

# RELIABILITY OF THE TIMELINE FOLLOW-BACK SEXUAL BEHAVIOR INTERVIEW<sup>1,2</sup>

Lance S. Weinhardt, M.S., Michael P. Carey, Ph.D., Stephen A. Maisto, Ph.D., Kate B. Carey, Ph.D.,  
Meredith M. Cohen, B.S., and Sohan M. Wickramasinghe, B.S.  
Syracuse University

## ABSTRACT

*The reliability of self-reported sexual behavior is a question of utmost importance to human immunodeficiency virus (HIV) prevention research. The Timeline Follow-Back (TLFB) interview, which was developed to assess alcohol consumption on the event level, incorporates recall-enhancing techniques that result in reliable information. In this study, the TLFB interview was adapted to assess HIV-related sexual behaviors and their antecedents, and its reliability was assessed. The interview was administered to 110 participants (46% women, *Age* = 19.7; *range* = 18–41), and 58 participants who reported sexual behavior during the previous three months returned one week later for a second interview. Test–retest intraclass correlations ( $\rho$ ) from the TLFB protocol showed that all sexual behaviors were reported reliably ( $\rho$  *range* = .86 to .97, *median* = .96). Bootstrapping, a non-parametric statistical technique, was used for significance testing in the reliability analyses. Reliability was equivalent across each of the three months assessed with the TLFB and was equivalent to conventional assessment methods (i.e. single-item questions). These findings show that the TLFB sexual behavior interview provides reliable reports of sexual behavior over three months and yields event-level data that are extremely valuable for sexual behavior and HIV-prevention research.*

(*Ann Behav Med* 1998, 20(1):25–30)

## INTRODUCTION

Measurement of sexual behavior is crucial to human immunodeficiency virus (HIV) prevention research, but obstacles related to recall are often not addressed in this domain. Croyle and Loftus (1) 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 related to recall that can contribute to unreliable self-reports. Further, other factors such as intentional overreporting and underreporting of sensitive behaviors are suspected to bias results (2). Because there are limited assessment alternatives, retrospective

self-report sexual behavior data are used to determine the prevalence of behaviors that place individuals at risk for HIV infection, track the transmission of HIV, identify antecedents of risk behavior, identify subpopulations to be targeted for intervention, and evaluate interventions designed to reduce risk behavior (3). Therefore, it is important that researchers use measures of sexual behavior that are psychometrically sound. However, measures typically used in HIV-prevention research are not generally standardized or tested for reliability.

A promising technique to assess sexual behavior is the Timeline Follow-Back method (TLFB) (4,5). The TLFB was first used to obtain retrospective self-reports of daily alcohol consumption during a specified reporting period. Several interview aids are used to facilitate recall, including presenting respondents with a calendar, marking salient events in the reporting period (e.g. hospitalizations), identifying lengthy periods of abstinence or patterned drinking (e.g. drinking only on weekends), and anchoring the highest and lowest quantity consumed in the target interval. In a brief interview, data are yielded regarding the number of drinking days, as well as the average and total quantity consumed during drinking episodes. Advantages of the TLFB procedure include the extensive information yielded on the patterning of drinking behavior, as well as the ability to summarize data on quantity and frequency over a variety of intervals.

Research on the TLFB resulted in its current use to assess alcohol, marijuana, cocaine, and inhalant use (6). Furthermore, it has been administered successfully to a variety of populations including problem drinkers in inpatient, outpatient, and residential alcohol treatment (7); normal drinkers (8); and psychiatric outpatients (9). In addition to the original face-to-face interview format, the protocol has been adapted for phone interviews and computerized assessment (10). Reporting periods ranging from one month to one year have been used. There is ample evidence that the TLFB method yields reliable and valid self-reports of socially-sensitive substance use behaviors (6,11).

A major strength of the TLFB method is that it yields event-level data. Event-level inquiry allows more detailed assessment and analysis than traditional single-item quantity or frequency assessments. To illustrate this benefit, consider investigations of the association between substance use and risky sexual behavior. Studies that have used global assessments of frequencies of substance use and sexual behavior typically have found an association between substance use and unprotected sex (e.g. 12,13). However, these studies do not provide evidence that drinking and unsafe sexual behavior occurred on the same occasions.

