



Published in final edited form as:

Int J STD AIDS. 2001 June ; 12(6): 365–375.

Assessing Sexual Risk Behavior with the Timeline Followback (TLFB) Approach: Continued Development and Psychometric Evaluation with Psychiatric Outpatients

M P Carey, PhD, K B Carey, PhD, S A Maisto, PhD, C M Gordon, PhD, and L S Weinhardt, PhD
Center for Health and Behavior, Syracuse University

Summary

This paper describes a series of four studies, designed to provide evidence of the feasibility, reliability, and validity of the Timeline Followback (TLFB) method when used to assess sexual risk behavior with psychiatric outpatients. This population was selected because patients often have difficulty completing assessments of sexual risk behaviors due to deficits in attention, memory, and communication skills. All four studies demonstrated the feasibility of the HIV-risk TLFB. Study 1 also demonstrated that it can be completed in 20 minutes, and scored in less than 10 minutes. Qualitative data revealed that both patients and assessors found the features of the TLFB helpful. Study 2 provided evidence that the HIV-risk TLFB can be reliably scored by interviewers whereas Study 3 demonstrated that this measure can be completed reliably by patients and that TLFB of sexual behavior were consistent over time. Study 4 provided initial evidence for the validity of the HIV-risk TLFB but also suggested that the TLFB may yield frequency estimates that are slightly less than those obtained with single-item measures. We conclude that the TLFB is feasible, reliable, and valid, even in a population known to have difficulty with self-report measures.

Keywords

sexual behavior; assessment; reliability; validity; HIV

Introduction

The reliable and valid assessment of sexual behavior is a major challenge to behavioral research on HIV and other STDs¹. Various forms of self-report remain the most practical and ethical method to assess sexual behavior, but there are concerns about the accuracy of such self-reports. Self-report may be inaccurate due to memory difficulties, including simple forgetting, telescoping (distorting the recency of salient events), and the use of estimation heuristics rather than exact episodic memory to report behavioral frequencies². Measurement strategies need to be developed and refined in order to minimize the influence of memory problems in the recall of sexual behaviors.

One promising assessment strategy is the Timeline Followback technique (TLFB)³. The TLFB, which was developed originally to assess alcohol use, has several advantages relative to traditional survey and interview methods. First, the TLFB was designed to benefit from research in cognitive psychology that has established the value of “landmark events”, calendars, and other memory aids to facilitate recall⁴. The use of memory aids is especially

useful when working with individuals who have difficulties with motivation, concentration, or communication. The structure of the TLFB as well as its interactive format encourages an iterative process whereby memory of one event may facilitate recall of similar or related events. Second, the TLFB method permits interviewers to obtain enriched contextual information regarding risk behavior. This ability to provide detailed event-level data is especially important for research on the co-occurrence of risky behaviors. For example, researchers or clinicians can investigate whether risk behavior (e.g., binge drinking) is more likely to occur in certain situations (e.g., public taverns), with specific partners (e.g., new acquaintances), or following certain affective states (e.g., depressed mood). Third, TLFB procedures yield data that document behavior patterns (e.g., quantity, frequency) in greater detail and over varying intervals. Unlike other measures, the TLFB can provide information regarding the range of risk behaviors such as alcohol use. Compared to diary methods (which share some of the advantages just noted), the TLFB (a) is not reactive (i.e., it does not influence the behavior being assessed), and (b) it is less burdensome to participants who may be unable to adhere to the demands of daily self-monitoring. Thus, the TLFB is well-suited to clinical trials where investigators seek to measure the efficacy of an intervention. Overall, the TLFB method appears to have great potential utility for a variety of research and clinical purposes.

Careful evaluation of the reliability and validity of TLFB reports of alcohol consumption has occurred with a variety of populations. Reliability evidence indicates that TLFB estimates of drinking behavior are consistent over time^{5,6}. Validity evidence comes from several sources. For example, TLFB data from participants in alcohol treatment correspond well with official records of hospitalized and incarcerated days⁷. Comparisons of TLFB alcohol consumption with reports of the same events from collateral informants yields moderate to high correlations⁸. Agreement between recent drinking estimates on the TLFB and commonly used averaging methods is good^{9,10}. In summary, the TLFB technique is a psychometrically sound, retrospective method for assessing alcohol use patterns and related events.

Recently, we modified the TLFB approach to assess sexual behavior with college students¹¹. Participants ($N = 58$) completed a 90-day TLFB interview on two occasions, separated by one week. Test-retest intraclass correlations from the TLFB showed that all sexual behaviors were reported reliably (range = .86 to .97). Reliability coefficients were equivalent across each of the three months assessed with the TLFB, and were equivalent to those obtained with conventional assessment methods (i.e., single-item questions). Frequency data obtained from the TLFB also corresponded well to data obtained with single-item assessment methods. This initial study showed that the sexual behavior TLFB interview provides reliable reports of sexual behavior when used with high functioning and verbal young adults. However, if the TLFB is to be useful in STD prevention contexts, its feasibility must be demonstrated in other populations.

Midanik and colleagues¹² used a similar 30-day TLFB to assess alcohol use, drug use, and sexual behavior in a sample of 418 gay or bisexual men in treatment for substance abuse. When compared to standard summary methods, the TLFB yielded lower reports of sexual behaviors. However, in this study, the assessment measures were confounded with the mode of administration; that is, the TLFB was administered in a face-to-face interview (FTFI) whereas the single items were obtained with a self-administered questionnaire (SAQ). Because prior research has found that FTFI administration leads to lower frequency estimates than SAQs, additional research is needed to clarify the whether the TLFB yields lower estimates of HIV risk behavior when the mode of assessment is held constant.

This paper describes the HIV-risk TLFB, an interview that we use to measure sexual behavior as well as alcohol and other drug use. The HIV-risk TLFB was designed to provide a comprehensive assessment of HIV and STD risk for both men and women, and to yield

summary scores for frequencies of protected and unprotected vaginal, oral, and anal sex. This study extends our earlier efforts¹¹, and those of Midanik and colleagues¹², in several ways. First, we evaluated the reliability of the coding as well as the reports of sexual risk behaviors. Second, we included both 30- and 90-day assessment intervals. Third, both the TLFB and the single item assessments were administered with a face-to-face interview. And, fourth, we sampled primarily heterosexual men and women from a clinical population known to be at high risk for HIV and other STDs; that is, participants in the current program of research were all psychiatric outpatients with severe and persistent mental illnesses. This population, often characterized by deficits in attention, memory, and communication skills, provides a stringent test for an event-level assessment of sexual and substance use behaviors. Severely mentally ill adults also experience increased prevalence of HIV infection¹³; thus, psychometrically sound and clinically sensitive sexual behavior assessments are particularly needed for this population.

