would have allowed us to assess the consistency of our findings across individuals who completed both diary modes and thereby further substantiate our observations. It would also have strengthened the diary acceptability

interview during which participants were asked to identify their preferred diary mode when they had completed only one of the two evaluated here. Nonetheless, a crossover design has its own limitations, as results could be biased by the order in which modes were completed. Other design-related limitations include the short follow-up period and small sample size, both of which may have reduced our ability to detect additional associations between mode and reporting of sensitive behaviors. In addition, because of the primarily clinic-based recruitment, it is possible that we observed a diary mode effect that is not generalizable beyond this cohort.

As increased reporting of risky sexual behavior through a particular data collection mode is believed to indicate more accurate responses, our experiment provides some evidence that an automated telephone diary may elicit more accurate reporting of sensitive sexual risk behavior than a written calendar diary. In particular, reported use of male condoms and spermicides appears to be influenced by diary mode. Most significantly, we conclude that significant differences exist between the information captured through the diary, regardless of mode, and the interview conducted at the end of the diary period. Specifically, the diary data appear to be less biased by a desire to have practiced socially desirable contraceptive behaviors, and thereby to offer more accurate reports. In comparing the two modes, the automated telephone diary seems to have several practical advantages over a written diary. The data collection mechanism (which provides direct data entry) facilitates the validation of the reported information as real-time data. In addition, in contrast to the written diary, the telephone diary does not appear to provide the respondent with a tangible summary of sensitive behavior that may modify their reporting and perhaps their behavior. The telephone diary also affords greater privacy and confidentiality for an adolescent population. Finally, and most important, adolescents expressed a preference for the telephone diary. Thus, the automated telephone diary may lessen participant burdens that are particularly pronounced for adolescents. Future sexual risk research among adolescent populations, including males, may consider employing this automated telephone technology, even potentially the Internet, to augment data for which accurate reports of frequency or timing of events is critical. Furthermore, investigators might experiment with the implementation of this data collection mode by modifying the length of the reporting period, including additional

sensitive items in the questionnaire (e.g., anal sex, steady vs. casual partner for a particular sexual event), and developing a diary reminder system for participants to increase response rates.

This research was supported by Institute for Child Health and Human Development, NIH, Grant 1R01HD32789-01A1. Many thanks are extended to the dedicated research staff: Jessica Bowen, Lilia Cardenas, Mary de la Roca, Sarah Glass, Heather Gould, Margaret Handley, Cío Hernandez, Jill Johnson, Janet Jonte, Abigail Norris, Andrea Raider, Judy Quan, April Vogensen, and Jane Wong. In addition, the authors are grateful to the study participants who volunteered for the Reproductive Health Project and to the staff at our clinical site: Planned Parenthood San Rafael. This study, including all consent forms and research protocol, was approved by the Committee for Human Research (Institutional Review Board) at University of California, San Francisco. All study participants provided written informed consent.

References

1. Kaiser Family Foundation and American Social Health Association. STDs in America: How many cases and at what cost? Menlo Park, CA: Author, 1998.
2. Donovan P. Confronting a hidden epidemic: The Institute of Medicine's report on sexually transmitted diseases. *Fam Plann Perspect* 1997;29:87-9.
3. Ebrahim SH, Peterman TA, Zaidi AA, Kamb ML. Mortality related to sexually transmitted disease in United States women, 1973-1992. *Am J Public Health* 1997;87:938-44.
4. Catania JA, Gibson DR, Chitwood DD, Coates TJ. Methodological problems in AIDS behavioral research: Influences on measurement error and participation bias in studies of sexual behavior. *Psychol Bull* 1990;108:339-62.
5. Weinhardt LS, Forsyth AD, Carey MP, et al. Reliability and validity of self-report measures of HIV-related sexual behavior: Progress since 1990 and recommendations for research and practice. *Arch Sex Behav* 1998;27:155-80.
6. Elish NJ, Weisman CS, Celentano D, Zenilman JM. Reliability of partner reports of sexual history in a heterosexual population at a sexually transmitted diseases clinic. *Sex Transm Dis* 1996;23:446-52.
7. Lauritsen JL, Swicegood CG. The consistency of self-reported initiation of sexual activity. *Fam Plann Perspect* 1997;29:215-21.
8. Newcomer S, Udry JR. Adolescents' honesty in a survey of sexual behavior. *J Adolesc Res* 1988;3:419-423.
9. Orr DP, Fortenberry JD, Blythe MJ. Validity of self-reported behaviors in adolescent women using biomarker outcomes. *Sex Transm Dis* 1997;24:261-6.
10. Padian NS. Sexual histories of heterosexual couples with one HIV-infected partner. *Am J Public Health* 1990;80:990-1.
11. Turner CF, Ku L, Rogers SM, et al. Adolescent sexual behavior, drug use, and violence: Increased reporting with computer survey technology. *Science* 1998;280:867-73.
12. Turner CF, Danella RD, Rogers SM. Sexual behavior in the United States, 1930-1990. *Sex Transm Dis* 1995;22:173-90.
13. Turner CF, Miller HG, Rogers SM. Survey measurement of sexual behavior: Problems and progress. In: Bancroft J (ed). *Researching Sexual Behavior: Methodological Issues*. Bloomington, IN: Indiana University Press, 1997:37-60.
14. Fortenberry JD, Cecil H, Zimet GD, Orr DP. Concordance between self-report questionnaires and coital diaries for sex-

- ual behaviors of adolescent women with sexually transmitted infections. In: Bancroft J (ed). *Researching Sexual Behavior: Methodological Issues*. Bloomington, IN: Indiana University Press, 1997:237-49.
15. Harvey SM, Beckman LJ. Alcohol consumption, female sexual behavior and contraceptive use. *J Stud Alcohol* 1986;47:327-32.
 16. Leigh BC. Alcohol consumption and sexual activity as reported with a diary technique. *J Abnorm Psychol* 1993;102:490-3.
 17. Kauth MR, St. Lawrence JS, Kelly JA. Reliability of retrospective assessments of sexual HIV risk behavior: A comparison of biweekly, three-month, and twelve-month self-reports. *AIDS Educ Prevent* 1991;3:207-14.
 18. McLaws ML, Oldenberg B, Ross MW, Cooper DA. Sexual behavior in AIDS-related research: Reliability and validity of recall and diary measures. *J Sex Res* 1990;27:265-81.
 19. Fortenberry JD, Orr DP, Katz BP, et al. Sex under the influence: A diary self-report study of substance use and sexual behavior among adolescent women. *Sex Transm Dis* 1997;24:313-9.
 20. Leigh BC, Gillmore MR, Morrison DM. Comparison of diary and retrospective measures for recording alcohol consumption and sexual activity. *J Clin Epidemiol* 1998;51:119-27.
 21. Zeger SL, Liang KY. Longitudinal data analysis for discrete and continuous outcomes. *Biometrics* 1986;42:121-30.
 22. Fleiss JL. *Statistical Methods for Rates and Proportions*. New York: Wiley, 1981:217-25.