In the first published application of a modified version of the TLFB to assess substance use and sexual behavior, Crosby et al. (14) studied 131 gay and bisexual men. Event-level data revealed, for example, that men who had unprotected anal sex consistently under the influence of alcohol and/or drugs had less education, lower income, and were more likely to use amyl nitrate and/or

<sup>1</sup> Preparation of this manuscript was supported in part by grants from the National Institute of Mental Health to Lance S. Weinhardt (MH11125) and Michael P. Carey (MH01101 and MH54929), National Institute on Alcohol Abuse and Alcoholism to Stephen A. Maisto (AA10291), and National Institute on Drug Abuse to Kate B. Carey (DA07635).

<sup>2</sup> The authors thank John R. Gleason and Andrew D. Forsyth for statistical consultation and Christopher M. Gordon for comments on an earlier draft of this paper.

Reprint Address: M. P. Carey, Ph.D., Department of Psychology, 430 Huntington Hall, Syracuse University, Syracuse, NY 13244-2340.

cocaine. Thus, the use of event-level data allowed for a more fine-grained analysis of the relationship between substance use and sexual behavior. Consequently, one advantage of the TLFB approach is that it allows event-level analysis of risk behavior, which is valuable for intervention planning and evaluation. Single-item frequency measures typically used in HIV-prevention research cannot provide such information.

The modified TLFB interview that Crosby et al. (14) used is a significant methodological advance and suggests that the TLFB can be applied to self-reports of sexual behavior. However, there were three limitations to Crosby et al.'s research. First, the interview assessed only protected and unprotected anal intercourse. This may reduce its value for use with populations other than men who have sex with men. Second, the interview assessed behavior occurring one month prior to the most recent sexual encounter. Because the frequency of sexual behavior can vary widely over time for an individual, a longer reporting period needs to be examined. Finally, Crosby et al. did not evaluate the reliability of their measure. Therefore, the purposes of this project were to extend the pioneering efforts of Crosby et al. by: (a) adapting the TLFB technique to assess a variety of sexual behaviors; (b) extending the interval to three months, a more common reporting period in HIV-prevention research (15); (c) evaluating the test-retest reliability of the instrument in a face-to-face interview format; and (d) comparing the reliabilities of the TLFB variables to those of commonly-used single-item sexual behavior frequency questions.

## METHOD

### Participants

One hundred ten college students (51 women, 59 men;  $M$  age = 19.7; range = 18–41) enrolled in an introductory psychology course and participated for course credit. This is a sexually active population, which is at risk for a variety of sexually transmitted diseases (STDs); among college students, 1 in 500 are HIV-positive (16). Ninety percent of the participants were in their first or second year of college.

### Materials

*TLFB Sexual Behavior Interview:* The protocol was adapted from the current TLFB manual and materials (6). Specific sexual behaviors (e.g. insertive and receptive vaginal, oral, and anal intercourse) were defined explicitly to minimize ambiguity. Consistent with recommendations for sexual behavior interviewing (17,18), participants were encouraged to use terms they preferred for these behaviors, and interviewers used the participants' terms throughout the interviews.

Next, interviewers presented a calendar that included the beginning and end dates of the reporting period, with campus events and holidays already identified. Participants indicated other days in the reporting period that were memorable for them, such as exams, road trips, newsworthy events, sporting events, and family visits, by writing the event on the calendar. Participants were instructed to identify all days on which they were sexually active. To do this, participants were prompted to consider the memorable events they had recalled and any patterns to their sexual encounters (e.g. weekend visits to partners). Beginning with the most recent sexual activity and for every sexual event, participants were asked a series of questions to assess: (a) the type of relationship with the sexual partner (i.e. monogamous, non-monogamous); (b) type and number of sexual activities; (c) the occurrence of possible antecedents of HIV-risk behavior (e.g. alcohol consumption); and (d)

HIV-preventive behavior (e.g. talking with a partner about using condoms before having sex). Interviewers coded participant responses directly on the TLFB calendar on each day that sexual activity was reported.