In this report, we describe a series of four studies that were designed to provide evidence of the feasibility, reliability and validity of the sexual behavior component of the HIV-risk TLFB. Feasibility would be demonstrated if participants were able to complete the HIV-risk TLFB in a timely fashion, without distress or confusion. Reliability would be demonstrated if raters provided equivalent summaries regarding behavioral frequencies, and if reports regarding the same interval but obtained on separate occasions were consistent. Validity would be suggested by moderate to strong correlations between estimates obtained with the TLFB and traditional (i.e., non-calendar-based) interview methods. We first present the methodological features that were common to all four studies. Next, we present each of the four studies separately, describing each study's unique aims, participants, procedures, analyses, and results. Finally, we summarize the evidence from all four studies and discuss the implications of this program of research.

Methods Common to all Four Studies

Source of participants

All participants were receiving outpatient care from psychiatric facilities in a medium-sized city in the northeastern United States. In addition, all participants were enrolled in Phase I of the "Health Improvement Project" (HIP), funded by the National Institute of Mental Health. Phase I of the HIP was designed to identify the prevalence and correlates of HIV-related risk behavior among the severely mentally ill. Phase II was designed to evaluate the efficacy of two risk reduction programs: an HIV-risk reduction program (i.e., to promote safer sexual behavior in order to avoid infection with HIV or other STDs), and a substance use reduction program (i.e., to promote reductions in the use of alcohol, tobacco, caffeine, and other non-prescribed drugs). All procedures for the HIP, including those described in this report, were approved by Institutional Review Boards at the two participating hospitals and at the authors' academic institution.

Interviewers

The interviewers were eight (7 female, 1 male) BA-level research assistants (RAs). Prior to assessing patients, all RAs were trained in the HIV-risk TLFB by the investigators, who are senior scientist-practitioners who had used this measure extensively in clinical work and research. The training involved the following steps: review a detailed manual, answer sheet, and coding sheet; listen to audiotaped, illustrative administration; meet with an experienced assessor to review the procedure for giving the TLFB as well as the forms used and scoring; observe an experienced assessor giving the TLFB; practice giving the measure to a research team member; review completed TLFBs to see proper coding on calendars; administer the measure to clinic volunteer who was not a research participant while an experienced assessor

observed and provided feedback; and administer TLFB to research participant while the Project Director observed and provided feedback.

Patient recruitment

Patients were invited to participate in the HIP if they reported (a) alcohol or illicit drug use and (b) sexual activity in the previous year, and (c) if they were between the ages of 18 and 65¹⁴. They were told that their initial experience would involve interviews and self-report measures designed to obtain diagnostic information, sexual behavior, substance use, and other health topics. They were also told they would receive modest compensation for their time and to offset travel and other expenses associated with their participation. Patients who agreed to participate provided informed written consent and were scheduled for the first of three sessions.

Session 1

During the first session, patients participated in an abbreviated version of the Structured Clinical Interview for the Diagnostic and Statistical Manual of Mental Disorders (SCID)-IV¹⁵. We used the psychotic, mood, and substance-use disorder modules of the SCID-Patient Version, which is the preferred form for psychiatric populations in which differential diagnosis of psychotic disorders is required. All interviews were administered by clinical psychologists, and were videotaped to allow determination of inter-rater reliability. The diagnostician also administered the Mini-Mental Status Exam (MMSE)¹⁶, which was used as a brief screen for cognitive dysfunction. The MMSE assesses orientation, memory, attention, naming, verbal comprehension, writing and copying abilities. Ample evidence of test-retest stability and validity is available¹⁶. Participants scoring 23 or lower (the standard cut-off score for determining signs of dementia) were excluded from the study.

Session 2

At the beginning of the second session, breathalyzer screens (Alcosensor IV, Intoximeters, Inc.) were administered to all participants to ensure sobriety at the time of the assessment. Next, the participants completed the HIV-risk TLFB with the assistance of the interviewer. The HIV-risk TLFB was adapted from the original TLFB³, to obtain sexual and substance-using behaviors over a 3-month interval. A structured manual was developed to guide the interview and subsequent scoring (available upon request from the authors).

The interviewer recorded the start time and then prepared participants explaining that use of a calendar and a set of memory aids would help them to recall sexual events. The interviewer then presented the calendar on which the assessment interval was marked, as well as civic and religious holidays. Participants identified special days (e.g., check receipt days, birthday) or salient periods (e.g., hospitalizations, incarcerations), which were marked on the calendar by the interviewer. Participants were encouraged to use personal date books, if available, to assist them. Next, the interviewer reassured participants that all information was confidential and encouraged them to complete the calendar as accurately as possible. The TLFB was completed in three separate “passes,” one each for sexual behavior, alcohol use, and drug use (this order was the same for all participants). Each type of behavior was recorded on the same calendar.

Assessment of sexual behavior had the following steps. First, the interviewer defined the sexual terms in language that was familiar to the participant, consistent with established guidelines¹⁷. Second, the participant was asked to provide the initials of all partners during the past three months. For each partner, the interviewer requested information regarding partner characteristics (e.g., new, casual, regular) and “risk” status (e.g., did their partner have sex with men [MSM]? had this partner injected drugs [IDU]? was the partner infected with HIV [HIV +]?) as well as the participant's perception as to whether the relationship was mutually monogamous or not. The interviewer recorded all information on a coding sheet. Third, all

penetrative sexual opportunities for each partner were recorded on the calendar, before moving on to the next partner. A discrete coding scheme allowed interviewers to summarize all of this information directly on the daily blocks on the calendar. Consistent with memory research, participants were encouraged to begin with the most recent event and then to work backward for that partner. Each sexual event was reviewed to determine type of sex (oral, anal, vaginal sex), type of protection (if any), time of day, whether alcohol or other drugs were involved, whether there was discussion of safer sex or HIV prior to sex, and whether sex trading or coercion was involved.

After the sexual behavior assessment, the interviewer began the substance use assessment, beginning with alcohol use and then proceeding to street drugs. For each substance class, the interviewer provided the necessary definitions and used language that was familiar to the participant. For example, for alcohol, pictures of a standard drink for each of the classes of alcohol beverages (i.e., beer, wine, and distilled spirits) was presented, according to the established TLFB instructions³. Extended periods of bingeing or abstinence were recorded on the calendar. The interviewer then systematically reviewed each substance use event and recorded information regarding time of day, minimum and maximum amounts consumed (for alcohol), and whether the participant had sexual relations before, during, or after the substance use.

Study 1

The purpose of Study 1 was to collect both quantitative and qualitative evidence regarding the feasibility of the HIV-risk TLFB. Unique procedures included measuring time to completion and time to code. We also obtained representative comments from both interviewers and participants regarding their subjective experience completing the TLFB assessment. Based on previous work with the substance abuse TLFB in this population⁵, we expected that the sexual TLFB would be feasible, and that it would be perceived as a useful and manageable assessment tool from both interviewer and participant perspectives.

