

# Accuracy of Audio Computer-Assisted Self-Interviewing (ACASI) and Self-Administered Questionnaires for the Assessment of Sexual Behavior

Dianne Morrison-Beedy · Michael P. Carey · Xin Tu

Published online: 23 May 2006  
© Springer Science+Business Media, Inc. 2006

**Abstract** This study examined the accuracy of two retrospective methods and assessment intervals for recall of sexual behavior and assessed predictors of recall accuracy. Using a 2 [mode: audio-computer assisted self-interview (ACASI) vs. self-administered questionnaire (SAQ)] by 2 (frequency: monthly vs. quarterly) design, young women ( $N = 102$ ) were randomly assigned to one of four conditions. Participants completed baseline measures, monitored their behavior with a daily diary, and returned monthly (or quarterly) for assessments. A mixed pattern of accuracy between the four assessment methods was identified. Monthly assessments yielded more accurate recall for protected and unprotected vaginal sex but quarterly assessments yielded more accurate recall for unprotected oral sex. Mode differences were not strong, and hypothesized predictors of accuracy tended not to be associated with recall accuracy. Choice of assessment mode and frequency should be based upon the research question(s), population, resources, and context in which data collection will occur.

D. Morrison-Beedy (✉)  
RN, School of Nursing, University of Rochester, 601 Elmwood  
Avenue, Box SON,  
Rochester, New York 14642  
e-mail: dianne.morrison-beedy@urmc.rochester.edu

M. P. Carey  
Center for Health and Behavior, Syracuse University,  
Syracuse, New York

X. Tu  
Biostatistics Computational Biology, University of Rochester,  
Rochester, New York

**Keywords** Audio-computer assisted self interview (ACASI) · Self-administered questionnaire (SAQ) · Sexual behavior · Women · Survey methods

## Introduction

Sexual health research relies almost exclusively on participants' self-report. Thus, identifying methods that optimize the accuracy of self-report is essential to advancing knowledge about sexual behavior and its health consequences, including unintended pregnancy, sexually transmitted disease (including HIV), sexual coercion, and dysfunction.

Frequently used self-report methods include face-to-face interviews, self-administered questionnaires, and computer-assisted interviewing. Each method has advantages and disadvantages. For example, face-to-face interviews (FTFIs) afford opportunities to establish rapport, to tailor questions to an interviewee's responses, and to minimize missing data. However, FTFIs are labor intensive and costly, and may lead to socially-desirable responses, less candid reporting, and lower quality data (Kurth *et al.*, 2004; Perlis, Des Jarlais, Friedman, Arasteh, and Turner, 2004). Concerns regarding FTFIs have led many investigators to prefer self- and computer-administered assessment methods.

Self-administered questionnaires (SAQs) are often used in field studies because they allow the efficient and low-cost collection of data from many participants. However, SAQs require that respondents be literate and, even with literate participants, SAQs can result in missing data because of skipped items. SAQs also require data entry, which introduces the opportunity for measurement error.

Like the SAQ, computer-administered methods provide greater privacy relative to FTFIs. With audio computer-assisted self-interviews (ACASI), the participant listens to interview questions privately using earphones attached to a computer. (Participants can simultaneously read the corresponding text on the screen.) Privacy is enhanced compared to FTFI because the participant responds to a computer rather than to an interviewer. High levels of literacy are not required because the questions are presented aurally as well as visually. Questions can be tailored so that participants will not be asked to respond to items that do not pertain to them. Encouragement and question clarification can be provided via preprogrammed voice prompts without the expense of personnel to provide them. In addition, out-of-range responses and missing data can be avoided because plausible responses to prior questions can be required for progression to the next item. Measurement error is also reduced because this method requires no data entry. Overall, ACASI has several advantages relative to other assessment methods.

Recent research also suggests that ACASIs may be more accurate than other methods. This suggestion emerges from several studies in which ACASIs yielded higher frequencies of sexual behavior relative to other methods (Des Jarlais *et al.*, 1999; Hewett, Mensch, and Erulkar, 2004; Newman *et al.*, 2002). Three studies have compared ACASI with SAQs. Turner *et al.* (1998) compared ACASI to SAQs with a large sample of male adolescents ( $n = 1690$ ). Participants who used ACASI reported more same-sex sexual behaviors and injection drug use relative to participants who used a SAQ. A second study of adolescent males ( $n = 928$ ) also found increased reporting of same sex behavior (Turner, Ku, Sonenstein, and Pleck, 1996), and a third study found that self-report of substance use were more frequent when adolescent girls used ACASI rather than a SAQ (Webb, Zimet, Fortenberry, and Blythe, 1999). Because higher reports of socially sensitive behavior are typically assumed to be more truthful, the inference drawn from these findings is that higher self-reports are more accurate.

In summary, several studies indicate that ACASI methods often yield higher reported risk behaviors than SAQs (Locke *et al.*, 1992; Romer *et al.*, 1997; Turner *et al.*, 1998). However, the assumption that ACASIs yield more *accurate* data than SAQs has not been directly evaluated using a “gold standard” (i.e., contemporaneous self-monitoring) to corroborate retrospective self-reports (Durant and Carey, 2000; Schroder, Carey, and Vanable, 2003; Weinhardt, Forsyth, Carey, Jaworski, and Durant, 1998).

A contemporaneous self-monitoring strategy that could be used to validate retrospective self-reports is use of a daily diary. Diaries allow brief yet detailed event level information in a format that is easy for respondents of all literacy levels. Diaries also appear to be a reasonable form of data collection without undue participant burden. In a study of 285 women

asked to return weekly surveys of their daily sexual behaviors for one year, the average amount of missing data for any given week was only 3.3% (Jaccard, McDonald, Wan, Dittus, and Quinlan, 2002). Measurement error should also be reduced because of the privacy of diaries, and the temporal proximity from the time of the sexual event to the recalling of the behavior. Daily diaries have been used to evaluate the validity of retrospective recall. In a study of 25 men assessing 1-month retrospective interview recall with their diaries, correlations approaching  $r = .9$  were reported for frequency of sexual intercourse (Reading, 1983). This was consistent even in those with the same, regular sex partner for whom “lack of distinctness” of routine may make recall more difficult than those with new partners. Similarly, 75 respondents were interviewed at either 1, 2, or 3 months after completing a 30-day diary to assess recall accuracy of sexual and substance behaviors (Graham, Catania, Brand, and Canchola, 2003). In comparison to the 1-month recall group, the only 2- or 3-month recall behavior that differed significantly from the diaries was recall of vaginal intercourse with 3-month recalls overestimating this behavior frequency. Despite the evidence that diaries can provide an important criterion to assess accuracy of retrospective recall, few studies have used a diary as a gold standard to evaluate other self-report measures. Thus, the primary purpose of this study was to compare the accuracy of two widely-used, self-administered methods of sexual risk behavior (viz., ACASI and SAQ) using a daily diary as the gold standard.