*Standard Sexual Behavior Questions:* A series of 28 single-item (SI) sexual behavior frequency questions was also included. For example, one SI was "How many times in the past three months did you have vaginal intercourse without using a condom?" These items were similar to those recommended for assessment by Kelly (19) and to those commonly used in HIV-prevention research (e.g. 20–23). Each SI frequency question corresponded to a sexual behavior variable yielded by the TLFB interview.

### Interviewers

One male and one female research assistant were trained to conduct interviews with participants of the same gender. Training focused on sexual behavior interviewing, HIV-prevention, and research ethics and confidentiality. Interviewers were trained through several role-played interviews simulating typical and difficult interview situations, which were observed by the first author. They received supervision throughout the course of the investigation.

### Procedure

Interviews were conducted individually in sound-attenuated private offices. After obtaining informed consent, the interviewer audiotaped the interview, which was later reviewed for adherence to the interview script. Participants completed the demographics questionnaire. Next, the interviewer conducted the structured interviews including the questions regarding general health, sexual functioning, history of sexually transmitted diseases, the single-item frequency questions, and then the TLFB interview was administered. The order in which questions were asked was planned to gradually increase the sensitivity of content, which is advised in sexual behavior interviewing (18). The order of the sexual behavior measures (single-item questions followed by the TLFB procedure) was planned to move from less detailed to more detailed questioning. Finally, participants completed a debriefing questionnaire. Participants who reported vaginal, oral, or anal intercourse during the reporting period were invited to return one week later for a second interview.

During the second interview, participants were reminded of the purpose and confidential nature of the project. They were reinterviewed about sexual behavior with the single-item frequency questions and the TLFB procedure. Behaviors were assessed for the same three-month reporting period used in the first interview.

### Data Analysis

Two characteristics of the data influenced our approach to reliability analysis. First, as is the case with sexual behavior and other low base-rate behaviors, the data distributions were positively skewed (i.e. for each behavior, many people reported a frequency of 0 or 1, whereas a smaller proportion reported a frequency of 2 or more). This violates the assumption of normality on which tests of significance of correlation coefficients are based. Second, each comparison of TLFB to SI reliability and each comparison of the reliability of individual months assessed with the TLFB involved four variables (e.g.  $r_{wx}$  and  $r_{yz}$ ). Therefore, a traditional approach to comparing dependent correlations (24, pp.

215–216), which is based on the relations among three variables (e.g.  $r_{xy}$  and  $r_{xz}$ ), could not be applied.

Because of these two characteristics, significance testing for the reliability analyses was done with the aid of bootstrapping (25) performed with Stata statistical software (26). The advantage of bootstrapping to evaluate the magnitude of correlation coefficients is that there is no assumption that the population distribution is Gaussian in form (27). Bootstrapping is a non-parametric technique that simulates the effects of repeated sampling of the variables being analyzed from the population of interest by repeatedly sampling from the available sample. Calculating a correlation from each such resample creates an approximation to the sampling distribution of the correlation, an approximation that imposes no assumptions about the population distribution. Thus, the magnitude of observed correlations can be evaluated relative to what would be expected from this distribution.

In our reliability analyses, the bootstrap technique was implemented as follows: A random sample of size  $N$  (in this case, 58) of the variable pairs (e.g. number of occasions of unprotected vaginal intercourse reported at first and second assessments) was selected from the data set, with replacement. The reliability coefficient was computed from this randomly-selected sample and converted to a  $z$  score using Fisher's  $r$  to  $z$  transformation. This procedure was repeated 1,000 times for each behavior, resulting in a distribution of Fisher's  $z$  transformations from which the mean, bias, standard error, and 95% confidence interval (CI) were calculated. When comparing two reliability coefficients (i.e. reliability of individual months of the TLFB assessment or TLFB versus SI reliability), the same procedure was implemented with random selection of sets of four variables with replacement; the Fisher's  $z$  distributions were generated simultaneously for the two coefficients. From the resulting distribution, the mean, bias, standard error, and 95% CI were obtained for each of the two reliability coefficients and for the difference between the two reliability coefficients.