Participants

The patient participants were 73 female and 35 male outpatients (M age = 36.9 years; see Table 1). They were diagnosed with schizophrenia (14%), schizoaffective disorder (14%), bipolar disorder (15%), and major depression (56%). The patients were primarily European-Americans (76%), unmarried (89%), with a high school education or less (58%). Over 88% of the sample reported sexual activity in the previous 3 months; however, only 44% reported that they had a steady sexual partner. Thirty-eight percent reported a STD in their lifetime.

Procedures Unique to Study 1

Three procedures were unique to Study 1. First, the interviewer recorded the time needed to complete the TLFB and to code the TLFB.

Second, a subset of 45 participants (23 women, 22 men) patients were invited to return for a third session. During this individual session, they were invited to provide their impressions of the assessment experience during individual “exit interviews” six months after their initial experience with the TLFB. Interviewers followed a semi-structured outline of open-ended questions (e.g., “Why did you decide to participate in this project?” and “What did you like/not like about the project?”). No specific prompting was provided with regard to the TLFB component; thus, the participants were free to respond about the whole experience of being in the assessment study. Within the broad outlines of the open-ended questions, participant responses also guided the flow of topics. Interviews lasted approximately one hour, and were

audiotaped for later transcription. Transcripts were content analyzed by two independent raters to determine participant perceptions of the TLFB.

Third, the eight interviewers who participated in the project were asked to provide written answers to open-ended questions about their experiences with TLFB administration. Questions included the following: Have you ever had someone not be able to complete the TLFB? How many? Did they discontinue because they were too frustrated? Too distressed? For any reason? More generally, describe your impressions of how participants react to doing the TLFB. Are there any facets of the TLFB that are particularly helpful in helping participants recall behavior (s)? Do they like or dislike it? The surveys were collected after interviewers had completed 30 or more TLFBs.

Results

Quantitative evidence—All participants (100%) completed the TLFB. The average time to administer the TLFB assessment was 19.7 minutes ($SD = 14.3$, range = 4 – 99 mins), and the average scoring time was 7.9 mins ($SD = 5.7$, range = 2 – 48 mins); 94% were scored in less than 15 minutes.

Qualitative evidence—In this section, we report only themes that were indicated by more than one interviewer or participant. The interviewer impressions and participant experiences indicated that participants generally found the TLFB to be acceptable. Occasional breaks (especially for sexually active participants who had a lot of behavior to record), and encouragement (for those who were not confident about their abilities to recall) were helpful. Samples of comments from the TLFB interviewers follow.

“I’ve never had anyone not complete the TLFB. However, I have had some participants need a break, for a snack or soda. Sometimes this is due to frustration at what they perceive to be ‘poor’ recall. Encouragement was important.”

“The calendars help a great deal - it gives both the participant and the interviewer a tangible representation of the past three months. Noting the “special days” on the calendar seems to help. They often say “I remember I did this on this day because it was right before such-and-such.” Determining holidays, payday, vacations, time in jail, is important.”

“I think that for the most part, the TLFB went amazingly well. I was always surprised when people I hardly knew were willing to give out such personal information. The calendars were an essential and necessary part of describing the process of recalling events to the participants. Many of them wouldn’t have understood what I was talking about without using the calendars. I found that if I really encouraged them to give me some anchor days of their own -- anything-- that this helped recall behaviors. I also found that the “lower functioning” individuals were usually the ones who didn’t have a whole lot of behaviors, if any, so the Timeline was actually fairly easy for them. It was the higher functioning participants who tended to have a lot more behaviors and used the calendars to help them in their recollection of events.”

“I found it very helpful to review the calendars and mark off special dates with the participant. It gives the administrator an idea of how the individual functions, who the participant spends time with, the context of behaviors, etc. Having the special dates is helpful. For women, asking them to recall their menstrual cycles is helpful. I also found it to be helpful to work forward on the TL when the participant had a new partner during the 3-month epoch. Working from the first sexual encounter to the present, I think it’s easier for the participants to remember when they first discussed condoms/HIV.”

From participants, we learned that features of the TLFB protocol (e.g., visual aids) and skills of the interviewer facilitated their completion. For example, the following exchange occurred:

Q: Was there anything about those assessment sessions that was difficult for you?

A: Well, with the calendars, sometimes it was hard to try and think back, that amount of time unless something really stuck in your head, but, laying out the calendars and having me write down personal holidays plus the holidays that were going on made it as easy as possible for me to remember.

Moreover, many participants reported that their TLFB experience was useful for self-monitoring and thinking about their sexual choices. The spontaneous responses to the exit interview questions, without prompting about the TLFB component of the assessments, illustrate how the TLFB assessment was useful, and how many liked the experience. Each of the following quotes is from a different participant:

Q: Were there things that you liked about the project?

A: I liked how they asked questions and you had to kind of use your memory to answer the questions. A lot of times a lot of things that she asked, I forgot, so I had to really sit there and think about when things had happened.

Q: Why did you decide to participate?

A: I thought it was a good way to make money at first, and then I liked the way that you tackled the use of alcohol and drugs and crack, and everything like that. I figured even that, just getting down on paper the dates was better than just leaving it in a mess. That was one of the reasons that I wanted to do it.

Q: Were there things about the project that you enjoyed?

A: It was very interesting once, when we were doing the three calendars. That was really interesting, things like that. Because I did answer the questions truthfully, as best I could.

Q: Were there any other things that you liked about the project?

A: I liked going through the calendars about the drug use because it showed me a kind of pattern.

Q: Did you learn anything about your own health from being in this project?

A: Partly the fact that I go and get a little crazy now and then. About once every other month. I'll start participating in strange sex, and drug deals.

Q: You learned that through being in the project?

A: Yeah, because I can't really think out one day from the next, unless I got those charts in front of me or I'm looking through them.

Study 2

The purpose of Study 2 was to assess the reliability of the coding scheme developed for the HIV-risk TLFB. We designed a descriptive notation scheme for interviewers to use during primary data collection; notations were placed on the calendar that contained information about type of sexual activity, presence of condom, partner identification, and time of day (notation system available from the authors). Then interviewers derived a series of summary scores for the purposes of data analyses. The aim of Study 2 was to demonstrate that summary scores can be obtained reliably across raters. We predicted good interrater reliability across summary scores.

Participants

Participants ($n = 25$) for Study 2 were selected randomly from the Study 1 cohort, from among those who reported being sexual active and using alcohol or other drugs at least once during the previous three months. The majority of participants were diagnosed with schizophrenia or schizoaffective disorder (56%). Most identified themselves as European-American (64%) or African-American (28%), had a high school education or less (76%), and were not married (96%). All were sexually active in the last three months and 44% had an STD in their lifetime (see Table 1).

