Published in final edited form as:

Arch Sex Behav. 1998 April ; 27(2): 155–180.

Reliability and Validity of Self-Report Measures of HIV-Related Sexual Behavior:

Progress Since 1990 and Recommendations for Research and Practice

Lance S. Weinhardt, M.S., Andrew D. Forsyth, M.S., Michael P. Carey, Ph.D., Beth C. Jaworski, B.A., and Lauren E. Durant, M.S.

Syracuse University

Abstract

The trustworthiness of self-reported sexual behavior data has been questioned since Kinsey's pioneering surveys of sexuality in the United States (Kinsey et al. 1948, 1953). In the era of HIV and AIDS, researchers and practitioners have employed a diversity of assessment techniques but they have not escaped the fundamental problem of measurement error. In this article, we review the empirical literature produced since Catania et al.'s (1990) review regarding reliability and validity of self-administered and automated questionnaires, face-to-face interviews, telephone interviews, and self-monitoring approaches. We also provide specific recommendations for improving sexual behavior assessment. It is imperative that standardized self-report instruments be developed and used for sexual risk-behavior assessment.

Keywords

sexual behavior; HIV-risk behavior; self-report assessment; reliability; validity

The fidelity of sexual behavior data obtained by self-report has been questioned repeatedly since Kinsey's pioneering surveys of sexuality in the United States (Kinsey et al. 1948, 1953). In recent years, commentaries on sexual behavior research have suggested that behavioral data produced by self-report methods are worthless. For example, Lewontin (1995) argued that self-reports of sexual behavior are inherently unreliable and invalid due to multiple sources of bias, including under-reports of stigmatized behaviors and over-reports of normative behaviors. Brody (1995) questioned the validity of self-report data for sexual behaviors that confer risk for HIV infection, suggesting that participants in behavioral research are prone to intentional misrepresentation. Recognition of the importance of these and related concerns has led to a National Institutes of Health-sponsored conference on "The Science of Self-Report", held in Washington, DC, in November, 1996.

Despite these criticisms and concerns, researchers and practitioners continue to rely on self-report methods to assess the topography of sexual behavior because ethical and practical considerations limit the use of more direct assessment methods. The goals of this paper are to describe the state-of-the-science and to provide recommendations for improving self-report assessment methodology. We begin by summarizing methodological issues presented by Catania et al. (1990); next, we summarize the research published since their review; then we make specific suggestions for the assessment of sexual behavior in research and practice.

Background

Catania et al. (1990) published a thoughtful review of the methodological challenges faced by researchers attempting to assess sexual behavior. Issues such as participation and response biases (e.g., intentional misrepresentation, inaccurate recall) and the effects of different modes of administration on responding were discussed regarding their potential to introduce error into measurement of sexual behavior. Table I details the practical advantages and disadvantages of different modes of administration that emerge from their discussion.

Catania et al. (1990) concluded that although of these methodological issues have been studied extensively in the assessment of other behaviors, rigorous research on assessment of sexual behavior is needed before decisions about the best measures for various purposes can be made with confidence. Because self-report sexual behavior data are used for tracking the spread of HIV, identifying populations at risk for HIV infection, and evaluating the effectiveness of HIV-risk-reduction interventions, the authors implored researchers to balance the urgency of the HIV epidemic with the methodological rigor necessary for quality assessment. A foundation of developing measures is psychometric evaluation, which generally begins with evaluations of reliability and validity.

Assessing Reliability and Validity

Reliability—Establishing the reliability of a measure is prerequisite to assessing its validity. A common method for determining the reliability of self-reports of behavior is the test-retest reliability (TR) study, wherein the same measure is administered to the same participants twice and the results are compared for congruency. Well-designed TR studies can suggest some of the variables that affect reporting accuracy. Studies that compare the reliability of different modes of administration also aid in determining procedures that reduce inconsistent responding. Another method of examining the reliability of reports is to evaluate internal consistency. Behavioral questionnaires can include the same question more than one time, and the concordance of responses can be examined.

Validity—One ever-present issue in assessment of sexual behavior is the absence of a “gold standard” with which to compare self-report data. As of 1990, there were no current data on the distribution of high-risk sexual behavior in the population of the United States. Even if such normative data were available, it would be difficult to draw conclusions on an individual basis regarding whether a participant has accurately reported his or her sexual behavior. Unlike health-risk behaviors that are directly observable (e.g., use of a bicycle helmet), leave permanent byproducts (e.g., cigarette butts), or are relatively unencumbered by societal proscription (e.g., sedentary lifestyle behavior), risky sexual behaviors are inherently private, frequently considered taboo, and situationally specific, characteristics that make them inherently inaccessible to direct assessment strategies. The fact that there are no physiological or psychophysiological data that correspond directly to frequency of sexual activity compounds the problem. Due to these obstacles, there have been few studies designed to examine evidence for the validity of self-report sexual behavior measures. However, methods such as comparing self-reports with partner reports and comparing self-reports of risk behavior to HIV serostatus, reinfection with bacterial STDs and other biochemical markers for sexual activity, have been employed in several studies.

Review of the Empirical Literature Since 1990

Method

We searched the psychological, psychiatric, and medical literature to obtain studies examining specifically the reliability and validity of self-report sexual behavior measures. We began with

computerized searches of the psychological (PsycLIT) and medical (MedLine) databases. The following search terms were used: sexual behavior, HIV-risk behavior, reliability, validity, self-report, and assessment. In order to review research produced concurrent with and since Catania et al.'s (1990) review, we collected relevant articles published from January 1, 1990 to December 31, 1995 (or later if made available to us by the author). Articles were considered relevant if they described assessment of sexual behavior using a self-report method, and included analyses examining the reliability and/or validity of those self-reports. We then studied each reference section in those articles to identify additional relevant research. This process was repeated for each new reference until all cited references were included or eliminated because they provided redundant information. Table II describes the 30 studies included in this review.

Review of Published Research

Our first observation from this literature is that studies of self-report sexual behavior measures are methodologically heterogeneous and often at odds conceptually. Few articles document assessments of sexual behavior using the same questions, and the administration of many measures were unstandardized (see Darke et al., 1991; McKinnon et al., 1993; Needle et al., 1995 for exceptions). Studies were conducted with structured and semi-structured face-to-face interviews (FTFIs), self-administered questionnaires (SAQs), biological markers, and collateral interviews. Reporting periods ranged from three weeks to several years and the number of sexual behavior items included ranged from 3 to 398. Behaviors commonly assessed include numbers of sexual partners; frequency of protected and unprotected oral, anal, and vaginal sex; non-consensual sexual experiences; and history of relationships. Assessment duration ranged from 6 - 90 minutes (see Boekeloo et al., 1994; Konings et al., 1995, respectively). Most measures we examined were appropriate for risk screening (e.g., dichotomous indication of risk behavior) and risk assessment (e.g., continuous indication of level of risk behavior).

We summarize evidence for reliability and validity in Table II. Evidence for the reliability of measures was most often reported as temporal stability resulting from test-retest studies and internal consistency in studies using a single administration. However, there was variation in the conceptualization of reliability. For example, researchers have compared the consistency of scores using different reporting periods, in effect measuring different behaviors, yet presented their results as an evaluation of test-retest reliability (e.g., McLaws et al., 1990). Validity was often demonstrated using convergent evidence, with comparison to other self-report data as the most common procedure. Some researchers compared partner and self-reports (e.g., Upchurch et al., 1991; Padian, 1990), and others compared self-reports to STD incidence (Doll et al., 1994; Cohen and Dent, 1992; Zenilman et al., 1995).