In addition to method of data collection, other factors may influence the accuracy of self-reports, including *cognitive demands* required by the method, the *social context* in which the self-report occurs, and *personal characteristics* of the respondent (Schroder *et al.*, 2003). The *cognitive demands* required by participants involved in sexual research include the length of time they are asked to recall and how often risk behaviors occurred. The length of the reference interval may impact the accuracy of self-reports with those requiring a longer period of recall to be more problematic than those with a shorter interval. Similarly, the frequency with which a behavior occurs may also impact recall accuracy.

Several studies have found that sexual behaviors can be recalled consistently for intervals of 1–3 months (Carey, Carey, Maisto, Gordon, and Weinhardt, 2001; Graham *et al.*, 2003; Jaccard *et al.*, 2002). Longer time frames may be more representative of a person’s sexual behavior patterns, but can be more difficult to recall. In contrast, errors for shorter time recall periods will receive greater weight because of the lower incidence of these behaviors. Behaviors that occur infrequently, and are thus thought to be memorable, are more likely to be accurately recalled than those that occur frequently making it more difficult to recall with precision how often they occurred. Additionally, participants asked

to recall behaviors over longer time periods may rely on a memory strategy such as “guestimation” of average number of dates per week they have with a specific partner. However, participants may use a process to recall specific even-by-event behaviors for shorter term or infrequent recall thus leading to increased error in the short term recalls (Jaccard and Wan, 1995).

The *social context* in which respondents are asked to recall behaviors that are personal and possibly carry negative societal connotations may influence the accuracy of such recall. In a study comparing response effects of a bystander on assessment of alcohol and drug use, adolescents were less likely to report substance use if a parent was present, especially for SAQs (Aquilino, Wright, and Supple, 2000). The presence of siblings had a small negative effect on reports and spouses had no effect. The differential effects of anonymity versus confidential responses were examined by Durant, Carey, and Schroder (2002). In the anonymous condition, more frequent risk behaviors were reported and non-response was lower than in the confidential condition. In addition, all 29 dropouts occurred in the confidential condition. These findings suggest that the ability of the data collection mode to provide as much privacy as possible and ensure anonymity or confidentiality of responses may improve accurate recall of responses.

At least four *person variables* (i.e., characteristics of the respondent) can be hypothesized to affect the accuracy of self-reported sexual behavior. First, “conscientiousness” (the psychological tendency toward meticulousness and precision) should be related to a person’s willingness to painstakingly recall past events. In a study of predominantly female participants examining conscientiousness, integrity, and honesty, conscientiousness predicted honest reporting of academic behavior (Horn, Nelson, and Brannick, 2004). Second, the regular use or abuse of alcohol and/or other drugs is likely to impair the accurate recall of such events. Concurrent alcohol use with sex predicted over reporting of some sexual behaviors in a study of college students assessing memory recall bias following a month-long daily diary (Graham *et al.*, 2003). Third, depressive mood might make accurate recall of behaviors more difficult because a depressed person has limited motivation for the mental tasks required to recall these activities (Fehnel, Bann, Hogue, Kwong, and Mahajan, 2004). Finally, the ability of the respondent to answer honestly without social demand for a desirable report of types and frequency of behaviors likely impacts accurate recall. Behaviors may be exaggerated or underreported depending on whether the respondent perceives them as socially desirable or socially disapproved. In a study of STD clinic patients, reporting agreement between ACASI and interview data on socially neutral variables was almost perfect but only moderately so for socially sensitive variables (Kurth *et al.*, 2004).

The primary purpose was to compare the accuracy of ACASI and SAQ; we predicted that ACASI would yield more accurate information than SAQs. The secondary purpose was to identify predictors of self-report accuracy. We predicted that accuracy will be associated with the following characteristics: less frequent sexual behavior, decreased duration of the assessment interval, higher levels of conscientiousness, lower depression scores, lower levels of social desirability, and less frequent substance use. We chose to study self-report accuracy and its predictors in a sample of young women because, previously, we found that young adult women reported feeling more threatened by questions about sensitive social behaviors and were more likely to refuse to answer such questions compared to similarly aged males (Durant *et al.*, 2002). Women were also less likely to admit socially undesirable sexual behaviors when assessed by face-to-face interviews than by computerized methods (Kissinger *et al.*, 1999).

## Methods

### Participants

Of the 159 women who expressed interest in the study, consented and provided baseline data, 64% completed the 3-month follow up. Of the 57 women who did not complete the study, 34 moved out of the area, 18 participants were lost to follow-up, and five said they were too busy to continue. The final sample consisted of 102 young women (range: 18–21 years;  $M = 19.6$ ,  $SD = 1.2$  years); 62% were White, 25% African American, 6% Hispanic, 6% Asian, and 1% other. All were unmarried, sexually active with a male partner in the past 3 months, not pregnant, and able to converse in English. The majority lived with either a parent(s) (33%) or friends (28%). Data were not collected on education or socioeconomic status because providers at the sites where we worked thought such items might be offensive to their more impoverished clients.

### Measures

Measures were administered at three times: (a) baseline, (b) contemporaneous self-monitoring phase, and (c) follow-up (i.e., retrospective recall phases). At baseline, a SAQ was administered to all participants; this SAQ assessed demographics, conscientiousness, depression, social desirability, substance use, and sexual behavior. During the self-monitoring phase, all data (i.e., daily substance use and sexual behaviors) were collected with a 3" × 5" diary card. At the follow-up(s), participants were assigned to take either a SAQ or an ACASI, dependent upon experimental condition, and

completed identical assessments of substance use and sexual behavior.

### *Demographics*

The demographics assessed included age, marital status, living situation, and race.

### *Conscientiousness*

The conscientiousness scale from the Comprehensive Personality and Affect Scales (COPAS) was used. This scale includes nine items and a 5-point Likert-type scale (1 = “not at all like me” to 5 = “very much like me”) (Lubin and Van Whitlock, 2002). Respondents considered nine different descriptors (e.g., organized, deliberate) to indicate how they generally think of themselves. Individual items are summed for a total score with higher scores indicating greater conscientiousness. Evidence for the validity and reliability of the measure have been compiled (Lubin and Van Whitlock, 2002). Cronbach  $\alpha$  for the current sample was .83.