Procedures Unique to Study 2

At the conclusion of the TLFB interview (as described in common methods section), the interviewer coded the sexual and substance use behavior, and transferred these data from the TLFB recording form to the data summary form. To allow for evaluation of intercoder reliability, a second coder used only the calendars created during the interviews to code the data independently. We computed the intraclass correlation coefficient (ICC) between the original data and the second coding for each sexual behavior.

Results

We examined the intercoder reliability of frequency of vaginal, oral, and anal sex, with and without a latex barrier, and several other detailed event-level behaviors (see Table 2). Across all of the behaviors that occurred during the 3-month period for the intercoder sample, the mean ICC was .98 and median ICC was also .98 (range = .80 to 1.00). These data provide evidence that the coding scheme used in this study could be interpreted consistently by different coders, and that the summary scores produced across a wide variety of sexual behaviors were reliable.

Study 3

The purpose of Study 3 was to evaluate the temporal stability of self-reported sexual behaviors obtained from the TLFB. We evaluated the test-retest reliability of summary scores from the full three-month assessment interval as well as those from the most recent one-month interval, both derived from the same three-month TLFB calendar. We predicted that sexual behaviors would be reported consistently over a one-week test-retest interval.

Participants

The test-retest sample consisted of 66 psychiatric outpatients (50% men); their mean age was 34 years (range = 18 - 60 yrs). As detailed in Table 1, diagnoses were 62% major mood disorder (Bipolar Disorder or Major Depression) and 38% Schizophrenia/Schizoaffective Disorder. Most participants (63%) had some high school education ($M = 12$ yrs). Sixty-four percent reported sexual activity in the previous 3 months but only 24% reported that they lived with their primary sexual partner. In addition, 80% of the sexually active participants reported unprotected vaginal intercourse, and 42% reported a STD in their lifetime.

Procedures Unique to Study 3

Potential participants were recruited in the way described in the common methods section, except that they were informed that they would be given some measures more than one time. The first TLFB was administered as described previously; the second TLFB was administered during an additional assessment session. Typically, the retest TLFB was scheduled within 1 week (M test-retest interval = 5 days; $SD = 4.2$; range = 1 - 19). Prior to both sessions, breathalyzer screens were administered to all participants to ensure sobriety. The instructions for the second TLFB included a reminder of why participants were being asked the questions

again, and instructions to minimize participants' attempts to use the first TLFB administration for memory cues, as follows:

“I'm going to ask about the same behaviors that I asked about the other day. We're asking people things twice, to help us find the best way of asking these questions. It is important that you do your best to remember what behaviors you did during the past 3 months. It's not important that you tell me the same things you told me the other day. As I asked you to do the last time, I want you to just try to remember what you did.”

Both the initial and the retest TLFB were completed in three separate passes, one each for sexual behavior, alcohol use, and drug use. Participants received modest compensation for their time and to offset travel and other expenses associated with their participation.

Initial examination of scatter plots revealed two characteristics common to sexual behavior data, namely, non-normal distributions and outliers¹⁹. Therefore, to examine the effects of removing these outlier participants, we report findings both with and without outliers. To examine stability between the two TLFB reports, we computed ICCs between the initial and subsequent assessment for each sexual behavior.

Results

Table 3 shows the test-retest ICCs for sexual partners, vaginal sexual events, vaginal sexual events with condom, oral sex receiving, and oral sex giving. (The frequencies of insertive or receptive anal sex were so low in this sample that stability coefficients could not be calculated.) Reliability coefficients for each behavior were calculated both for the most recent month, and for the last 3 months. Generally, reports over both intervals were stable, although the ICCs were occasionally affected by the presence of outliers (values that were 5 or more standard deviations away from the mean Time 1 – Time 2 discrepancy score). Visual inspection of selected test-retest figures illustrate this pattern. Figure 1 plots the initial (Time 1) TLFB data for vaginal sex occasions (3-months) along the y-axis, and the retest (Time 2) data along the x-axis. If the single most extreme outlier (see point in lower right of Figure 1) is removed, then the ICC improves from .73 to .87. Similarly, the test - retest ICC for vaginal sex events with condoms (one-month) with one outlier included is .52; without that outlier, the ICC is .97. In Table 3, the absence of a parenthetical value indicates that there was no obvious outlier for that variable. ICCs calculated separately by gender revealed no obvious patterns of differential reliability.

Study 4

The purpose of Study 4 was to compare summary scores obtained from the TLFB with responses to commonly-used survey questions (i.e., “single-item” questions). The latter elicit an average frequency estimate for sexual behaviors over a given time frame; in contrast, the TLFB elicits recall of individual sexual events, and frequency estimates are derived by summing these event-level data. Based on our previous work with the TLFB¹¹, we expected to find moderate to strong correlations between the two methods. To follow-up the findings reported by Midanik et al.¹², we analyzed whether self-reports from the TLFB were lower than estimates yielded by the single-item interview in a set of exploratory analyses.

Participants

Participants were 230 outpatients (43% men); mean age was 37 years (SD = 10.0; range = 18 - 60 yrs; see Table 1). SCID diagnoses were 73% major mood disorder (Bipolar Disorder or Major Depression) and 27% Schizophrenia/Schizoaffective Disorder. The majority were European-American (72%), most were unemployed (81%), and had a high school education or less (62%). Regarding sexual risk behavior, 79% reported sexual activity in the previous 3

months, and 44% were married or had a current sexual partner. In addition, 80% of the sexually active participants reported unprotected vaginal intercourse, and 38% reported a STD in their lifetime.

Procedures Unique to Study 4

Recruitment and data collection procedures were nearly identical to those described in the common methods section except that, during the second session, participants first completed the Sexual History Form (SHF), a traditional sexual history interview that asks participants about the frequency of unprotected and protected oral (giving and receiving), anal (insertive and receptive), and vaginal intercourse, and the number of male and female sexual partners using separately for the last 30 and the last 90 days. The SHF is characterized by an open response format to reduce unreliability due to memory distortion²⁰. All items have been used in prior research with this population^{21,22}, and are similar to those used routinely in sexual behavior research²³⁻²⁵. After the SHF was administered, participants then completed the TLFB interview as described in Study 1.

Initial examination of scatter plots again revealed non-normal distributions and outliers (as in Study 3, outliers were defined as being more than 5 SD away from the mean discrepancy score for each variable). As with the test-retest results, we report ICCs between the TLFB and single-items both with and, for variables where an obvious outlier existed, without outliers to demonstrate the effects of outliers on these ICCs. To explore whether the TLFB yields systematically lower frequency reports compared to the single item approach, we proceeded in two steps. First, we computed a discrepancy score (i.e., SHF minus TLFB) for each participant for each behavior. Second, we examined the discrepancy scores with the non-parametric Wilcoxon signed-rank test to determine if, relative to what would be expected by chance, participants were more likely to report higher values on the SHF compared to the TLFB.