The measures were used in research on a variety of subpopulations in the U.S. and abroad, including ethnic minority groups (Boekeloo et al., 1994; Cohen & Dent 1992; Dowling-Guyer et al., 1993; Kalichman et al., 1997; Schopper et al., 1993; Wyatt et al., 1992; Zenilman et al., 1995), psychiatric patients (McKinnon et al., 1993; Sacks et al., 1990), gay men (Doll et al., 1994; McLaws et al., 1990; Schneider et al., 1991; Siegel et al., 1994), college students (Anderson and Pollack, 1994; McEwan et al., 1992; Schneider et al., 1991), STD clinic patients (Cohen and Dent, 1992; Zenilman et al., 1995; Upchurch et al., 1991), sex traders (McLaws et al., 1990; Konings et al., 1995), and intravenous drug users (IDU; Dowling-Guyer, 1994; McElrath et al., 1994; Needle et al., 1995).

Eight studies directly examined the effects of administration mode on reported frequency of sexual behavior. Five compared different modes of assessment (Boekeloo et al., 1994; Kalichman et al., 1997; McEwan et al., 1992; Nebot et al., 1994; Siegel et al., 1994) and three compared different versions of the same self-report assessment mode (Kauth et al., 1991;

Konings et al., 1995; Weinhardt et al., in press). These studies are detailed in Table II. The following descriptions highlight four of the studies that examined different modes of assessment.

Noteworthy Investigations

Boekeloo et al. (1995) conducted a mode effect study of HIV-risk assessment that examined differences in reported risk behaviors between participants in a self-administered questionnaire condition and participants who responded to the same questions presented by audiotape. After the initial assessment, all participants completed face-to-face interviews based on the same questionnaire items. Boekeloo et al. concluded that audiotape administration of a culturally sensitive sexual behavior measure was preferable to written or face-to-face interview versions of the same measure because it resulted in fewer missing responses for several behaviors, including unprotected vaginal sex with steady or nonsteady partners, unprotected receptive anal sex with steady or nonsteady partners, multiple partners, and sex with a homosexual or bisexual man. The audiotape administered questionnaire also resulted in more reports of risk factors including unprotected vaginal sex with steady or nonsteady partners, unprotected receptive anal sex with steady partners, and sex with an HIV positive partner. Although methodological confounds (e.g., length of assessment covaried with assessment mode) probably influenced their results, Boekeloo et al.'s article serves as a model for mode-effects studies in applied environments. This study also highlights the potential advantages of using technologically advanced assessment methods.

Kalichman et al. (1997) examined differences in response rates, reliability, and changes in HIV-preventive behavioral intentions between groups of women who were administered an HIV-risk behavior assessment that was either self-administered, displayed on an overhead projector, or conducted as a face-to-face interview. In each condition, participants were administered long or short forms of the measure. Results indicate that: (a) effects of the interview and SAQ method were similar, and both enhance sensitization, whereas the lack of time to reflect during the projected assessment, resulted in minimized sensitization; and (b) longer assessments may have greater sensitizing effects (which may prepare people to change their behavior).

McEwan et al. (1992) compared a postal survey to a face-to-face interview. Results indicated that most responses and participation rates were similar across modes, but that the interviews were more costly more to administer. However, the interviews yielded better responses to open-ended questions and resulted in fewer reported "socially unacceptable" behaviors. Although the two assessments were administered to different samples, the authors concluded that postal questionnaires are better for obtaining information on sexual behavior if in depth answers are not needed. McEwan et al. recommended the postal method for most surveys in the HIV/AIDS field.

Finally, Kauth et al. (1991) reviewed duration of retrospective reporting periods used in sexual behavior research and found that many studies used reporting periods of 12 months or longer, and that reporting period varied widely across studies. The authors tested reliability of 2-week, 3-month, and 1-year reporting and found that participants tended to report reliably at 2-weeks and 3-months, but not at 12 months. Men tended to underreport all behaviors using the 12-month reporting period. The authors recommended the use of briefer reporting periods to improve the reliability of self-reports.

Gaps in the Literature

Few studies with direct implications for future assessment have been conducted. Most studies that have examined the psychometric properties of self-report sexual behavior measures have

done so in the context of larger empirical investigations, where evidence of reliability and validity is reported only as part of the preliminary analyses. Although more attention has been paid to measurement error in sexual behavior research, it remains as challenging today to select an empirically evaluated self-report measure of sexual behavior as when Catania et al. (1990) issued their call to action. Further, there is little consensus among researchers regarding which administration modes yield the most reliable and valid sexual behavior data. Consequently, contemporary sexual behavior research may employ measures that yield low fidelity data. Additional research focusing on modes of administration with diverse populations will be necessary to reach such a consensus.

In addition to study design, greater attention to statistical analyses of self-report data is worthy of attention. For example, optimal performance of standard tests of correlation and mean differences depend upon a number of important conditions that, when violated, limit confidence in the results. Among these is the assumption that observed scores come from normal distributions. It is widely recognized that the distributions for many sexual behaviors tend to be positively skewed; many people report zero or a few occasions of target behaviors whereas other participants report high frequencies. In some cases, transforming scores prior to examining their covariation with other variables may be appropriate if the transformation produces a distribution that approximates the normal (see Tukey, 1977, for a thorough discussion of data transformation). However, data for these behaviors are often sufficiently non-normal to suggest the need for alternative analytical approaches (e.g., negative binomial regression; Gardner et al., 1995) that are appropriate for highly skewed, count data. Other issues worthy of attention include modeling data that contain a disproportionate number of specific values relative to the rest of the distribution (e.g., zeros or ones), the implications of relying on least squares estimation for modeling low base rate behavior, and the elaboration of alternative analytical strategies for modeling non-normally distributed, sexual behavior data. Clearly, sexual behavior data require more sophisticated analytic procedures that yield test statistics that can be interpreted with greater confidence.

Recommendations

The quality of sexual-behavior assessment is an important but daunting challenge to HIV-related research. From the research produced to date, there is evidence suggesting that well-designed interviews and questionnaires can provide acceptable data when administered appropriately. However, the problematic issues discussed herein (see also Catania et al., 1990, 1995), coupled with the public health implications of the use of sexual self-report data, require active research that focuses on the design and evaluation of measures for a variety of populations. To improve the quality of self-report assessment of sexual behavior, we offer the following recommendations, which are based on the reviewed studies, information from literacy experts, and our experiences with survey design and administration. These recommendations are likely to be useful for selecting among existing measures, adapting existing measures for different applications, or designing new measures.