### *Depression*

The 20-item Center for Epidemiological Studies-Depression (CESD) scale was used. Respondents indicated how often in the past week they experienced feeling blue, hopeless, lonely, and similar symptoms of emotional distress (Radloff, 1977). Individual items scores are summed with higher totals (range 0–60) indicating greater depressive symptoms. Total scores of 23 or greater indicate depressive symptoms in adolescents but are not diagnostic for clinical depression (Furukawa, Hirai, Kitamura, and Takahashi, 1997; Roberts, Andrews, Lewinsohn, and Hops, 1990). The validity and reliability of the CES-D is well-established (Radloff, 1977). Cronbach’s  $\alpha$  for the current sample was .91.

### *Social desirability*

The Social Desirability Response Scale (SDRS) was used to assess social desirability. Examples of these five items included “I sometime feel resentful when I don’t get my way” and “No matter who I’m talking to I’m always a good listener.” Individual items are summed using a 5-point Likert scale and higher total scores indicate higher levels of socially desirable responses. This scale has demonstrated reliability and validity (Hays, Hayashi, and Stewart, 1989). Cronbach  $\alpha$  for the current sample was .63.

### *Substance use*

Thirteen items using “yes/no” answer responses to “ever using” 13 drugs in your lifetime were used to assess drug

use (e.g., PCP; amphetamines such as speed; hallucinogens such as LSD, peyote; inhalants such as amyl nitrate, poppers; painkillers/opiates such as oxycodone). These same dichotomous items are repeated asking for use in the “past month.” Higher scores indicated more frequent drug use. A measure of alcohol use included one question asking for any alcohol use in the past month (yes/no/not sure). Yes responses were then followed by three questions asking about frequency and quantity of alcohol intake (number of drinks on typical day, number of days of alcohol intake in past month, number of days of 4 or more drinks in past month). Participants were also asked the number of times in the past 3 months that they had intentionally (a) drank less before sex or (b) use drugs less before sex.

### *Sexual risk behavior*

Four items assessed the frequency of unprotected and protected vaginal and anal intercourse, and four items assessed giving and receiving oral sex (both unprotected and protected with a condom or dental dam) using a count format. At baseline participants recalled these behaviors over the past 3 months and at the follow-ups either over the past month (ACASI/SAQ monthly) or past 3 months (ACASI/SAQ quarterly). These measures have been used in previous studies (Carey *et al.*, 1997, 2000; Kalichman, Rompa, and Coley, 1996).

### *Diary cards*

A 3" × 5" diary card was used to record substance use and sexual behaviors during the self-monitoring phase (see Fig. 1). The substance use questions included: total number of drinks that day, if any alcohol intake occurred before sex (yes/no), any drug use that day (yes/no), and any drug use before sex (yes/no). The sexual behavior questions included: number of episodes of vaginal, anal, and oral sex with and without condoms, partner type with each sexual act (new, occasional, regular) and whether that partner was male or female. We have used these cards in a previous study assessing sexual behavior in young women (Durant and Carey, 2000). These cards were written in code to maintain privacy and were small enough to carry with the participant for ease of completion. Each diary card had the participant’s ID number and date of birth but no other identifying data.

### *Procedures*

#### *Recruitment*

In 2003 potential participants were recruited from an inner-city family planning agency or adolescent health clinic

**Fig. 1** Diary card for daily behavior

DOB: ___/___/___	Code: _____						
	Mon	Tues	Wed	Thur	Fri	Sat	Sun
	___/___	___/___	___/___	___/___	___/___	___/___	___/___
1. Al (#)							
2. Alb4sx (y/n)							
3. Dr (y/n)							
4. Drb4sx (y/n)							
5. Part (new, occ, or reg)							
6. Part (m/f)							
7. v with c (#)							
8. v without c (#)							
9. a with c (#)							
10. a without c (#)							
11. o with c (#)							
12. o without c (#)							

**KEY** (not included on actual card):

1. drink alcohol
2. drink alcohol before sex
3. drugs
4. drugs before sex
5. sexual partner type (new, occasional, regular)
6. male/female partner
7. vaginal sex with condom
8. vaginal sex without condom
9. anal sex with condom
10. anal sex without condom
11. oral sex with condom
12. oral sex without condom

...serving impoverished youths within a city in the Northeast as well as by posters, flyers, and word-of-mouth. Seventy percent of those who consented and 67% of those who completed the study were recruited through the health agencies. Participants were screened for eligibility (i.e., unmarried, sexually active with a male partner in the past 3 months, not pregnant or trying to become pregnant, able to converse in the English language) in a private section of the waiting area or office. Those women interested in participating and who met the inclusion criteria then completed the consent process. We used complete randomization procedures to assign participants to one of the four conditions: ACASI monthly, SAQ monthly, ACASI quarterly, and SAQ quarterly.

*Baseline*

Participants completed a SAQ on various health-related behaviors, knowledge, attitudes, and beliefs. The survey took approximately 30 min to complete and participants were paid \$10 to help defray the cost of transportation or lost wages. Next, participants received detailed instructions on use of the diary cards. Both verbal and written instructions were provided regarding how to complete the cards documenting daily sexual and substance use behaviors using abbreviations to describe each behavior (i.e., “alb4sx” for “alcohol before sex”). They were given the 12 stamped, addressed cards and asked to return these on a weekly basis.

*Contemporaneous self-monitoring phase*

Participants completed the diary card daily, and returned one card each week for 12 weeks. For each diary card returned, participants received \$5 for a total of up to \$60 for this phase of the study.

*Follow-up (retrospective recall phase)*

As part of the experimental manipulation, one-half of the participants returned monthly for 3 months whereas the other one-half of participants returned only once, 3 months after baseline. At these follow-ups, they completed either ACASI or SAQ, as appropriate, with items related to sexual behavior and substance use. Each survey took less than 30 min to complete and participants were compensated \$10 for each occasion (i.e., \$10–30 in total depending on condition assignment).

*Data analysis*

Demographic, personality, substance use, and sexual behaviors were examined at baseline using ANOVA and Fisher’s exact test for differences between the four groups. To assess accuracy in retrospective reporting, we used the generalized linear model (GLM) with inference based on the generalized estimating equations (GEE). We first modeled the marginal means of each mode of assessment (diary cards and

retrospective reporting) for each variable within each condition (within-group comparison), where the response variable of the GLM was either the diary cards or retrospective assessment and the predictor was the time of assessment. This analysis allowed us to obtain the monthly means for the ACASI monthly and SAQ monthly conditions and the 3-month average for the ACASI quarterly and SAQ quarterly conditions. By averaging the model-based monthly means over the 3-month period, we obtained a model-based mean over the 3-month time for the ACASI monthly and SAQ monthly conditions. This strategy allowed us to estimate the mean response for the 3-month period. Standard deviations were calculated based on the empirical variance–covariance matrix of the model estimates under the exchangeable working correlation structure.