Results

Table 4 presents the ICCs -- reflecting level of agreement -- between the TLFB and single-item methods for 5 separate variables, each evaluated over 2 time intervals. The level of agreement was good to excellent by established standards²⁶, when the single most extreme outlier was removed. ICCs calculated separately by gender produced a similar pattern of relationships.

Table 5 presents (a) descriptive data (raw means, standard deviations, medians, and ranges) for both the TLFB and single-item measures, (b) difference scores between the TLFB and single-item assessments, and (c) p-levels for the Wilcoxon sign rank tests comparing the difference scores. Results from the Wilcoxon sign rank tests indicated that more participants reported higher frequencies on the SHF compared to the TLFB for three of the four vaginal sex items, and on all four of the oral sex items. The ordering of means for 6 of the 7 non-significant findings were in the same direction (i.e., TLFB < Single Item). Overall, however, the magnitude of these differences was small; the range of the mean discrepancy scores was 0.03 to 0.7 sexual events during a 1 month period, and 0.1 to 1.8 sexual events during a 3 month period (see Table 5, 7th column).

Discussion

This series of four studies provides substantial evidence that the HIV-risk TLFB can be used to simultaneously gather sexual behavior and substance use data, even with psychiatrically impaired participants who have difficulty with tasks involving recall and reporting. Across these four studies, the TLFB was administered on more than 400 occasions without a single

refusal or failure to complete the interview. Moreover, the results indicate that the HIV-risk TLFB can be completed in 20 minutes, and scored in less than 10 minutes. Qualitative data provided by both interviewers and participants revealed that the structural features of the TLFB (e.g., use of calendars and landmark events) facilitated the task of recalling sexual behaviors that occurred up to 90 days earlier. This demonstration of feasibility extends previous reports^{11,12} to another population that is known to be vulnerable to infection with HIV and other STDs.

Studies 2 and 3 of this series also provided evidence that the HIV-risk TLFB can be reliably scored by interviewers and completed by participants. Study 2 showed that trained interviewers agreed on interview results at a high level across a large number of sexual events (see Table 2). This is the first demonstration of inter-rater reliability with the HIV-risk TLFB, and builds on previous research conducted with the alcohol (only) TLFB³.

Study 3 showed that TLFB self-reports of sexual behavior were consistent over time. We obtained test-retest correlations that are at least equivalent to (and often surpass) correlations obtained using traditional, single-item measures (see Table 3). For example, Sohler et al.²⁷ used single-item questions in an interview with 39 mentally ill men in New York City. They also used a retest interval of approximately one week, and reported ICCs ranging from .54 to .87 for partner type, .74 to .82 for specific sexual behaviors, and .49 to .59 for condom use. The ICCs in the current study ranged from .82 to .94 for total sexual partners, .47 to .98 for vaginal sexual events, and from .47 to .82 for oral sex events. The latter two ranges improved to .86 - .98 and .65 - .94, respectively, with the removal of a single outlier. These levels of association meet or exceed conventional standards regarding test-retest stability²⁶, and indicate that the TLFB yields reliable data on socially sensitive behaviors, even in a population known for its cognitive deficits.

Finally, Study 4 provides some evidence for the validity of the HIV-risk TLFB by showing moderate to strong associations between the results obtained with the TLFB and results obtained by traditional, single-item measures. However, when frequency estimates obtained from the TLFB were compared to those yielded by single-item methods, a pattern emerged that suggested slightly lower values on the TLFB relative to the traditional measures. This finding is consistent with one earlier report¹², and warrants brief discussion.

The current data do not allow us to determine if the TLFB leads to under-reporting, if single item measures lead to over-reporting, or if a combination of processes accounts for these findings. Both the TLFB and the single-item approaches strive to elicit accurate information from episodic memory. Episodic memory is vulnerable to memory errors, such as simple forgetting and telescoping²⁸. Recall of episodic memory can be improved through use of multiple questions about an event (made possible in an interview format), use of landmark events (rather than dates) as retrieval cues, and reconstruction of past events using multiple modalities (e.g., verbal and visual).²⁸ These are strategies associated with the TLFB approach, and would seem to favor this approach. A disadvantage of this approach is the extra cost associated with its administration. In contrast, single item approaches would seem to encourage the use of estimation heuristics, which would be expected – theoretically – to lead to less accurate memories. Additional research is needed to continue to examine this question, and to obtain corroborating evidence of both single-item and TLFB estimates.

We wish to acknowledge the limitations of this program of research, and suggest directions for future research. Most importantly, the current research does not provide strong evidence of the validity of the TLFB. Although the primary goals of this investigation were to assess the feasibility and reliability of the HIV-risk TLFB, future research needs to obtain evidence of the validity of the TLFB, for example by using collateral partner interviews, concurrent self-

monitoring (e.g., with a diary), biomedical markers, or other strategies. Second, we evaluated test-retest stability at one week intervals for reports ranging back one to three months. These reporting intervals were selected because they are common end-points for randomized prevention trials; however, investigation of test-retest stability over a wider range of test-retest and reporting intervals will help to establish the temporal limits of this type of self-report. A good model for this type of research (using surveys) was reported by Kauth and colleagues²⁰. Third, we did not examine the association between specific cognitive deficits and self-report reliability or validity; future research might explore the influence these deficits have on self-report data. Finally, we did not obtain comparable feasibility and test-retest evidence for the traditional interview method.

In conclusion, our results combined with those obtained in similar investigations suggests that both the TLFB and the single item methods are feasible and reliable methods. There is no compelling, psychometric reason – at this time – to prefer one method over the other. In the absence of a strong scientific rationale, the choice of an assessment strategy is best made on the basis of how the resulting data are to be used. When working with clinical populations or those with cognitive difficulties, or when event-level analyses are anticipated (e.g., to evaluate the relationship between alcohol use and sexual risk behavior²⁹), then the TLFB approach would appear most useful. In contrast, when greater anonymity is desirable, or brief assessments designed to yield very specific and limited data are indicated, then traditional single-item interviews would appear to be more appropriate. We encourage continued research to optimize the quality of the self-report measures; the behavioral data yielded by these methods are essential if we are to determine with confidence the prevalence of risk behavior and the efficacy of risk reduction interventions.