Use psychometrically evaluated measures

Table II provides psychometric evidence for the latest measures developed. Research on the Risk Behavior Assessment (RBA; Needle et al., 1995), which includes sexual and drug-related HIV risk factors, provides an example of how psychometric evaluation can be conducted systematically. Psychometric properties of the RBA measure have been evaluated with different populations. When adapting a measure for use in a population, or when creating a new measure, it is crucial to conduct appropriate psychometric evaluation. In studies reporting test-retest analyses, the focus should be on the consistency of reporting rather than the consistency of target behaviors. Consequently, investigators should be careful to assess behaviors occurring in the same reporting period at both administrations. As others have

pointed out, shorter assessment intervals (e.g., one week) limit the impact that memory deterioration and other variables may have on retest coefficients.

Validate the measure

There is an urgent need for a standardized method for validating findings from sexual behavior self-reports. One approach, suggested by Schopper et al., (1993), is to (i) use the concordance of numbers of regular and casual partners and sex acts across gender to identify gaps; (ii) use the degree of concordance between couples on the type of relationship (e.g., monogamous or polygamous), and on number of sex acts with the partner to measure accuracy at the individual level; and (iii) compare patterns related to participant demographics and behavior with results from other surveys with similar samples. Seidman and Reider (1994) reviewed sexual behavior surveys conducted in the United States, and provided data on “normative” sexual behavior across the life cycle. A more comprehensive survey has recently been reported by Laumann et al. (1994). This normative data may allow researchers to compare aggregate patterns of behavior in their samples with appropriate population parameters. Other methods can be used to provide evidence for validity of self-report measures are biochemical evidence of sexual activity (e.g., reinfection with STDs) and comparison with concurrent self-monitoring data.

Use a measure appropriate for the purpose of the assessment

Because measurement instruments address different levels of behavioral specificity, attention to the needs of the assessment procedure are warranted. It may be useful to conceptualize measures of HIV-risk behavior as falling into three concentric categories: risk screening, risk assessment, and risk-event data. Risk screening describes dichotomously whether a respondent has engaged in risk behavior during the reporting period (e.g., Boekeloo et al., 1994). Risk assessment allows description of the level of risk, based on the frequency of risk behavior (e.g., Downey et al., 1995). Risk-event data is the most detailed, and allows event-level examination of the co-occurrence of potential risk factors with risk behaviors (e.g., Crosby et al., 1996). Clearly, a measure designed for risk-screening may yield data that are inadequate if the goal of an investigation is to assess the level of risk reported by the target sample. Alternatively, although an event-level assessment provides risk screening data, the additional resources required for assessment would be wasted if event-level analysis is not required.

Use language that is easily understood

Researchers interested in studying at-risk behaviors among under-served samples are likely to encounter problems associated with the terminology used in assessments. On the one hand, terms that are unfamiliar to participants may lead to increased measurement error, and on the other hand, using slang terms from the argot of the target sample may reduce researchers' credibility or be misinterpreted as condescension. One solution is to provide parenthetical descriptions of clinical terms. For example, “How many times in the past three months have you had vaginal sex (when a man puts his penis in a woman's vagina)?” Careful pilot research with the target sample prior to measurement development and implementation is essential. Specifically, the literacy level of the population should be assessed, and used to guide vocabulary used in the questionnaire. Oral and written instructions should be as concise as possible, and should be reviewed with participants before they begin the questionnaire. If multiple questionnaires are used, response formats should be as consistent as possible. Finally, offering an audiotape-administered version may be advantageous for groups with low reading levels.

Use focus groups, pilot data, and other formative methods to adapt the assessment protocol for sensitivity to cultural issues of the participants

Just as consideration of the reading and comprehension level of the target sample is necessary, so is attention to the lifestyles, prevailing customs, and religious and cultural traditions of target samples. For example, sexual behavior questionnaires should be designed without inherent sexual orientation bias. Caution should be exercised to include response options that are suitable for a wide variety of responses. Filler questions should be culturally relevant. Also, emphasizing the personal benefits of the research to participants and their community is likely to engender cooperation and elicit candid reporting. Conducting focus groups with men and women from the population to be assessed often reveals unanticipated cultural and contextual issues that can be used to inform sexual behavior assessment (Carey et al., 1997). The need for qualitative research during measurement development cannot be over-emphasized.

Include techniques that improve recall of behavior

Croyle and Loftus (1993) detailed the influence of the constructive nature of memory on self-reports of sexual behavior. Simple forgetting, telescoping (distorting the recency of particularly memorable events), exposure to misleading information since the event, and the use of heuristics to estimate behavior frequencies are some of the factors that can contribute to inaccurate self-reports. Useful techniques include: (a) providing anchor dates for reporting periods, (b) encouraging participants to use appointment books and calendars to recall other memorable events during the reporting period, and (c) recalling extensive periods of abstinence or consistent sexual activities. Recent use of the timeline follow-back procedure (Sobell & Sobell, 1996), which utilizes many of these techniques, to assess sexual behavior have shown that it provides valuable event-level data (Crosby et al., 1996) and that these self-reports are reliable (Weinhardt et al., in press).

Establish a working trust with interviewees and questionnaire respondents

Assessment of risk should take place after a participant and interviewer have established rapport, and the interviewer has assured the participant of confidentiality. Specific sexual behavior assessment should always begin with an appropriate introduction for the participants (Carey, 1998). During this time the reasons for asking questions about sexual (and other socially sensitive) behaviors should be provided. For example, in a clinical context, one might say that a standard practice is to inquire about risk for HIV just as one routinely inquires about suicidal ideation, personal safety, and other important matters; thus, all participants get asked and no one will feel singled out as being at unique risk. In a public health or research context, investigators might identify the overall purpose of the research, including how this will improve public health or enhance the scientific knowledge base. After these introductory remarks, participants should be invited to ask any questions they might have.

Ask questions in a direct fashion, without apology or hesitancy (Kinsey et al., 1948)

If an interviewer or investigator appears embarrassed about or unsure of the appropriateness of the questions, participants may detect this and provide incomplete or ambiguous responses.

Adopt default assumptions to gather the most accurate information efficiently (Wincze & Carey, 1991)

These assumptions reflect the preferred direction of error. Thus, for example, it is better to assume minimal understanding on the part of the participant so that language is clear and concrete. Other useful assumptions include: (i) participants will be embarrassed about and have difficulty discussing sexual matters; (ii) participants will not understand all sexual behavior terms, medical terminology, etc.; (iii) participants will be misinformed about sexual material, including STDs and other threats to sexual health; and (iv) questions will provoke concern

regarding a participant's health or well-being (thus, the interviewer should be prepared with information, referral sources, and other reassuring materials). As the interviewer or investigator learns more about the client, these assumptions can be adjusted.

Sequence the inquiry from the least to most threatening questions

Thus, questions about oral or vaginal sex might precede questions regarding sexual behaviors that are less socially approved (e.g., anal intercourse, fisting, or rimming).

Place the “burden of denial” on the participant (Kinsey et al., 1948)

Rather than ask “if” a clients has engaged in a particular activity, ask “how many times have you ...” engaged in it. Use of open response formats on questionnaires is encouraged (Catania et al., 1990). Such an approach will communicate an investigator's (or clinician's) expectation that such behaviors do occur and are not abnormal.

Be sensitive to contextual issues in administration

Among these factors are the experimenter's demeanor during interactions, administration setting, relevance of the studies aims to participants' lives, and perceptions of trust regarding the purpose and personnel of the study. Clearly, these “non-specific” variables may impact participants' the assessment protocol and affect measurement error.