We used the GLM and GEE framework to assess differential accuracy in retrospective reporting across the four conditions (between-group comparison). For these analyses, the dependent variable was the difference between the diary cards and retrospective assessments.

We used the identity link function (or linear model) in the GLM because the Poisson model was not appropriate given the excessive number of zero values in these variables. Although the zero-inflated Poisson (ZIP) model can be used to account for the much inflated zero counts (Cheung, 2002; Lambert, 1992; ), implementation of such a model is not currently available for analyzing bivariate longitudinal data. In addition, to assess accuracy of retrospective reporting, it is more meaningful to use an index based on the absolute difference between the two assessments, rather than a Poisson model with one count variable used as the response and the other as a predictor. Because GEE provides robust inference regardless of data distributions, the linear GLM still gives rise to valid inference when modeling count data or difference between two such variables.

## Results

### Baseline data

Baseline demographic, personality, substance use, and sexual behavior are summarized in Table 1. Collapsed across conditions, participants scored high on conscientiousness,  $M = 35.9$ , and low on social desirability,  $M = 11.4$ . Mean scores on the CESD were 18.4 (high range of normal for depressive symptoms for this age group). Although most participants were legally underage, most consumed alcohol (76%) with binge drinking (i.e., four or more drinks on one occasion) occurring on average 4.2 times in the past month. Approximately one-third of the young women had used drugs in the past 30 days with 31% of the sample using marijuana and 13% using ecstasy. (Only those drugs that

10% or more of our sample reported using in the past month, three out of the 13 assessed, are reported in Table 1).

No group differences were found on any of these baseline characteristics, including substance use and sexual behavior. In their baseline retrospective recall of behaviors over the past 3 months (see Table 1), they averaged equivalent numbers of episodes of unprotected vaginal sex (range = 7–12,  $M = 9.7$ ) and protected vaginal sex (range = 6–11,  $M = 9.0$ ) with somewhat fewer episodes of unprotected oral sex (range = 5–10,  $M = 6.2$ ). Unprotected and protected anal sex and protected oral sex (giving and receiving) were reported infrequently by the sample as a whole ( $M = 0.29$  or fewer episodes/month).

### Contemporaneous self-monitoring data

Descriptive summaries of the contemporaneously-recorded diary data are presented in Table 2. Of the 102 women returning diary cards, 81% returned 12 completed cards and an additional 10% returned 11 cards; only one participant completed fewer than eight cards. Almost all (98%) of the women reported sexual behaviors on their diary cards during the 3-month self-monitoring phase (data not tabled). Because the number of participants reporting unprotected or protected anal sex (similar to Graham *et al.*, 2003), as well as protected oral sex was low, these data are excluded from further analyses. On the diary cards the majority (55%) reported having only a primary (regular) sexual partner, 36% reported having both regular and new or occasional partners, and 7% reported having only a secondary (new or occasional) partner (not tabled). During the course of 3 months, this sample reported on average 12.6 acts of protected vaginal sex, 13.5 episodes of unprotected vaginal sex, and 13.0 unprotected episodes of oral sex.

Although incentives were paid differently depending on whether they were assigned to the monthly or quarterly conditions, attrition did not differ as a function of payment schedule,  $\chi^2 = .90$ , *ns*. Table 2 provides the mean responses for the contemporaneous and retrospective self-reports over the 3-month self-monitoring period by experimental condition. Means are provided for the three variables, namely, protected vaginal sex, unprotected vaginal sex, and unprotected oral sex. Because preliminary analyses revealed that the responses of two participants were extremely atypical (i.e., outliers;  $\geq 3$  standard deviations from all other participants), data from these two participants were not included in subsequent analyses. To determine if random assignment produced equivalent groups, we compared the contemporaneously-recorded data (aggregated over the self-monitoring phase) for each sexual behavior using ANOVA-like approach. No differences across conditions were found for protected vaginal sex,  $\chi^2(3) = 0.53$ , unprotected vaginal sex,  $\chi^2(3) = 0.32$ , and unprotected oral sex,  $\chi^2(3) = 0.72$ , all  $p$ 's  $> .10$ .

**Table 1** Demographic, personality, substance use, and sexual behavior data by condition at baseline

	Overall		ACASI				SAQ				<i>F</i> or $\chi^2$
	<i>M</i>	<i>SD</i>	Monthly (3×)		Quarterly (1×)		Monthly (3×)		Quarterly (1×)		
			<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
Age	19.6	1.2	19.7	1.2	19.7	1.1	19.3	1.1	19.5	1.3	<i>F</i> = 0.69 <sup>a</sup>
Race (% White)	62%		59%		55%		67%		67%		
Conscientiousness	35.9	5.0	35.0	5.3	35.4	4.3	36.4	5.8	37.0	4.5	<i>F</i> = 0.81
Depression	18.4	6.5	20.2	6.3	18.1	8.4	17.3	3.9	17.9	7.2	<i>F</i> = 0.93
Social desirability	11.4	3.3	11.3	3.4	11.1	3.6	12.1	3.5	11.1	3.2	<i>F</i> = 0.50
<i>Alcohol use (past 30 days)</i>											
Any alcohol use (%)	76%		83%		67%		71%		81%		$\chi^2$ = 1.52 <sup>b</sup>
Quantity, number of drinks, typical use day	3.9	2.2	3.8	1.6	3.9	2.8	4.0	1.8	4.1	2.8	<i>F</i> = 0.12
Frequency, drinking days	6.4	6.5	6.5	6.3	5.4	7.2	5.5	5.5	7.5	7.2	<i>F</i> = 0.42
Quantity estimate (# days × typical use)	31.0	41.1	25.9	27.0	32.1	59.9	27.5	36.7	39.1	46.1	<i>F</i> = 0.44
Binge Frequency, ≥ 4 drinks at one time	4.2	6.0	4.2	5.7	4.4	9.3	3.6	4.9	4.6	4.6	<i>F</i> = 0.09
<i>Other substance use (% using in past 30 days)</i>											
Marijuana	31%		33%		29%		42%		22%		$\chi^2$ = 2.37
Ecstasy	13%		13%		14%		13%		11%		$\chi^2$ = 0.12
Opiates	12%		10%		14%		8%		18%		$\chi^2$ = 0.74
<i>Sexual behavior (past 3 months)</i>											
Oral, unprotected	6.2	11.0	7.0	11.3	5.6	16.1	9.6	8.2	5.5	8.2	<i>F</i> = 0.13
Oral, condom protected	0.29	2.0	0.90	3.69	0	0	0.04	0.21	0.04	0.20	<i>F</i> = 1.29
Vaginal, unprotected	9.7	15.8	12.0	18.7	7.0	13.3	9.4	10.5	9.4	18.1	<i>F</i> = 0.71
Vaginal, condom protected	9.0	12.8	9.5	12.9	11.3	15.9	9.8	14.0	6.0	8.7	<i>F</i> = 0.41
Anal, unprotected	0.26	1.2	0.10	0.31	0.24	1.1	0.78	2.1	0	0	<i>F</i> = 2.31
Anal, condom protected	0.10	0.36	0.10	0.40	0.10	0.49	0.17	0.49	0.04	0.20	<i>F</i> = 0.56