## RESULTS

No participants refused to complete the initial interview. Overall, participants rated their experience with the study as "somewhat" interesting and "not-at-all" to "slightly" embarrassing. Participants indicated that they were very comfortable with the protocol and that they viewed the study as "somewhat" to "very" important. Sixty-four (58%) of the participants reported sexual behavior during the previous three months; these individuals were invited to return for a second interview approximately one week later (M test-retest interval = 9.2 days, SD = 2.8, range = 6–16 days). Five eligible men declined to participate in the second interview because they had completed the required hours of research participation for course credit; one woman could not be recontacted. Thus, the retest sample consisted of 58 individuals (50% women), 91% of the eligible participants. The single-item frequency questions required approximately five minutes to administer. The time required for the TLFB portion of the interview varied from 5 to 20 minutes as a function of frequency of sexual behavior.

### Frequencies of Sexual Behaviors

Analyses using the Bonferroni adjustment for multiple comparisons indicated that the SI and TLFB assessments yielded equivalent frequencies on all sexual behavior variables. Table 1 reports the percentage of the sexually active retest sample that reported each behavior. Several behaviors, including the use of a

TABLE 1  
Proportion of Participants Endorsing Sexual Behaviors at First Assessment, by Mode ( $N = 58$ )

Behavior (in the past three months)	TLFB	SI
1 or more vaginal sex partners	* 83%	81%
1 or more times with a condom	* 50%	55%
1 or more times without a condom	* 52%	54%
1 or more male oral sex partners	* 36%	38%
1 or more times giving oral sex	* 30%	36%
1 or more times receiving oral sex	* 28%	33%
1 or more female oral sex partners	* 47%	45%
1 or more times giving oral sex	* 41%	26%
1 or more times receiving oral sex	* 43%	41%
1 or more times using a barrier during any type of oral sex	4%	2%
1 or more anal sex partners	5%	5%
1 or more times giving anal sex (insertive partner)	0%	0%
1 or more times receiving anal sex (receptive partner)	5%	5%
Had sex after drinking alcohol	* 47%	47%
Talked with a partner about HIV/AIDS	* 36%	45%
Talked with a partner about using a condom before having sex	* 54%	60%

Note: TLFB = data obtained with Timeline Follow-Back procedure; SI = data obtained with single-item frequency questions; \* = item retained for reliability analyses.

barrier (e.g. dental dam) during any type of oral sex and anal sex (protected and unprotected, insertive and receptive), were not reported by a sufficient number of participants to compute reliability coefficients and were not included in the reliability analysis.

### Test-Retest Reliability

The test-retest reliability of participants' responses on each behavior variable was computed using the intraclass correlation ( $\rho$ ). When computing the intraclass correlation for each behavior, we retained data from participants who reported a frequency of zero for that behavior (see Table 1 for proportion of participants who endorsed each item), because all participants in the final sample were sexually active. Table 2 reports the test-retest reliability intraclass correlation coefficients for each behavior by gender. Table 3 reports the test-retest reliability intraclass correlation coefficients for each reported behavior by assessment method and by month assessed with the TLFB.

Analysis of reliability by assessment mode and gender for each behavior (see Table 2) revealed no significant differences between women's (TLFB range  $\rho = .79$  to  $.98$ , median  $\rho = .94$ ; SI range  $\rho = .71$  to  $.98$ , median  $\rho = .90$ ) and men's (TLFB range  $\rho = .86$  to  $.97$ , median  $\rho = .96$ ; SI range  $\rho = .62$  to  $.96$ , median  $\rho = .83$ ) reports. Using the bootstrapping technique described earlier, the magnitude of difference between reliability of TLFB and single-item frequency questions by gender was evaluated for each behavior by using bootstrapping to construct a 95% normal theory<sup>3</sup> CI around the difference between the mean Fisher's  $z$  transformation scores. Zero was contained within the 95% CI for each behavior. Comparisons of the reliabilities between TLFB

<sup>3</sup>Stata's bootstrapping function produces three versions of confidence intervals: normal theory, percentile, and bias corrected. Because the three types of confidence intervals were indistinguishable for practical purposes and the Fisher's  $z$  transformation distributions were Gaussian, the normal theory confidence intervals were used for all analyses.