Additional studies employing these steps will increase understanding of the issues affecting sexual behavior self-reports, and facilitate the development of standardized self-report measures that are appropriate for specific purposes in a variety of populations. Researchers and practitioners can then place more confidence in sexual behavior data gathered with self-report methods.

References

- Anderson SL, Pollack RH. The making of an oral sex questionnaire. *J Sex Educ. Ther* 1994;20:123–133.
- Boekeloo BO, Schiavo L, Rabin DL, Conlon RT, Jordan CS, Mundt DJ. Self-reports of HIV risk factors by patients at a sexually transmitted disease clinic: Audio vs. written questionnaires. *Am. J. Public Health* 1994;84:754–759. [PubMed: 8179044]
- Brody S. Patients misrepresenting their risk factors for AIDS. *Int. J. STD. AIDS* 1995;6:392–398. [PubMed: 8845395]
- Carey, MP. Assessing and reducing risk of infection with the human immunodeficiency virus (HIV). In: Koocher, GP.; Norcross, JC.; Hills, SS., editors. *Psychologist's desk reference*. Oxford University Press; New York: 1998.
- Carey MP, Gordon CM, Morrison-Beedy D, McLean DA. Low-income women and HIV-risk reduction: Elaborations from qualitative research. *AIDS and Beh* 1997;1:163–168.
- Catania JA, Binson D, Van Der Straten A, Stone V. Methodological research on sexual behavior in the AIDS era. *Ann. Rev. Sex Res* 1995;6:77–125.
- 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;108:339–362. [PubMed: 2270232]
- Cohen DA, Dent C. The validity of self-reported condom use. *Am. J. Public Health* 1992;82:1563–1564. [PubMed: 1443312]
- Crosby GM, Stall RD, Paul JP, Barrett DC, Midanik LT. Condom use among gay/bisexual male substance abusers using the timeline follow-back method. *Addict. Beh* 1996;21:249–257.
- Croyle, RT.; Loftus, EF. Recollection in the kingdom of AIDS. In: Ostrow, DG.; Kessler, RC., editors. *Methodological issues in AIDS behavioral research*. Plenum; New York: 1993.
- Darke, Hall, Heather, Ward, Wodak. The reliability and validity of a scale to measure HIV risk-taking behaviour among intravenous drug users. *AIDS* 1991;5:181–1185. [PubMed: 2031690]

- Davoli M, Perucci CA, Sangalli M, Brancato G, Dell'Uomo G. Reliability of sexual behavior data among high school students in Rome. *Epidemiology* 1992;3:531–535. [PubMed: 1420521]
- Doll IS, Harrison JS, Frey RL, KcKirnan D, Bartholow BN, Douglas JM, Joy D, Bolan G, Doetsch J. Failure to disclose HIV risk among gay and bisexual men attending sexually transmitted disease clinics. *Am. J. Prev. Med* 1994;10:125–129. [PubMed: 7917436]
- Dowling-Guyer S, Johnson ME, Fisher DC, Needle R, Watters J, Andersen M, Williams M, Kotranski L, et al. Reliability of drug users' self-reported HIV risk behaviors and validity of self-reported drug use. *Assessment* 1994;1:383–392.
- Downey L, Ryan R, Roffman R, Kulich M. How could I forget? Inaccurate memories of sexually intimate moments. *J. Sex Res* 1995;32:177–191.
- Gardner W, Mulvey EP, Shaw E. Regression analyses of count and rates: Poisson, overdispersed Poisson, and negative binomial models. *Psychol. Bull* 1995;118:392–404. [PubMed: 7501743]
- Geisser ME, Jefferson TW, Spevak M, Boaz T, Thomas RG, Murray FT. Reliability and validity of the Florida Sexual History Questionnaire. *J. Clin. Psychol* 1991;47:519–528. [PubMed: 1939696]
- Kalichman SC, Kelly JA, Stevenson LY. Priming effects of HIV risk assessments on related perceptions and behavior: An experimental field study. *AIDS and Behav* 1997;1:3–8.
- Kauth MR, 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. Prev* 1991;3:207–214. [PubMed: 1834142]
- Kinsey, AC.; Pomeroy, WB.; Martin, CE. Sexual behavior in the human male. W. B. Saunders; Philadelphia: 1948.
- Kinsey, AC.; Pomeroy, WB.; Martin, CE.; Gebhard, PH. Sexual behavior in the human female. W. B. Saunders; Philadelphia: 1953.
- Konings E, Bantebya G, Carael M, Bagenda D, Mertens T. Validating population surveys for the measurement of HIV/STD prevention indicators. *AIDS* 1995;9:375–382. [PubMed: 7794542]
- Lewontin RC. Sex, lies, and social science. *New York Review of Books* 1995;42:24–29.
- Laumann, EO.; Gagnon, JH.; Michael, RT.; Michaels, S. The Social Organization of Sexuality: Sexual Practices in the United States. University of Chicago; Chicago: 1994.
- McElrath K, Chitwood DD, Griffin DK, Comerford M. The consistency of self-reported HIV-risk behavior among injection drug users. *Am. J. Public Health* 1994;84:1965–1970. [PubMed: 7998638]
- McEwan RT, Harrington BF, Bhopal RS, Madhok R, McCallum A. Social surveys in HIV/AIDS: telling or writing? A comparison of interview and postal methods. *Health Education Research* 1992;7:195–202. [PubMed: 10148739]
- McKinnon K, Cournos F, Meyer-Bahlburg HFL, Guido JR, Caraballo LR, Margoshes ES, Herman R, Gruen R, Exner TM. Reliability of sexual risk behavior interviews with psychiatric patients. *Am. J. Psychiatry* 1993;150:972–974. [PubMed: 8494081]
- McLaws ML, Oldenburg B, Ross MW, Cooper DA. Sexual behavior in AIDS-related research: Reliability and validity of recall and diary measures. *J. Sex Res* 1990;27:265–281.
- Nebot M, Celentano DD, Burwell L, Davis A, Davis M, Polacsek M, Santelli J. AIDS and behavioral risk factors in women in inner city Baltimore: A comparison of telephone and face to face surveys. *J. Epidem. Comm. Health* 1994;48:412–418.
- Needle R, Fisher DG, Weatherbee N, Chitwood D, Brown B, Cesari H, Booth R, Williams ML, Watters J, Andersen M, Braunstein M. Reliability of self-reported HIV risk behaviors of drug users. *Psychol. Addict. Beh* 1995;9:242–250.
- Padian NS. Sexual histories of heterosexual couples with one HIV-infected partner. *Am. J. of Public Health* 1990;80:990–991. [PubMed: 2368866]
- Sacks MH, Perry S, Graver R, Shindledecker R, Hall S. Self-reported HIV-related risk behaviors in acute psychiatric inpatients: A pilot study. *Hosp. Comm. Psychiatry* 1990;41:1253–1255.
- Schneider DJ, Taylor EL, Prater LM, Wright MP. Risk assessment for HIV infection: Validation study of a computer assisted preliminary screen. *AIDS Educ. Prev* 1991;3:215–229. [PubMed: 1931425]
- Schopper D, Doussantousse S, Orav J. Sexual behaviors relevant to HIV transmission in a rural African population: How much can a KAP survey tell us? *Soc. Sci. Med* 37:401–412. [PubMed: 8356488]