<sup>a</sup>*F* denotes the *F*-statistic with the numerator degree of freedom equal to 3 and the denominator degree of freedom equal to the available sample size for the variable under consideration which varies from 78 to 100 across the variables.

<sup>b</sup> $\chi^2$  denotes the  $\chi^2$  statistic with the degree of freedom equal to 3.

Because we used a variance estimate for each group rather than a pooled variance estimate across the groups and did not assume normality in the data with inference based on the GEE,  $\chi^2$  rather than *F*-statistics are reported.

Retrospective assessments occurred only once for the quarterly conditions; therefore, the monthly columns (labeled *a*, *b*, and *c*) in Table 2 for these conditions have no data. Also, as noted earlier, the mean responses and associated deviations reported in the column (*d*) labeled “total” for the ACASI monthly and SAQ monthly were calculated from the model-based monthly means for these two conditions to permit comparisons with the quarterly conditions for an equivalent time period.

**Within-conditions comparisons**

To compare accuracy of retrospective reporting *within* conditions, we assessed whether the retrospective assessments diverged from the diary accounts. As depicted in Table 2, four of the 12 within-conditions contrasts revealed significantly discrepant reports: (a) protected vaginal sex for ACASI

monthly; (b) unprotected vaginal sex for SAQ monthly; (c) unprotected oral sex for ACASI monthly; and (d) unprotected oral sex for SAQ monthly.

**Between-conditions comparisons**

To compare accuracy of self-reports across the four conditions, we again conducted two-factor ANOVA-like analyses. For these analyses, we used the difference scores (retrospective – diary) between retrospective assessment and diary as the response variable. As before, we used a variance estimate for each group rather than a pooled variance estimate across the groups so that  $\chi^2$  rather than *F*-statistics were reported from the GEE procedure.

Table 3 presents the means and standard deviations (*SDs*) of the cumulative events for each of the four groups for the three sexual behaviors. Because the *SDs* varied among conditions, variance estimates based on each individual group were used to assess the between-group differences using GEE. In addition, the marginal means and *SDs* for the main effects of methods and frequency of assessment show the effect of each factor when the other is ignored.

**Table 2** Within-condition comparison of contemporaneous with retrospective assessments<sup>a</sup>

Condition	Mean (SD)				$\chi^2$
	(a) 1st Month	(b) 2nd Month	(c) 3rd Month	(d) Total	
Protected vaginal sex					
ACASI Quarterly <sup>b</sup>					
<i>Retrospective recall</i>				4.95 (5.7)	
<i>Diary</i>	4.25 (5.0)	5.00 (5.0)	3.85 (5.0)	13.10 (12.4)	$\chi^2 = 10.83^{***c}$
ACASI Monthly					
<i>Retrospective recall</i>	4.6 (6.8)	2.8 (4.3)	3.12 (5.2)	10.46 (13.3)	
<i>Diary</i>	3.6 (5.6)	2.5 (4.2)	3.73 (5.2)	9.82 (11.8)	$\chi^2 = 0.09$
SAQ Quarterly					
<i>Retrospective recall</i>				9.77 (13.4)	
<i>Diary</i>	3.96 (4.8)	4.54 (6.0)	4.00 (6.2)	14.50 (14.6)	$\chi^2 = 1.97$
SAQ Monthly					
<i>Retrospective recall</i>	4.5 (7.4)	4.59 (6.8)	5.59 (8.2)	14.66 (19.5)	
<i>Diary</i>	4.7 (5.8)	4.17 (5.8)	4.05 (6.0)	12.91 (14.5)	$\chi^2 = 1.17$
Unprotected vaginal sex					
ACASI Quarterly					
<i>Retrospective recall</i>				8.8 (13.2)	
<i>Diary</i>	4.60 (7.5)	4.70 (7.0)	3.60 (6.0)	12.9 (19.2)	$\chi^2 = 1.64$
ACASI Monthly					
<i>Retrospective recall</i>	4.9 (7.1)	4.3 (5.4)	4.9 (5.9)	14.1 (14.3)	
<i>Diary</i>	6.0 (8.6)	4.9 (5.7)	4.0 (5.4)	14.85 (17.2)	$\chi^2 = 0.29$
SAQ Quarterly					
<i>Retrospective recall</i>				8.9 (12.2)	
<i>Diary</i>	3.35 (4.4)	3.15 (4.4)	4.27 (6.9)	10.8 (12.1)	$\chi^2 = 4.49$
SAQ Monthly					
<i>Retrospective recall</i>	6.3 (6.9)	6.9 (5.7)	7.4 (7.2)	20.5 (14.7)	
<i>Diary</i>	5.1 (6.1)	5.2 (6.2)	4.9 (5.2)	15.2 (13.8)	$\chi^2 = 15.14^{**}$
Unprotected oral sex					
ACASI Quarterly					
<i>Retrospective recall</i>				9.0 (18.4)	
<i>Diary</i>	4.15 (6.0)	4.85 (6.7)	2.90 (4.6)	11.9 (16.7)	$\chi^2 = 0.62$
ACASI Monthly					
<i>Retrospective recall</i>	8.5 (11.1)	8.3 (8.9)	6.7 (10.1)	23.5 (24.4)	
<i>Diary</i>	5.0 (5.8)	4.4 (5.0)	4.1 (4.7)	13.6 (13.2)	$\chi^2 = 4.22^*$
SAQ Quarterly					
<i>Retrospective recall</i>				11.8 (15.6)	
<i>Diary</i>	3.84 (4.4)	2.65 (3.3)	3.65 (4.0)	10.0 (9.4)	$\chi^2 = 0.38$
SAQ Monthly					
<i>Retrospective recall</i>	10.6 (14.4)	8.0 (10.8)	10.6 (16.5)	29.1 (24.5)	
<i>Diary</i>	5.9 (9.2)	5.3 (5.9)	5.1 (6.5)	16.4 (18.2)	$\chi^2 = 4.71^*$

<sup>a</sup>For the ACASI Monthly and SAQ Monthly conditions, the condition means are calculated as the averages of the model-based conditions means over the assessments at 1st, 2nd, and 3rd months.