Acknowledgements

This study was supported by a grant from the NIMH to the first author (R01-MH54929). The authors thank Kristin Barnes, Connie Basta, Jennifer Becker, Susan Bland, Brian Borsari, Christopher Correia, Lauren Durant, Don Fredericks, Julie Fuller, JulieAnn Hartley, Deborah Kahkejian, Jaejin Kim, Pat Lewis, William Licurse, Jeanette Mattson, Dan Neal, Teal Pedlow, David Peppel, MaryBeth Pray, Eileen Ryan, Kerstin Schroder, Peter Vanable, Emily Wright, and Denise Zona for their assistance with the Health Improvement Project.

References

1. Auerbach JD, Coates TJ. HIV prevention research: Accomplishments and challenges for the third decade of AIDS. *Am J Public Health* 2000;90(7):1029–1032. [PubMed: 10897177]
2. Weinhardt LS, Forsyth AD, Carey MP, Jaworski BC, Durant LE. Reliability and validity of self-report measures of HIV-related sexual behavior: Progress since 1990 and recommendations for research and practice. *Arch Sex Behav* 1998;27(2):155–180. [PubMed: 9562899]
3. Sobell, LC.; Sobell, MB. Timeline followback user's guide: A calendar method for assessing alcohol and drug use. Toronto, Ontario, Canada: Addiction Research Foundation; 1996.
4. Hammersley R. A digest of memory phenomena for addiction research. *Addiction* 1994;89:283–293. [PubMed: 8173494]
5. Carey KB. Reliability and validity of the time-line follow-back interview among psychiatric outpatients: A preliminary report. *Psychol Addict Behav* 1997;11:26–33.
6. Maisto SA, Sobell MB, Cooper AM, Sobell LC. Test-retest reliability of retrospective self-reports in three populations of alcohol abusers. *J Behav Assess* 1979;1:315–326.
7. Cooper AM, Sobell MB, Sobell LC, Maisto SA. Validity of alcoholics' self-reports: Duration data. *Int J Addictions* 1981;16:401–406.
8. Maisto SA, Sobell LC, Sobell MB. Comparison of alcoholics' self-reports of drinking behavior with reports of collateral informants. *J Consult Clin Psychol* 1979;47:106–122. [PubMed: 429642]

9. Cervantes EA, Miller WR, Tonigan JS. Comparison of timeline follow-back and averaging methods for quantifying alcohol consumption in treatment research. *Assessment* 1994;1:23–30. [PubMed: 9463496]
10. Maisto SA, Sobell LC, Cooper AM, Sobell MB. Comparison of two techniques to obtain retrospective reports of drinking behavior from alcohol abusers. *Addict Behav* 1982;7:33–38. [PubMed: 7080882]
11. Weinhardt LS, Carey MP, Maisto SA, Carey KB, Cohen MM, Wickramasingee SM. Reliability of the Timeline Followback sexual behavior interview. *Ann Behav Med* 1998;20:25–30. [PubMed: 9755348]
12. Midanik LT, Hines AM, Barrett DC, Paul JP, Crosby GM, Stall RD. Self-reports of alcohol use, drug use and sexual behavior: Expanding the timeline follow-back technique. *J Stud Alcohol* 1998;59:681–689. [PubMed: 9811089]
13. Carey MP, Weinhardt LS, Carey KB. Prevalence of infection with HIV among the seriously mentally ill: Review of research and implications for practice. *Prof Psychol: Res Prac* 1995;26:262–268.
14. Carey MP, Carey KB, Maisto SA, Gleason JR, Gordon CM, Brewer KK. HIV-Risk behavior among outpatients at a state psychiatric hospital: Prevalence and risk modeling. *Behav Ther* 1999;30:389–406.
15. First, MG.; Spitzer, RL.; Gibbon, M.; Williams, JBW. Structured clinical interview for the DSM-IV--patient version (SCID-I/P, Version 2.0). New York: New York State Psychiatric Institute; 1995.
16. Folstein MF, Folstein SE, McHugh PR. "Mini-Mental State" A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res* 1975;12:189–198. [PubMed: 1202204]
17. Carey, MP. Assessing and reducing risk of infection with the human immunodeficiency virus (HIV). In: Koocher, GP.; Norcross, JC.; Hill, SS., editors. *Psychologist's desk reference*. New York: Oxford University Press; 1998.
18. Fleiss, JL. *Statistical methods for rates and proportions*. 2nd. New York: Wiley; 1981.
19. Neter, J.; Wasserman, W.; Kutner, MH. *Applied linear statistical models*. 3rd. Boston, MA: Irwin; 1990.
20. Kauth MR, St Lawrence JS, Kelly JA. Reliability of retrospective assessments of sexual HIV risk behavior: A comparison of biweekly, three-month, and twelve-month self-reports. *AIDS Educ Prev* 1991;3:207–214. [PubMed: 1834142]
21. Weinhardt LS, Carey MP, Carey KB, Verdecias RN. Increasing assertiveness skills to reduce HIV risk among women living with a severe and persistent mental illness. *J Consult Clin Psychol* 1998;66:680–684. [PubMed: 9735586]
22. Carey MP, Carey KB, Weinhardt LS, Gordon CM. Behavioral risk for HIV infection among adults with a severe and persistent mental illness: Patterns and psychological antecedents. *Community Ment Health J* 1997;33:133–142. [PubMed: 9145255]
23. Carey MP, Maisto SA, Kalichman SC, Forsyth AD, Wright EM, Johnson BT. Enhancing motivation to reduce the risk of HIV infection for economically disadvantaged urban women. *J Consult Clin Psychol* 1997;65:531–541. [PubMed: 9256553]
24. Kalichman SC, Cherry C, Browne-Sperling F. Effectiveness of a video-based motivational skills-building HIV risk-reduction intervention for inner-city African American men. *J Consult Clin Psychol* 1999;67:959–966. [PubMed: 10596517]
25. Kelly JA, Murphy DA, Washington CD, et al. The effects of HIV/AIDS intervention groups for high-risk women in urban clinics. *Am J Public Health* 1994;84:1918–1922. [PubMed: 7998630]
26. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment* 1994;6:284–290.
27. Sohler N, Colson PW, Meyer-Bahlburg HFL, Susser E. Reliability of self-reports about sexual risk behavior for HIV among homeless men with severe mental illness. *Psychiatr Serv* 2000;51:814–816. [PubMed: 10828118]
28. Croyle, RT.; Loftus, EF. Recollection in the Kingdom of AIDS. In: Ostrow, DG.; Kessler, RC., editors. *Methodological issues in AIDS behavioral research*. New York: Plenum; p. 163-180.
29. Weinhardt LS, Carey MP, Carey KB, Maisto SA, Gordon CM. The relation of alcohol use to sexual HIV risk behavior among adults with a severe and persistent mental illness. *J Consult Clin Psychol*. in press