- Seha A, Klepp KI, Ndeki SS. Scale reliability and construct validity: A pilot study among primary school children in Northern Tanzania. *AIDS Educ. Prev* 1994;6:524–534. [PubMed: 7702963]
- Seidman SN, Rieder RO. A review of sexual behavior in the United States. *Am. J. of Psychiatry* 1994;151:330–341. [PubMed: 7619092]
- Siegel K, Krauss B, Karus D. Reporting recent sexual practices: Gay men's disclosure of HIV risk questionnaire and interview. *Arch. Sex. Beh* 1994;23:217–230.
- Sobell, LC.; Sobell, MB. Timeline followback user's guide. Alcohol Research Foundation; Toronto: 1996.
- Solstad K, Davidsen M. Sexual behaviour and attitudes of Danish middle-aged men--methodological considerations. *Maturitas* 1993;17:139–149. [PubMed: 8231905]
- Taylor JF, Rosen RC, Lieblum SL. Self-report assessment of female sexual function: Psychometric evaluation of the brief index of sexual functioning for women. *Arch. Sex. Beh* 1994;23:627–643.
- Tripp JH, Mellanby AR, Phelps FA, Curtis HA, Crichton NJ. A method for determining rates of sexual activity in schoolchildren. *AIDS Care* 1994;6:453–457. [PubMed: 7833363]
- Tukey, JW. *Exploratory Data Analysis*. Addison-Wesley; Reading, MA: 1977.
- Upchurch DM, Weisman CS, Shepherd M, Brookmeyer R, Fox R, Celentano D, Colletta L, Hook EW. Interpartner reliability of reporting of recent sexual behaviors. *Am. J. Epidemiol* 1991;134:1159–1166. [PubMed: 1746526]
- Weinhardt LS, Carey MP, Maisto SA, Carey KB, Cohen MM, Wickramasinghe SM. Reliability of the timeline followback sexual behavior interview. *Ann. Behav. Med.* in press
- Wincze, JP.; Carey, MP. *Sexual Dysfunction: A Guide for Assessment and Treatment*. Guilford; New York: 1991.
- Wyatt GE, Lawrence J, Vodounon A, Mickey MR. The Wyatt sex history questionnaire: A structured interview for female sexual history taking. *J. Child Sex. Abuse* 1992;1:51–67.
- Zenilman JM, Weisman CS, Rompalo AM, Ellish N, Upchurch DM, Hook EW, Celentano D. Condom use to prevent incident STDs: The validity of self-reported condom use. *Sex. Transm. Dis* 1995;22:15–21.

Table I**Summary of Practical Advantages and Disadvantages of Modes of Self-Report Sexual Behavior Assessment**

Assessment Mode	Advantages	Disadvantages
Self-Administered Questionnaire (SAQ)	Privacy Efficiency	Literacy dependent No potential to probe responses
Face-to-Face Interview (FTFI)	Ability to build rapport Ability to probe ambiguous, internally inconsistent, or non-response Ability to explain unfamiliar terms Enhanced credibility Minimizes non-responding	Less privacy (especially with household interviews) Reaction to interviewer Inefficiency Possibility of inaccurate interpretation of responses by interviewer
Telephone Interview	Ability to probe ambiguous, internally inconsistent, or missing-responses Ability to explain unfamiliar terms More privacy than FTFI (no visual contact) Less interpersonal reaction to interviewer	Obscene phone-caller phenomenon (less credibility) Many high risk individuals are homeless or transient and have no phone
Postal Questionnaire	Privacy Efficiency	Literacy dependent Less credibility Low response rate
Self-Monitoring	Minimizes effects of memory error: Less time between event and recall	Reactive - responding may impact sexual behavior and reporting of sexual behavior. Requires more effort and commitment from participant
Audiotape administered interview / questionnaire	Privacy No interpersonal reaction to interviewer	Literacy dependent No potential to probe responses Relatively difficult to repeat questions Reaction to voice on tape: Who is this person, and how would they judge me?
Computer administered interview / questionnaire	Ability to probe ambiguous, internally inconsistent, or missing-responses Ability to explain unfamiliar terms Privacy	Literacy dependent Expense and inefficiency: One participant per computer Level of comfort/familiarity with computers may effect responses

Table II
Studies Reporting Evidence of Reliability or Validity of Sexual Behavior Self-Report Measures (1990-1995)^a

Study	Participants	Sexual Behaviors Included in Analysis / Response Format	Mode(s) / Reading Level	Reporting period	Number of Items / Duration	Evidence for Reliability	Validation Method	Evidence for Validity	Comments
Boekeloo et al. (1994)	N = 305. Gender: 60% male. Age: M = 25. Race: 73% Af Am. Source: STD clinic.	Multiple male or female partners; homosexual or bisexual male partner; one night stand; prostitution with man or woman; IDU partner; alcohol and illicit drug use with sex; unprotected vaginal and anal sex with steady and nonsteady partner / "Yes" or "no"	FTFI, written SAQ, audio SAQ / NR	3 months	16 / written SAQ and FTFI: 6 min; Audio SAQ: 14 min.	NR	Compared audio SAQ and FTFI responses and written and FTFI responses.	No differences between audio SAQ and FTFI on 10 of 16 risk factors and between written SAQ and FTFI on 12 of 16 risk factors.	Participants less likely to report unprotected vaginal or receptive anal sex with a nonsteady partner, receptive anal sex with a regular partner, and alcohol use with sex, during FTFI than either audio or written SAQs.
Cohen & Dent (1992)	N = 509. Gender: 63% male. Age: NR Race: "pre-dominantly" Hispanic and Af. Am. Source: STD clinics.	Frequency of condom use; frequency of intercourse / "Always" or "not always", NR	SAQ / NR	NR	4 / NR	Internal consistency for 3 items regarding frequency of intercourse, $\alpha = .59$	Compared self-report with evidence of STD treatment in 9 months following SAQ.	Odds Ratio (95% CI) Men's frequency of sex and subsequent STD treatment: 1.2 (0.7, 2.1) Women's frequency of sex and subsequent STD treatment: 0.8 (0.4, 1.6). Frequency of condom use and STD treatment NR.	Assumes correspondence between failure to use condoms & STDs.
Darke et al. (1991)	N = 175. Gender: ~66% male. Age: M = 30. Race: NR Source: Opiate treatment and needle exchange program.	# partners, # occasions of anal sex; frequency of condom use with regular, casual, and paying partners / None, 1, 2, 3-5, 6-10, or >10 times; Every time, often, sometimes rarely, never.	FTFI, collateral reports / NA	NR	5 / NR	7 day test-retest $r = .86$. Internal consistency, $\alpha = .70$	Collateral reports	Exact agreement across 11 items ranged from 82% - 100%	Exact agreement provides a liberal test statistic.
Davoli et al. (1992)	N = 383. Gender: 70% female. Age: M = 16. Race: NR Source: High school classes in Rome.	Ever had sex, age of first partner, method of contraception / NR	FTFI and equivalent SAQ / NR	Lifetime	NR / NR	NR	Comparison of SAQ to FTFI. Half of sample randomly assigned to SAQ first, FTFI 2-4 weeks later. Other half FTFI then SAQ.	Agreement on "Have you ever had sexual intercourse" was $\kappa = 0.88$.	Few behaviors assessed, frequency not assessed
Doll et al. (1994)	N = 1,063. Gender: 100% male Age: M = 32. Race: Caucasian, African-American, Hispanic, Asian, & Native-American	Screening interview: items differed across sites / dichotomous: engaged in behavior or not FTFI: Frequencies of sexual behaviors	FTFI (screening SAQ used in one city) / NR	Screen: 4 months. FTFI: Lifetime and 4 months	Screen: NR / 5 minutes. FTFI: NR / NR	NR	Compared screening with FTFI for unprotected anal intercourse. Also compared rectal gonorrhea (RG) diagnosis with results of screening and FTFI.	29% of men who reported unprotected anal sex during the FTFI, did not during the screening (failed to disclose). Results consistent with unprotected oral sex. 8.2% who failed to disclose at screen tested positive for RG. 9.6% who	Authors assume FTFI was more valid measure. Predictors of failure to disclose unprotected anal sex during screening: membership in a gay-identified organization, more peer support for