<sup>b</sup>For ACASI Quarterly  $N = 20$ , SAQ Quarterly  $N = 26$ , ACASI Monthly  $N = 28, 27, 26$ , respectively, and SAQ Monthly  $N = 23, 23, 22$ , respectively.

<sup>c</sup> $\chi^2$  denotes the  $\chi^2$  statistic with the degree of freedom equal to 1.

\* $p < .05$ ; \*\* $p < .01$ .

GEE indicated that there were no interactions between frequency and method of assessment; however, several main effects were observed. First, monthly assessments yielded a more accurate account of unprotected vaginal sex,  $\chi^2(1) = 5.92, p < .05$ , and protected vaginal sex,  $\chi^2(1) = 7.55, p < .05$ ) relative to quarterly assessment,

which resulted in underreporting. Second, for unprotected oral sex, the quarterly assessment was more accurate, with monthly recall over reporting oral sex events,  $\chi^2(1) = 8.28, p < .01$ . It is interesting to note that all the marginal estimates of monthly recall are all positive and those of quarterly recall are all negative, indicating that monthly recall had a



**Table 3** Between-condition comparisons of accuracy of retrospective assessments

	Monthly mean (SD)	Quarterly mean (SD)	Monthly + quarterly mean (SD)
Protected vaginal sex			
ACASI	0.67 (10.81)	−8.15 (9.43)	−2.86 (11.08)
SAQ	1.57 (7.69)	−2.73 (9.73)	−0.71 (9.01)
ACASI + SAQ	1.06 (9.51) <sup>a</sup>	−5.09 (9.87) <sup>a</sup>	
Unprotected vaginal sex			
ACASI	−0.77 (7.32)	−3.37 (11.15)	−1.78 (8.98) <sup>c</sup>
SAQ	4.96 (6.53)	−1.85 (9.63)	1.42 (8.89) <sup>c</sup>
ACASI + SAQ	1.78 (7.49) <sup>b</sup>	−2.49 (10.21) <sup>b</sup>	
Unprotected oral sex			
ACASI	6.89 (12.97)	−2.95 (16.36)	4.38 (18.34)
SAQ	6.56 (13.68)	1.81 (14.68)	6.70 (23.52)
ACASI + SAQ	10.48 (23.84) <sup>d</sup>	−0.26 (15.45) <sup>d</sup>	

*Note.* Means in a row or column sharing alphabets (on the baseline) are significantly different, as detailed below. For all measures, positive values indicate overreporting and negative values indicate underreporting, relative to the diary assessment.

<sup>a</sup>Main effect of frequency of assessment:  $\chi^2(1) = 7.55, p < .01$ .

<sup>b</sup>Main effect of frequency of assessment:  $\chi^2(1) = 5.92, p < .05$ .

<sup>c</sup>Main effect of method:  $\chi^2(1) = 5.25, p < .05$ .

<sup>d</sup>Main effect of frequency of assessment:  $\chi^2(1) = 8.28, p < .01$ .

tendency to overreport and quarterly recall had a tendency to under report the events.

The only significant main effect of method was for unprotected vaginal sex, with ACASI under- and SAQ overreporting this sexual event,  $\chi^2(1) = 5.25, p < .05$ . However, the absolute magnitude of this difference was trivial. The main effect of method was not significant for protected vaginal and unprotected oral sex. As depicted in Table 3, both ACASI and SAQ biased the recall in the same direction; that is, ACASI and SAQ underreported the events for protected vaginal sex, whereas both methods over-reported the events for the unprotected oral sex.

#### Predictor analyses

The final set of analyses evaluated whether the hypothesized predictors of self-report accuracy have any effect on accuracy of retrospective recall. For these analyses, we used a regression model involving the main effects of groups and predictors and their interactions with inference based on the GEE. This approach identified predictors and examined whether they predicted accuracy differentially. With this sample size, we can detect a medium effect size of 0.28 with 0.80 power in a single-predictor regression analysis with a two-sided  $\alpha = .05$  (Cohen, 1988). The effect sizes for the accuracy predictor analyses for the three sexual variables analyzed ranged from .002 (social desirability for protected vaginal sex) to .16 (conscientiousness for unprotected oral sex). Although none of the hypothesized predictors (conscientiousness, depression, social desirability, or substance use) forecast the accuracy

of any of the retrospective recalls of the three sexual behaviors some of these findings may be the result of insufficient power. We then examined whether the baseline values of the three outcomes (that is, the frequency of the sexual behavior) predicted accuracy of retrospective reporting. Among the frequency of these sex behaviors, the study was sufficiently powered only for the baseline frequencies of unprotected oral sex, with an effect size of 0.30 and 0.43 for the main and interaction effects, respectively. We found main and interaction effects for unprotected oral sex (main effect of baseline unprotected oral sex,  $\chi^2(1) = 9.21, p < .005$ , and interaction with group,  $\chi^2(3) = 18.35, p < .001$ ). With posthoc analyses, we found that baseline unprotected oral sex predicted accuracy of retrospective recall with higher baseline values leading to less accurate recall of unprotected oral sex for the monthly ACASI and SAQ groups.

#### Discussion

This study examined the accuracy of two commonly-used methods of collecting retrospective behavioral data. Employing a randomized design, we compared participants' ability to recall sexual behaviors (i.e., vaginal sex with and without condoms, oral sex without condoms) recorded in daily diaries. Participants were unmarried women aged 18–21 years. Behaviors reported by this cohort suggest considerable risk for HIV and other STDs. During the 3 months of the study, almost one-half reported multiple or casual sex partners.

Consistent condom use was rare even among those young women with multiple partners.