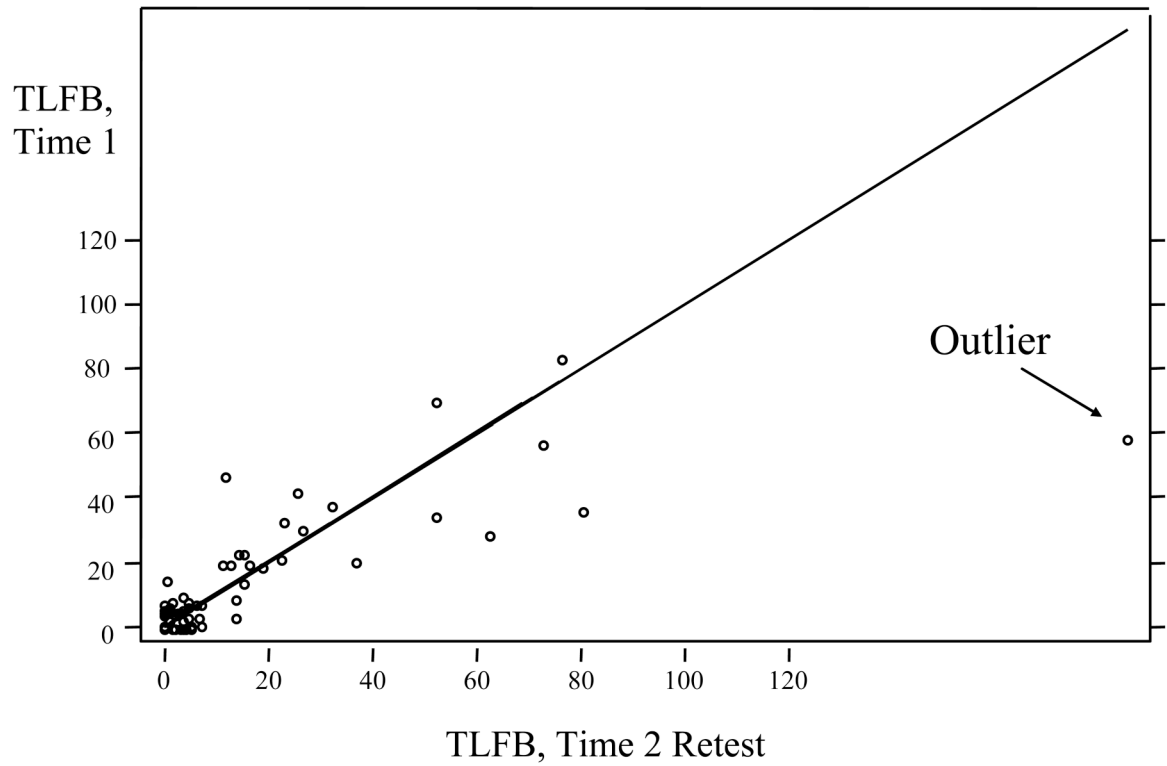


Figure 1. Test-retest for vaginal sex occasions, 3 months. This figure plots the Time 1 TLFB data for vaginal sex events (3 months) on the y-axis and the retest data on the x-axis. If the outlier is removed, the intraclass correlation coefficient improves from 0.73 to 0.87.

Table 1
Demographic, Psychiatric, and Behavioral Characteristics of Study Samples

| | Study 1: Feasibility n = 108 (%) | Study 2: Intercoder Reliability n = 25 (%) | Study 3: Test-Retest Reliability n = 66 (%) | Study 4: TLFB - Single Item Validity n = 230 (%) |
|----------------------------------|---|---|--|---|
| Gender | | | | |
| Men | 35 (68%) | 13 (52%) | 33 (50%) | 98 (43%) |
| Women | 73 (32%) | 12 (48%) | 33 (50%) | 132 (57%) |
| Ethnicity | | | | |
| European-American | 82 (76%) | 16 (64%) | 47 (71%) | 166 (72%) |
| African-American | 17 (16%) | 7 (28%) | 14 (21%) | 41 (18%) |
| Other | 9 (8%) | 2 (8%) | 5 (8%) | 23 (10%) |
| Age (M, SD) | 36.9 (9.7) | 34.8 (9.2) | 34.2 (8.9) | 36.6 (10.2) |
| Education (M, SD) | 12.5 (2.1) | 11.0 (2.9) | 12.1 (2.3) | 12.2 (2.6) |
| Income (M per year) | \$4926 | \$6871 | \$6228 | \$6228 |
| DSM-IV Diagnosis | | | | |
| Bipolar | 16 (15%) | 4 (16%) | 11 (17%) | 40 (17%) |
| Depression | 58 (54%) | 7 (28%) | 30 (45%) | 121 (53%) |
| Schizophrenia | 15 (14%) | 8 (32%) | 16 (24%) | 30 (13%) |
| Schizoaffective | 15 (14%) | 6 (24%) | 9 (14%) | 29 (13%) |
| Other | 4 (4%) | 0 (0%) | 0 (0%) | 10 (4%) |
| Relationship Status (current) | | | | |
| Married | 12 (11%) | 1 (4%) | 6 (9%) | 35 (15%) |
| Current partner | 36 (33%) | 10 (40%) | 18 (27%) | 67 (29%) |
| No current partner | 60 (56%) | 14 (56%) | 42 (64%) | 128 (56%) |
| Sexually Active (3 months) | | | | |
| No | 24 (22%) | 0 (0%) | 24 (36%) | 48 (21%) |
| Yes | 84 (88%) | 25 (100%) | 42 (64%) | 182 (79%) |
| Vaginal Sex Occasions (3 months) | | | | |
| Mean (SD) | 12.6 (22.3) | 16.9 (23.2) | 12.5 (19.5) | 13.9 (24.2) |
| STD (lifetime) | | | | |
| No | 67 (62%) | 14 (56%) | 38 (58%) | 142 (62%) |
| Yes | 41 (38%) | 11 (44%) | 28 (42%) | 88 (38%) |

Notes. TLFB = Timeline Followback; DSM-IV = Diagnostic and Statistical Manual of Mental Disorders, 4th edition; STD = sexually transmitted disease; M = mean; SD = standard deviation.