Study	Participants	Sexual Behaviors Included in Analysis / Response Format	Mode(s) / Reading Level	Reporting period	Number of Items / Duration	Evidence for Reliability	Validation Method	Evidence for Validity	Comments
	Source: homosexual men attending STD clinics in 3 cities.							disclosed consistently tested positive for RG.	condom use, fewer episodes of unprotected anal sex, and no history of treatment for alcohol or drug abuse.
Dowling-Guyver et al. (1994)	N = 218. Gender: 74% male. Age: M = 40. Race: 49% Af. Am., 28% Caucasian; 18% Hispanic. Source: Multi-site, IDU or cocaine users not in treatment.	# of partners; # of drug-using partners; frequencies of protected and unprotected vaginal, oral (cunnilingus and fellatio), and anal sex / Open format	FTFI / NA	30 days	NR / 30-45 min.	48 hour test-retest, # of days having sex: $r = .89$, # of partners: $r = .92$. Specific sexual behaviors: r s ranged from .07-1.0. Median test-retest $r = .75$	NR	NR	Eleven of seventeen test-retest coefficients for sexual behaviors were greater than $r = .7$
Downey et al. (1995)	N = 87. Gender: male. Age: Median = 35. Race: NR Source: group HIV-risk counseling program.	# of occasions of protected (no condom) and unprotected insertive and receptive oral and anal penetration. Oral penetration, protected and unprotected, with or without ejaculation/	SAQ and weekly diary / NR	Diary: weekly retro-spective for 3 months; SAQ 3-month	NR / NR	NR	Compared weekly diary reports to 3-month SAQ	Several statistical methods used. For combination of ten items: Exact hits = 58%, Misses = 21.6%, Kappa = .37, Rank-order correlation = .82	Unprotected anal intercourse resulted in most discrepancies; protected sex resulted in fewest discrepancies. Higher frequency resulted in more discrepancies across behaviors. Includes discussion on determining concordance of data from different modes of self-report assessment.
Geisser et al. (1991)	N = 91. Gender: male Age: 24 - 83. Race: NR Source: Medical center outpatient clinic and "community service clubs".	Ejaculation, premature ejaculation, erectile disorder, frequency of intercourse, time of last intercourse / 6-point scales	FTFI / NA	NR	20 / NR	Internal consistency, $\alpha = .9$, split half reliability, Spearman Brown $r = .86$	NR	NR	Questionnaire for assessment of sexual dysfunction.
Kalichman et al. (1997)	N = 267. Gender: 100% women. Age: M = 33. Race: 100% African-American. Source: Inner-city community outreach.	Frequencies of vaginal, anal, and oral intercourse, with and without condoms; # of partners / Open format	SAQ, overhead facsimile projection, FTFI (short and long form of each measure) / Flesch = 6.8 - 6.9.	3 months; 2 weeks	Short form: 8; Long form: 16 / NR	Two week TR intertrial kappas: unprotected vaginal intercourse SAQ = .67, overhead = .58, FTFI = .72; vaginal intercourse SAQ = .62, overhead = .45, FTFI = .39; # of partners SAQ = .61, overhead = .41, FTFI = .71	NR	NR	The SAQ format was the most consistently reliable across behaviors. Compared different 2 week reporting periods; this assumes that behavior is consistent.
Kauth et al. (1991)	N = 61. Gender: male. Age: NR.	Frequencies of protected and unprotected oral &	SAQ / NR	Past two weeks, past three months, and past year.	NR / NR	Pearson r s between adjusted 2 week and 3-month, 3-month and 12-month, and 2-	NR	NR	Biweekly and 3-month data were multiplied by 26, and

Study	Participants	Sexual Behaviors Included in Analysis / Response Format	Mode(s) / Reading Level	Reporting period	Number of Items / Duration	Evidence for Reliability	Validation Method	Evidence for Validity	Comments
	Race: NR Source: HIV seropositive support group.	anal sex or contact, fisting, mutual masturbation, frottage, # of male partners / Open format.				week and 12-month reports: Consolidated safer activities: .90, .61, .31; unsafe activities: .91, .46, .10.			4 respectively to represent adjusted 12 month intervals. Higher frequencies for shorter recall periods compared to 12-month retrospective reports. Shorter time frames provided most consistent reports.
Konings et al. (1995)	N = 520, n=75 at Time 2. Gender: 61% Age: M = 24 Race: 100% African. Source: Ugandan household survey.	Intercourse with a non-regular sex partner, condom use during last sexual intercourse with a non sexual partner/ "yes" or "no"	FTFI - short, long and in-depth versions / NA	12 months	NR / short: 10-15 min.; long: 25-30 min.; in-depth: 90 min..	3 week intertrial interval; condom use (kappa=.1), No. of partners (kappa=.47)	Compared reports from 3 versions of FTFI. Known groups evidence: female prostitutes compared to demographically representative sample	In depth FTFI resulted in higher reports of # of sex partners, # of "non-regular" partners, and and prevalence of condom use, relative to short and long versions. Prostitution confirmed with in-depth FTFI, not short or long versions.	Demographic data were consistent across methods. Authors recommend use of SAQs for efficiency and standardization, but that they should be complemented with small qualitative studies to detect biases in interpreting results.
McElrath et al. (1994).	N = 366. Age: M = 35. Race: 62% Caucasian, 24% Af. Am., 13% Hispanic. Source: IDU treatment programs	# of sexual partners: condom use / Open format; never, rarely, sometimes, half, most, always.	FTFI / NA	6, 12, and 18 months	NR / NR	6-, 12-, or 18-month TR intertrial interval: kappas ranged from .25 to .65	NR	NR	Higher reliability for dichotomous questions, but still low kappas. Non-significant decay in memory comparing 6 month to 12 and 18 month intertrial interval.
McEwan et al. (1992)	Two samples: Postal survey (N = 345; Gender: 51% male), FTFI (N = 300; Gender: 60% male) Age: 18 - 22+. Race: NR. Source: college subject pool	"behaviors related to HIV" / NR	FTFI postal SAQ / NR	NR	4 / NR	NR	Compared counts from postal surveys with data from FTFI's	Postal-survey participants were more likely to report that "drinking too much has made me not take contraceptive precautions while having sex" (18%) than interview participants (13%)	Assessment method confounded with samples, although samples were similar. Authors concluded that postal methods should be used for most HIV/AIDS surveys.
McKinnon et al. (1993)	N = 24. Gender: 67% male Age: 18 - 59. Race: 50% Hispanic, 42% Af. Am., 8% Caucasian. Source: inpatient and day treatment psychiatric facilities.	# of sexual partners: # of sexual episodes; % of sexual episodes in which vaginal intercourse occurred; % of occasions of vaginal intercourse in which condoms were used / Open format	FTFI / NA	6-months	NR / 45 min.	One month TR intertrial interval kappas ranged from 0.61 to 0.89	NR	NR	High reliability even with lengthy reporting period and intertrial interval (M = 33 days).