This study provided mixed evidence regarding the differential accuracy of ACASI and SAQs for the assessment of sexual risk behavior. Although we predicted that ACASI would yield more accurate data than SAQ, this result was not obtained consistently across the three behaviors or across duration of assessments. Protected and unprotected vaginal intercourse are key outcomes in behavioral risk reduction interventions for STDs, yet retrospective versus diary recalls differed on each of these behaviors in regard to what method and what recall period were most accurate. The recall error differed for the ACASI quarterly group which underestimated protected vaginal sex, the SAQ monthly group which overestimated unprotected vaginal sex, and the ACASI monthly and SAQ quarterly groups which overestimated unprotected oral sex. Based on between-condition comparisons, frequency of assessment affected the accuracy of retrospective recall across the three behaviors, with the monthly assessment yielding more accurate reporting for the two behaviors most related to HIV and STDs, namely, protected and unprotected vaginal sex. There was also a difference between the methods for assessing unprotected oral sex, with ACASI under- and SAQ overestimating the events. For protected and unprotected oral sex, there was no significant difference between the methods; both methods biased the recall of events in the same direction. Although these findings differ from those of previous studies, including that by Graham *et al.* (2003) in which only vaginal intercourse was overestimated in the 3-month recall, no other study has compared two methods (ACASI vs. SAQ) and two different durations of recall (monthly vs. quarterly), and used the daily diary as the gold standard. Other studies which have compared ACASI and SAQ with adolescents reported increased self-reports of risk behaviors with the inference being that these self-reports were, therefore, more accurate; these investigations did not employ diaries as a form of validation of these behaviors. In addition, only one of these studies had female adolescents in their sample. This study is the first to identify the small differences in precision that may exist between these methods and recall periods. However, because our findings provide mixed evidence, and because of the importance of replication in science, studies with larger, more diverse samples assessing these as well as other risk behaviors are needed before definitive conclusions regarding the differential accuracy of assessment methods can be reached.

The relationships of several hypothesized predictors of recall accuracy were also examined. Most of the effect sizes of the hypothesized predictors of recall accuracy were small and our sample size did not provide sufficient power to detect statistically significant findings. Only the effect size of the frequency of unprotected oral sex was large enough to allow statistical testing of the impact of this predictor on accuracy.

As hypothesized, less frequent sexual behavior (i.e., unprotected oral sex) was associated with greater recall accuracy. Those participants with higher frequency of this behavior at baseline tended to overestimate it on retrospective recall. Another factor that may influence recall accuracy is the regularity of behavior; that is, when individuals have a regular pattern such that they have sex the same day each week with the same partner, individual instances may blend together in memory. In contrast, new partners are likely to be more distinctive and increase recall accuracy. Future research might investigate the importance of this factor by coding diaries for regularity of sexual behaviors/sexual partners and examine the relationship to recall accuracy. Future work might also include assessments of the other predictors with larger samples to test those that had small effect sizes. However, the magnitude of the effect sizes of these predictors of accuracy might reassure sexual health scientists regarding the impact of their effect on recall accuracy.

Several limitations of this study should be noted. First, the sample was restricted to young women. Second, frequency counts of anal sex and protected oral sex were low. Future research sampling populations for whom these behaviors are more frequent is needed. Third, because diaries were returned to us weekly, it is possible that participants provided weekly rather than daily data. However, our instructions coupled with input from participants who called to ask for clarification on how to document their behavior for that day, use of different writing instruments, and comments from participants describing how they developed reminder systems to document their daily behaviors suggest that diaries were completed daily. Fourth, our protocol required that participants note whether or not sexual behavior occurred on a daily basis, and later to recall (retrospectively) those same behaviors. It is possible that maintaining a diary led participants to encode sexual event memories with greater strength relative to a purely retrospective method. This limitation is impossible to avoid without employing observers or other intrusive validation strategies. Fifth, the internal consistency on the social desirability scale was only moderate ( $\alpha = .63$ ). Thus, measurement of this construct may be imprecise and should be considered when interpreting these results. Lastly, our ability to detect small predictor effects was limited by the size of our sample size. However, none of these psychological predictors of accuracy had effect sizes larger than 0.16 suggesting that their impact may be minor.

In conclusion, ACASI and SAQ are two frequently used modes to assess retrospective recall of sexual behavior. Questions posed by investigators in this area are: (a) over what period of time can participants recall such personal behaviors, (b) is one mode of data collection superior to the other, and (c) do personality variables affect recall accuracy? This study provides evidence from young women that the answer to the

first two questions may vary depending on the behavior one is asked to recall. Findings from these 102 participants suggest that episodes of protected and unprotected vaginal sex may best be recalled using an ACASI mode and a recall period of 1 month. In regard to the third question, conscientiousness, depression, social desirability, substance use, and frequency of sexual behaviors (except for unprotected oral sex) did not predict greater or lesser accuracy in this group of young women. The small effect sizes may indicate that their impact on recall accuracy will be modest, in any event.

The choice of a specific method for data collection cannot be made based only on these findings. Instead, investigators and clinicians need to consider the population being sampled and the context in which data collection will occur. Our findings do suggest that the daily diary recall is a useful tool for both methodological as well as substantive questions related to sexual and substance use behaviors. Although implementing diary recalls in a study protocol requires time and effort for both the investigator as well as the participants, similar to work by Jaccard *et al.* (2002), participants in this study completed the diaries conscientiously with few missing data. Effect sizes of common psychological predictors of accuracy (i.e., conscientiousness, depression, substance use, social desirability) were small and may, therefore, have limited impact on recall accuracy. Continued investigation is needed if we are to accurately measure important outcomes of interventions that cannot be assessed in ways other than by retrospective recall.

**Acknowledgments** This research was funded by the National Institutes of Health (R01 NR008194 to DMB) as well as the National Institute of Mental Health (K02 MH01582 to MPC). The authors would like to thank the participants of the study for their important contribution to HIV-prevention science.