**TABLE 2**  
Timeline Follow-Back and Single-Item Test-Retest Intraclass Correlation Coefficients, by Gender

Behavior (in the past three months)	Women (N = 29)		Men (N = 29)	
	SI	TLFB	SI	TLFB
How many partners did you have vaginal sex with?	.71	.94	.96	.99
How many times with a condom?	.98	.93	.91	.99
How many times without a condom?	.90	.98	.79	.86
How many people of the opposite sex did you have oral sex with?	.93	.94	.84	.91
How many times did you give oral sex?	.90	.96	.94	.92
How many times did you receive oral sex?	.79	.97	.80	.96
Had sex after drinking alcohol	.90	.92	.83	.89
Talked with a partner about HIV/AIDS	.89	.97	.62	.81
Talked with a partner about using a condom before having sex	.79	.79	.62	.84

*Note:* Reliability coefficients are intraclass correlations; TLFB = data obtained with Timeline Follow-Back procedure; SI = data obtained with single-item frequency questions.

responses from men and from women and between SI responses from men and from women were conducted in a similar manner. Again, zero was contained within the 95% CI for each comparison. Thus, the magnitude of reliability did not differ as a function of gender. Because there were no gender differences, men's and women's data were combined for subsequent analyses.

In the total sample, the TLFB yielded high reliability coefficients for all behaviors (TLFB range  $\rho = .86$  to  $.97$ , median  $\rho = .96$ ) (see Table 3). Using bootstrapping, we constructed 95% CI around the mean Fisher's  $z$  transformation for each behavior. Zero was not included in the CI for any behavior, indicating that TLFB test-retest coefficients were significant for all behaviors, taking into account the positively-skewed distributions. The single-item frequency questions resulted in reliability coefficients ranging from  $\rho = .81$  to  $.96$ ; median  $\rho = .89$  (see Table 3). The magnitude of difference between reliability of TLFB and single-item frequency questions was evaluated for each behavior by using bootstrapping to construct a 95% CI around the difference between the mean Fisher's  $z$  transformation scores. Zero was contained within the 95% CI for each behavior. Thus, the magnitude of the correlations did not differ as a function of method.

Table 3 also contains the test-retest reliability coefficients for each behavior by each month assessed with the TLFB protocol. Bootstrap analysis revealed that zero was contained within the 95% CI of the difference between the mean Fisher's  $z$  transformation scores compared for each behavior. Thus, reliability did not differ significantly across the three months for any behavior.

## DISCUSSION

The primary finding emerging from this investigation was that the Timeline Follow-Back techniques resulted in test-retest reliability coefficients between  $\rho = .86$  and  $.97$  for all behaviors reported. These coefficients, as well as those from the SI frequency questions, were well within the acceptable range for both research and clinical use (28, pp. 264-265). Reliability remained stable for each of the three months assessed, indicating that the TLFB protocol resulted in reliable reporting for a period of up to three months with no appreciable drop-off in reliability after the first or second month. Kauth, St. Lawrence, and Kelly (29) suggested that

**TABLE 3**  
TLFB and Single-Item Test-Retest Intraclass Correlation Coefficients (N = 58)

Behavior (in the past three months)	Month of TLFB Assessment			Assessment Method	
	Most Recent Month	Middle Month	Most Distant Month	TLFB	SI
How many partners did you have vaginal sex with?	.87	.95	.92	.97	.81
How many times with a condom?	.98	.88	.93	.96	.93
How many times without a condom?	.94	.86	.90	.93	.87
How many men did you have oral sex with?	.85	.86	.94	.96	.96
How many times did you give oral sex?	.97	.94	.96	.96	.92
How many times did you receive oral sex?	.89	.78	.80	.91	.82
How many women did you have oral sex with?	.86	.86	.90	.96	.94
How many times did you give oral sex?	.88	.92	.91	.93	.95
How many times did you receive oral sex?	.88	.94	.90	.94	.83
Had sex after drinking alcohol	.84	.88	.84	.91	.86
Talked with a partner about HIV/AIDS	.84	.84	.96	.97	.84
Talked with a partner about using a condom before having sex	.94	.80	.78	.86	.89

*Note:* TLFB = data obtained with Timeline Follow-Back procedure; SI = data obtained with single-item frequency questions.

three months may be the upper limit for acceptably reliable reporting of sexual behavior using SI questions. The ability of the TLFB to produce reliable reports for longer reporting periods warrants further investigation.