Study	Participants	Sexual Behaviors Included in Analysis / Response Format	Mode(s) / Reading Level	Reporting period	Number of Items / Duration	Evidence for Reliability	Validation Method	Evidence for Validity	Comments
McLaws et al. (1990)	<u>N</u> = 30. <u>Gender</u> : male. <u>Age</u> : NR. <u>Race</u> : NR. <u>Source</u> : Australian brothel and response to advertisement.	Number of regular and casual partners, frequency of condom use, non-condom use, and withdrawal for anal insertive and receptive sex, frequency of masturbation of and by participant, and frequency of oral sex with ejaculation of partner by participant / Open format.	FTFI, daily diary / NR	4 weeks	NR / NR	48- to 72-hour intertrial interval test-retest with SAQ: Spearman r s range from .35 to .98. All behaviors but receptive anal sex with withdrawal and masturbation of subject by partner resulted in reliability of $r > .7$	Compared first SAQ to diary	Correlations ranged from $r = -.28$ (frequency of receptive anal sex using a condom) to $r = .99$ (frequency of insertive anal sex without a condom). 6 of 13 behaviors resulted in SAQ-diary correlations of $r > .6$.	Diary was given to participants after completing the SAQ. Diary data represented the month following the SAQ, whereas SAQ data represented data from the prior month.
Nebot et al. (1994)	Two samples. Telephone interview: <u>N</u> = 775. <u>FTFI</u> : <u>N</u> = 416. <u>Gender</u> : 100% female. <u>Age</u> : range 17-35. <u>Race</u> : 83% Af. Am. <u>Source</u> : Inner-city Baltimore	Condom use, drinking, using drugs, during last sexual episode. Two or more partners in last year. Asked STD in past year, participants about STD among any of last 5 partners / Dichotomous.	FTFI, Telephone interview / NR	Last partner, last year, and last 5 partners	NR / NR	Internal consistency	Comparison of responses from telephone interview participants with FTFI participants	FTFI participants, relative to telephone respondents, were more likely to report using alcohol (10.6 vs. 2.4%) and drugs (30.5 vs. 16.3%) during last sexual intercourse	Telephone respondents were older, better educated, more often married, less likely to live in subsidized housing, more likely to report HIV testing, a previous abortion, or surgical sterilization. Authors conclude that despite sample differences, both interviews yielded comparable sexual behavior data
Needle et al. (1995)	<u>N</u> = 209, <u>Gender</u> : 70% male. <u>Age</u> : <u>M</u> age = 38. <u>Race</u> : 66% Af. Am., 19% Hispanic, 11% Caucasian <u>Source</u> : "drug user hang-outs" in 5 cities.	# of days having sex; # of partners; # of IDU partners; frequency of vaginal sex, # of times condoms were used during vaginal sex / Open format.	FTFI	30 days	NR / 30 - 45 min	48 hour test-retest intertrial interval, kappas ranged from .75 to 1.00.	NR	NR	Assumes drug & sexual behavior is constant. Different intervals of sexual behavior examined.
Padian (1990)	<u>N</u> = 98 serodiscordant couples. <u>Gender</u> : 77% male. <u>Age</u> : NR <u>Race</u> : NR <u>Source</u> : Various	Frequency of vaginal and anal sex, and condom use / Open format.	FTFI / NR	A "typical" month	NR / NR	NR	Partner interviews	Partner agreement on frequency of vaginal intercourse: $r = .72$, anal intercourse = .44. "Ever had anal intercourse": kappa = .76. "Ever used condoms": kappa = .74.	Couples were aware of the content in advance of interviews. Typical sexual activity does not purport to measure actual behavior.
Sacks et al. (1990)	<u>N</u> = 113 <u>Gender</u> : 33% female. <u>Age</u> : <u>M</u> = 31.	Homosexual anal intercourse without condoms; vaginal intercourse without condoms with an injection drug user	SAQ / NR	Past 5 years	NR / NR	NR	Compared SAQ to HIV status from chart review.	Of 19 participants who were at self-reported risk, seven were HIV-seropositive.	42% (48 / 113) of the sample of psychiatric inpatients reported HIV-risk behaviors in the five years prior to the survey.

Study	Participants	Sexual Behaviors Included in Analysis / Response Format	Mode(s) / Reading Level	Reporting period	Number of Items / Duration	Evidence for Reliability	Validation Method	Evidence for Validity	Comments
	Race: 74 Caucasian, 22 Af. Am, 10 Hispanic. Source: Psychiatric inpatient ward	or HIV- positive partner; intercourse without condoms with a prostitute; unprotected heterosexual intercourse without condoms / “Yes” or “No”							
Schneider et al. (1991)	N = sample 1: 70, sample 2: 17, sample 3: 74. Gender: 1. Male; 2. NR; 3. NR. Age: NR. Race: NR. Source: 1. HIV testing facility; 2. US military reserve unit; 3. College subject pool.	# of partners, race of partners, frequency of sexual events; frequency of receptive anal intercourse, frequency of condom use / 6- point scale - reduced to “at risk” or “not at risk”	Computer administered SAQ / NR	Since 1980, in 3 year segments.	NR / NR	NR	Comparison with confirmed HIV serostatus	Computer program identified 87.5% of seronegative participants as “not at risk” for infection based on their self-reports.	Detailed discussion of statistical risk levels of specific behaviors, and problems with validating self-reports with HIV test results.
Schopper et al. (1993)	N = 1,486. Gender: 75% female. Age: M = 29, Race: 100% African. Source: Rural Ugandan household.	Age at first intercourse, history of premarital sex and extramarital sex / various	FTFI / NR	Lifetime	NR / NR	NR	Partner interviews; compared results to other KAP surveys in region.	86% of 392 couples agreed on being polygamous or monogamous. Partners reported equal total # of extramarital and marital sex acts, but men reported more frequent vaginal intercourse in the past month and number of casual partners in the past year.	Significant interviewer effects: Women interviewed by women reported more sex in past year, more casual sex in past month, and less desire to use condoms than those interviewed by men. Paper includes excellent discussion on validating results from sexual behavior surveys.
Seha et al. (1994)	N = 464. Gender: 51% male. Age: M = 14.. Race: 100% African (Northern Tanzania). Source: Fifth and sixth grade classes.	Lifetime intercourse; frequency of intercourse / “Yes” or “no”; 3-point scale.	SAQ / NR	Lifetime	72 / 40 min.	Internal consistency. 10% of participants gave inconsistent answers on the two sexual behavior items.	Construct validity. Evaluated relationships between sexual behaviors and other sections of the survey.	Sexual behavior related to (a) attitudes about AIDS ($r = -.24$, $p < .001$), (b) social norms related to AIDS ($r = -.17$, $p < .001$), and (c) alcohol use in the past 3 months ($r = .19$, $p < .001$)	Pilot test of World Health Organization’s standardized surveys among primary school children.
Siegel et al. (1994)	N = 154. Gender: male. Age: M = 34. Race: NR. Source: Homosexual men from New York City metro area.	Frequency of protected and unprotected oral and anal intercourse with and without ejaculation, fingering, fisting, rimming, and mutual masturbation / 4-point scale.	FTFI, SAQ / NR	6 months	NR / NR	NR	Compare between SAQ, FTFI, partner interview, and HIV serostatus.	Correlation between SAQ and FTFI: Fim’s $r = 0.51$; $p < .001$; SAQ and FTFI long-form scoring (from transcriptions) : $r = .73$. No significant relationship between serostatus and self-reports.	Relative to the SAQ, during the FTFI participants were less likely to report behaviors such as rimming, fisting, unprotected anal and oral sex. Authors recommended the use of combined interview/