## References

- Aquilino, W. S., Wright, D. L., and Supple, A. J. (2000). Response effects due to bystander presence in CASI and paper-and-pencil surveys of drug use and alcohol use. *Substance Use and Misuse*, *35*, 845–867.
- Carey, M. P., Braaten, L. S., Maisto, S. A., Gleason, J. R., Forsyth, A. D., Durant, L. E., *et al.* (2000). Using information, motivational enhancement, and skills training to reduce the risk of HIV infection for low-income urban women: A second randomized clinical trial. *Health Psychology*, *19*, 3–11.
- Carey, M. P., Carey, K. B., Maisto, S. A., Gordon, C. M., and Weinhardt, L. S. (2001). Assessing sexual risk behavior with the timeline followback (TLFB) approach: Continued development and psychometric evaluation with psychiatric outpatients. *International Journal of STDs and HIV*, *12*, 365–375.
- Carey, M. P., Maisto, S. A., Kalichman, S. C., Forsyth, A. D., Wright, E. M., and Johnson, B. T. (1997). Enhancing motivation to reduce the risk of HIV infection for economically disadvantaged urban women. *Journal of Consulting and Clinical Psychology*, *65*, 531–541.
- Cheung, Y. B. (2002). Zero-inflated models for regression analysis of count data: A study of growth and development. *Statistics in Medicine*, *21*, 1461–1469.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd Ed.). Hillsdale, NJ: Erlbaum.
- Des Jarlais, D. D., Paone, D., Milliken, J., Turner, C. F., Miller, H. G., Gribble, J. N., *et al.* (1999). Audio-computer interviewing to measure risk behaviors for HIV among injecting drug users: A quasi-randomized trial. *Lancet*, *353*, 1657–1661.
- Durant, L. E., and Carey, M. P. (2000). Self-administered questionnaires versus face-to-face interviews in assessing sexual behavior in young women. *Archives of Sexual Behavior*, *29*, 309–322.
- Durant, L. E., Carey, M. P., and Schroder, K. E. E. (2002). Effects of anonymity, gender, and erotophilia on the quality of data obtained from self-reports of socially sensitive behaviors. *Journal of Behavioral Medicine*, *25*, 439–467.
- Fehnel, S. E., Bann, C. M., Hogue, S. L., Kwong, W. J., and Mahajan, S. S. (2004). The development and psychometric evaluation of the motivation and energy inventory (MEI). *Quality of Life Research*, *13*, 1321–1336.
- Furukawa, T., Hirai, T., Kitamura, T., and Takahashi, K. (1997). Application of the center for epidemiologic studies depression scale among first-visit psychiatric patients: A new approach to improve its performance. *Journal of Affective Disorders*, *46*, 1–13.
- Graham, C. A., Catania, J. A., Brand, R., and Canchola, J. A. (2003). Recalling sexual behavior: A methodological analysis of memory recall bias via interview using the diary as a gold standard. *Journal of Sex Research*, *40*, 325–332.
- Hays, R., Hayashi, T., and Stewart, A. (1989). A five-item measure of socially desirable response set. *Educational and Psychological Measurement*, *49*, 629–636.
- Hewett, P. C., Mensch, B. S., and Erulkar, A. S. (2004). Consistency in the reporting of sexual behaviour by adolescent girls in Kenya: A comparison of interviewing methods. *Sexually Transmitted Infections*, *80*(S2), ii43–ii48.
- Horn, J., Nelson, C. E., and Brannick, M. T. (2004). Integrity, conscientiousness, and honesty. *Psychological Reports*, *95*, 27–38.
- Jaccard, J., McDonald, R., Wan, C. K., Dittus, P. J., and Quinlan, S. (2002). The accuracy of self-reports of condom use and sexual behavior. *Journal of Applied Psychology*, *32*, 1863–1905.
- Jaccard, J., and Wan, C. K. (1995). A paradigm for studying the accuracy of self-reports of risk behavior relevant to aids: Empirical perspectives on stability, recall bias, and transitory influences. *Journal of Applied Social Psychology*, *25*, 1831–1858.
- Kalichman, S. C., Rompa, D., and Coley, B. (1996). Experimental component analysis of a behavioral HIV-AIDS prevention intervention for inner-city women. *Journal of Consulting and Clinical Psychology*, *64*, 687–693.
- Kissinger, P., Rice, J., Farley, T., Trim, S., Jewitt, K., Margavio, V., *et al.* (1999). Application of computer-assisted interviews to sexual behavior research. *American Journal of Epidemiology*, *149*, 950–954.
- Kurth, A. E., Martin, D. P., Golden, M. R., Weill, N. S., Heagerty, P. J., Spielberg, F., *et al.* (2004). A comparison between audio computer-assisted self-interviews and clinician interviews for obtaining the sexual history. *Sexually Transmitted Diseases*, *31*, 719–726.
- Lambert, D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics*, *34*, 1–14.
- Locke, S. E., Kowaloff, H. B., Hoff, R. G., Safran, C., Popovsky, M. A., Cotton, D. J., *et al.* (1992). Computer-based interview for screening blood donors for risk of HIV transmission. *JAMA*, *268*, 1301–1305.
- Lubin, B., and Van Whitlock, R. (2002). Development of a measure that integrates positive and negative affect and personality: The comprehensive personality and affect scales. *Journal of Clinical Psychology*, *58*, 1135–1156.

- Newman, J. C., Des Jarlais, D. C., Turner, C. F., Gribble, J., Cooley, P., and Paone, D. (2002). The differential effects of face-to-face and computer interview modes. *American Journal of Public Health, 92*, 294–297.
- Perlis, T. E., Des Jarlais, D. C., Friedman, S. R., Arasteh, K., and Turner, C. F. (2004). Audio-computerized self-interviewing versus face-to-face interviewing for research data collection at drug abuse treatment programs. *Addiction, 99*, 885–896.
- Radloff, L. S. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement, 1*, 385–401.
- Reading, A. E. (1983). A comparison of the accuracy and reactivity of methods of monitoring male sexual behavior. *Journal of Behavioral Assessment, 5*, 11–23.
- Roberts, R., Andrews, J., Lewinsohn, P., and Hops, H. (1990). Assessment of depression in adolescents using the center for epidemiologic studies depression scale. *Psychological Assessment, 2*, 122–128.
- Romer, D., Hornik, R., Stanton, B., Black, M., Li, X., Ricardo, I., et al. (1997). “Talking” computers: A reliable and private method to conduct interviews on sensitive topics with children. *Journal of Sex Research, 34*, 3–9.
- Schroder, K. E. E., Carey, M. P., and Venable, P. A. (2003). Methodological challenges in research on sexual risk behavior: II. Accuracy of self-reports. *Annals of Behavioral Medicine, 26*, 104–123.
- Turner, C. F., Ku, L., Rogers, S. M., Lindberg, L. D., Pleck, J. H., and Sonenstein, F. L. (1998). Adolescent sexual behavior, drug use, and violence: Increased reporting with computer survey technology. *Science, 280*, 867–873.
- Turner, C. F., Ku, L., Sonenstein, F. L., and Pleck, J. H. (1996). Impact of ACASI on reporting of male–male sexual contacts: Preliminary results from the 1995 national survey of adolescent males. In R. B. Warnecke (Ed.), *Health survey research methods: Conference proceedings* (pp. 996–1013). Hyattsville, MD: National Center for Health Statistics.
- Webb, P. M., Zimet, G. D., Fortenberry, J. D., and Blythe, M. J. (1999). Comparability of a computer-assisted versus written method for collecting health behavior information from adolescent patients. *Journal of Adolescent Health, 24*, 383–388.
- Weinhardt, L. S., Forsyth, A. D., Carey, M. P., Jaworski, B. C., & Durant, L. E. (1998). Reliability and validity of self-report measures of HIV-related sexual behavior: Progress since 1990 and recommendations for research and practice. *Archives of Sexual Behavior, 27*, 155–180.