Table 2

Intercoder Reliability

| TLFB Item | ICC |
|---|------|
| Vaginal sex | .99 |
| Vaginal sex, with (latex) condom | .99 |
| Vaginal sex, (don't know) type of condom | .98 |
| Give oral sex, to male partner | .99 |
| Give oral sex, to male partner, with barrier | 1.00 |
| Receive oral sex, from male partner | .99 |
| Receive oral sex, from male partner, with barrier | 1.00 |
| Give oral sex, to female partner | 1.00 |
| Receive oral sex, from female partner | 1.00 |
| Insertive anal sex, male partner | 1.00 |
| Insertive anal sex, male partner, with condom | 1.00 |
| Anal sex, with (don't know) condom type | 1.00 |
| Receptive anal sex | 1.00 |
| Sexual event, with non-monogamous partner | .98 |
| Sexual event, with (unsure) partner monogamy | .99 |
| Sexual event, with HIV+ partner | .99 |
| Sexual event, with (don't know) partner's HIV serostatus | .96 |
| Sexual event, discuss HIV before sex | .86 |
| Sexual event, with IDU partner | 1.00 |
| Sexual event, after alcohol use | .91 |
| Sexual event, after any substance use | .99 |
| Sexual event, after substance use, discuss/ use condom | .95 |
| Sexual event, after substance use, discuss/ no use condom | .80 |
| Sexual event, after substance use, no discuss condoms | .99 |

Notes. ICC = Intraclass correlation coefficient; TLFB = Timeline Followback; IDU = injection drug user. ICCs were computed using frequency counts (e.g., vaginal sexual occasions) for 3-month interval.

Table 3
Test-Retest Reliability (ICCs) for Sexual Behaviors

| TLFB Item | All (<u>n</u> = 66) | Women (<u>n</u> = 33) | Men (<u>n</u> = 33) |
|------------------------------------|-------------------------|---------------------------|-------------------------|
| Total Sexual Partners | | | |
| 1 Month | .87 | .93 | .82 |
| 3 Months | .91 | .94 | .86 |
| Vaginal Sexual Events | | | |
| 1 Month | .78 (.90) | .71 (.94) | .89 |
| 3 Months | .73 (.87) | .66 (.91) | .86 |
| Vaginal Sexual Events, with Condom | | | |
| 1 Month | .52 (.97) | .98 | .47 (.95) |
| 3 Months | .95 | .92 | .98 |
| Oral Sex, Giving | | | |
| 1 Month | .69 (.76) | .67 (.76) | .77 |
| 3 Months | .52 (.91) | .47 (.94) | .80 |
| Oral Sex, Receiving | | | |
| 1 Month | .64 (.71) | .51 (.65) | .77 |
| 3 Months | .80 | .80 | .82 |

Notes. TLFB = Timeline Followback; ICC = Intraclass correlation coefficient; ICCs in parentheses were computed with a single outlier removed.

Table 4
Correlations between TLFB and Single-Items for Sexual Behavior

| Sexual Behaviors | All (<i>n</i> = 230) | Women (<i>n</i> = 132) | Men (<i>n</i> = 98) |
|------------------------------------|--------------------------|----------------------------|-------------------------|
| Total Sexual Partners | | | |
| 1 Month | .84 | .85 | .83 |
| 3 Months | .79 | .69 (.77) | .81 |
| Vaginal Sexual Events | | | |
| 1 Month | .69 (.76) | .76 | .64 (.76) |
| 3 Months | .82 (.86) | .75 (.82) | .88 |
| Vaginal Sexual Events, with Condom | | | |
| 1 Month | .49 (.83) | .85 | .21 (.76) |
| 3 Months | .80 | .88 | .70 |
| Oral Sex, Giving | | | |
| 1 Month | .58 (.74) | .67 (.77) | .53 (.79) |
| 3 Months | .54 (.67) | .67 (.75) | .54 (.68) |
| Oral Sex, Receiving | | | |
| 1 Month | .54 (.60) | .49 (.61) | .59 (.63) |
| 3 Months | .74 (.79) | .65 (.76) | .81 |

Notes. TLFB = Timeline Followback; ICC = Intraclass correlation coefficient; ICCs in parentheses are calculated with an outlier removed; ICCs for anal sex are not computed due to small number of cases (≤ 5).

Table 5
Comparison of TLFB and Single-Item Report for Sexual Behaviors

| Behavior | TLFB | | | Single Item | | | SI-TLFB | | p |
|--------------------------|-------------|-----|-------|-------------|-----|-------|------------|--|-------|
| | M (SD) | Mdn | Range | M (SD) | Mdn | Range | M (SD) | | |
| Sexual Partners | | | | | | | | | |
| 1 Month | 0.9 (1.5) | 1 | 0-18 | 1.0 (1.5) | 1 | 0-15 | 0.03 (0.9) | | .10 |
| 3 Months | 1.5 (2.4) | 1 | 0-50 | 1.4 (3.6) | 1 | 1-25 | -0.1 (2.0) | | .85 |
| Vaginal Sex | | | | | | | | | |
| 1 Month | 4.4 (8.1) | 2 | 0-57 | 4.9 (8.2) | 2 | 0-60 | 0.5 (6.4) | | .001 |
| 3 Months | 13.9 (24.2) | 5 | 0-175 | 14.4 (22.1) | 5 | 0-120 | 0.5 (14.0) | | .008 |
| Vaginal Sex, with condom | | | | | | | | | |
| 1 Month | 0.9 (3.2) | 0 | 0-34 | 1.1 (3.5) | 0 | 0-40 | 0.2 (3.4) | | .04 |
| 3 Months | 2.3 (2.4) | 0 | 0-40 | 2.4 (7.6) | 0 | 0-60 | 0.2 (4.5) | | .45 |
| Oral Sex, Giving | | | | | | | | | |
| 1 Month | 2.0 (4.4) | 0 | 0-43 | 2.8 (6.7) | 0 | 0-60 | 0.8 (5.2) | | .017 |
| 3 Months | 5.6 (13.0) | 0 | 0-57 | 7.4 (16.4) | 0 | 0-90 | 1.8 (13.2) | | .006 |
| Oral Sex, Receiving | | | | | | | | | |
| 1 Month | 1.4 (3.2) | 0 | 0-26 | 2.1 (4.3) | 0 | 0-30 | 0.7 (3.6) | | .0002 |
| 3 Months | 5.1 (14.3) | 0 | 0-165 | 6.3 (14.4) | 1 | 0-100 | 1.2 (10.3) | | .0005 |
| Anal Sex, Insertive | | | | | | | | | |
| 1 Month | 0.8 (0.6) | 0 | 0-7 | 0.2 (0.8) | 0 | 0-10 | 0.1 (0.8) | | .16 |
| 3 Months | 0.2 (1.7) | 0 | 0-12 | 0.3 (1.5) | 0 | 0-22 | 0.1 (1.2) | | .29 |
| Anal Sex, Receptive | | | | | | | | | |
| 1 Month | 0.08 (0.5) | 0 | 0-5 | 0.13 (0.8) | 0 | 0-10 | 0.1 (0.6) | | .20 |
| 3 Months | 0.3 (1.1) | 0 | 0-9 | 0.4 (2.0) | 0 | 0-25 | 0.1 (1.7) | | .32 |

Note. TLFB = Timeline Followback. Raw means and SDs are reported; p = significance level from Wilcoxon sign rank test.

All significant values indicate that that more people reported higher values on the Sexual History Form (SHF) relative to the TLFB than would be expected by chance.