Study	Participants	Sexual Behaviors Included in Analysis / Response Format	Mode(s) / Reading Level	Reporting period	Number of Items / Duration	Evidence for Reliability	Validation Method	Evidence for Validity	Comments
									questionnaire techniques.
Solstad and Davidsen (1993)	<u>N</u> = 439. Gender: male. <u>Age</u> : Time 1: <u>M</u> = 40. Time 2: <u>M</u> = 51 Race: NR <u>Source</u> : 4 Danish communities.	# of partners, # of male partners, ever paid for sex with a woman, masturbation / Mixed	FTFI, SAQ / NR	NR	FTFI: 40; SAQ: 4	NR	SAQ compared to results from 100 participants randomly selected from original sample and interviewed.	Higher # of lifetime sex partners, frequency of intercourse, and masturbation reported during interview than on SAQ. No further statistical information provided.	Article designed to describe the external validity of the results from a populations study.
Taylor et al. (1994)	<u>N</u> = 269. Gender: female. <u>Age</u> : <u>M</u> = 42. <u>Race</u> : 96% Caucasian. 2% Af. Am., 2% Hispanic. <u>Source</u> : gynecological clinic, human sexuality program, nursing staff.	Frequency of sexual fantasy, kissing, masturbation alone, mutual masturbation, petting and foreplay, oral sex, vaginal penetration or intercourse, anal sex / 6-point scale (“Not at all” to “More than once a day”)	SAQ / NR	1 month	20 / NR	1-month test retest intertrial interval: $r = .78$ for sexual activity factor (consisting of 9 items)	Compared SAQ with the Derogatis Sexual Functioning Inventory (Derogatis & Melisaratos, 1979)	Sexual behavior items correlated significantly with information, experience, drive, attitude, affect, fantasy, body image, satisfaction, and general satisfaction DSFI subscales.	Measure includes questions related to sexual dysfunction.
Tripp et al. (1994)	<u>N</u> = 1,025. Gender: 50% female. <u>Age</u> : 14 - 16. Race: NR <u>Source</u> : 9 public and private schools in Wales.	Sexual behavior data derived from social norms for sexual activity and the individual's history of relationships.	FTFI and SAQ at time 1, postal SAQ one year later / NR	2 years	NR / NR	Post-questionnaire follow-up with one year inter-trial interval. This approach measures behavioral consistency rather than reliability.	Concordance between FTFI and SAQ at time 1. Concordance between SAQ's.	Sexual activity varied by 3% across the two modes.	Assessment was acceptable to schools, parents, and students. Designed to provide population estimates rather than detailed assessment for each participant.
Upchurch et al. (1991)	<u>N</u> = 71. Heterosexual couples. <u>Age</u> : <u>M</u> = 26. Race: 87% Af. Am. <u>Source</u> : STD clinic.	# of days and # of acts of any sexual behavior, vaginal sex, and vaginal sex with condom / Open format	FTFI calendar-aided assessment / NA	4 weeks	NR / NR	NR	Comparison with partner self-report over the entire 4-week reporting period by each individual week in reporting period	Paired <i>t</i> -test analyses revealed that women reported more acts of vaginal sex in past 4 weeks than men	Overall, partner reports were very consistent. Multivariate analyses showed that socioeconomic status, marital status, and age had no effects on interpartner consistency
Weinhardt et al. (in press)	<u>N</u> = 110 in original sample; 58 retested. Gender: 54% male. <u>Age</u> : <u>M</u> = 19.7. Race: 90% Caucasian. <u>Source</u> : College subject pool.	# of male and female oral, vaginal, & anal sex partners, frequency of protected and unprotected oral, vaginal and anal sex, sex after drinking alcohol, and sex with non-monogamous partner / Open format	2 FTFI versions: (1) single-item frequency questions and (2) timeline follow-back (TLFB) method / NR	3 months	29 single-item frequency questions, equivalent variables coded from TLFB interview / NA.	1-week TR inter-trial interval: single-item frequency questions r s ranged from .81 to .96; median $r = .89$. TLFB ranged from .86 to .97; median $r = .96$.	NR	NR	No gender differences in reliability. Bootstrapping, a nonparametric statistical technique, was used for reliability analyses because of nonnormal sexual behavior distributions.

Study	Participants	Sexual Behaviors Included in Analysis / Response Format	Mode(s) / Reading Level	Reporting period	Number of Items / Duration	Evidence for Reliability	Validation Method	Evidence for Validity	Comments
Wyatt et al. (1992)	N = 70. Gender: 100% female. Age: 18 - 40. Race: 57% Caucasian, 43% Af. Am. Source: College campus and community.	# of consensual and nonconsensual sexual behaviors	FTFI, SAQ, Telephone interview, randomized response technique (RRT) / NR	NR	398 / NR	Internal consistency: $r = .88$ ($p < .001$). Randomized response technique used for marital status, history of abortion, and history of STD. Not used for sexual behavior questions.	Comparison between FTFI and Bender Sexual Behavior Inventory (BSBI), which was completed after FTFI.	Concordance with BSBI: $r = 1.00$	Questionnaire designed to elicit non-consensual sexual abuse. Authors conclude that in their multi-ethnic sample, the FTFI maximized accuracy and specificity of reports. SAQ and telephone interview recommended as validity checks. RRT produced poor results for sensitive topics.
Zenilman et al. (1995)	N = 598. Gender: 54% male. Age: NR Race: 94.3% Af. Am. Source: STD clinic.	Condom use, # of partners, types of partners, frequency of intercourse, any oral or anal sex / Open format, categorized for analysis	FTFI calendar-aided assessment / NR	30 days	NR / NR	NR	Relationship of sexual behaviors to new STD infections	Self-reported condom use was related to new infections of Chlamydia among men. Associations between self-reported condom use and other STD infections were not significant for male or female participants	Results indicate the need to also inquire about whether condoms were used properly

Note. NR = not reported; NA = not applicable; FTFI = face-to-face interview; SAQ = self-administered questionnaire; CI = confidence interval; STD = sexually transmitted disease; TR = test-retest.