Based on these data and other reliability-evaluation studies involving single-item sexual behavior frequency questions (e.g. 30-32), it appears that both the SI and the TLFB methods are reliable. However, the TLFB interview has the added benefit of producing more specific and useful event-level information (33). This is particularly important for investigating the correlates of HIV-related risk-taking and for evaluating the effectiveness of risk-reduction interventions. It would be difficult to determine with single-item frequency questions, for example, whether participants are more or less likely to talk with a sexual partner about condom use when they have been drinking and whether this relationship changes as a result of an intervention. However, data collected with the TLFB procedure allow this type of analysis (14). Other examples include evaluating the degree of co-occurrence of HIV risk behavior with risk factors such as substance use and/or mood changes (34), as well as preventive factors, such as assertive communication about prevention with partners (35). Such relationships can be identified in target populations and addressed in interventions. The effectiveness of interventions designed to target the hypothesized determinants of risky sex can then be evaluated at the event level.

The need for this type of event-level information must be weighed against the resources available for assessment. At present,

the TLFB sexual behavior assessment has been examined only in a face-to-face interview format, which requires more personnel resources than does a survey administered to groups of participants. Also, because of the specificity of the data gathered, a single administration of the TLFB typically takes more time than traditional single-item frequency questions. Given these relative costs, the unique level of analysis afforded by TLFB data will not be necessary or desirable in all studies of sexual behavior.

Two limitations of this study should be acknowledged. First, it is possible that the order in which content was presented affected participant self-reports. However, prior studies with college students found that order did not influence measurement error for sexual questions in face-to-face interview format (36). Second, we did not assess the validity of the TLFB. The goal of this initial investigation was to assess the reliability of the TLFB sexual behavior interview, a prerequisite to establishing its validity. Validation of self-report measures of sexual behavior is challenging because direct observation is neither ethical nor practical. Validation using strategies such as collateral partner interviews and concurrent self-monitoring data is needed.

We conclude that the TLFB sexual behavior interview yields reliable event-level data on self-reports of sexual HIV-risk behaviors and HIV-preventive behaviors for a three-month reporting period. Further research is encouraged to determine if the TLFB provides reliable data over longer intervals and to evaluate evidence for the validity of self-reported sexual behavior. In addition, it would be useful to determine the importance of the interview-facilitated memory aids for eliciting reliable self-reports. Although both the SI and TLFB approaches provide reliable estimates of sexual behavior, the TLFB also provides additional event-level information that is needed for more precise understanding of the context of risk behavior and the design and evaluation of HIV-prevention interventions.

## REFERENCES

- (1) Croyle RT, Loftus EF: Recollection in the kingdom of AIDS. In Ostrow DG, Kessler RC (eds), *Methodological Issues in AIDS Behavioral Research*. New York: Plenum, 1993, 163–180.
- (2) Catania JA, Binson D, van der Straten A, Stone V: Methodological research on sexual behavior in the AIDS era. *Annual Review of Sex Research*. 1995, 4:77–125.
- (3) 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. *Psychological Bulletin*. 1990, 108:339–362.
- (4) Sobell LC, Maisto SA, Sobell MB, Cooper AM: Reliability of alcohol abusers' self-reports of drinking behavior. *Behaviour Research and Therapy*. 1979, 17:157–160.
- (5) Sobell LC, Sobell MB: Timeline follow-back: A technique for assessing self-reported alcohol consumption. In Litten RZ, Allen J (eds), *Measuring Alcohol Consumption: Psychosocial and Biological Methods*. Totowa, NJ: Humana Press, 1992, 41–72.
- (6) Sobell LC, Sobell MB: *Timeline Followback User's Guide*. Toronto, Canada: Alcohol Research Foundation, 1997.
- (7) Maisto SA, Sobell MB, Cooper AM, Sobell LC: Test-retest reliability of retrospective self-reports in three populations of alcohol abusers. *Journal of Behavioral Assessment*. 1979, 1:315–326.
- (8) Connors GJ, Watson DW, Maisto SA: Influence of subject and interviewer characteristics on the reliability of young adults' self-reports of drinking. *Journal of Psychopathology and Behavioral Assessment*. 1985, 7:365–374.
- (9) Carey KB: Reliability and validity of the timeline follow-back interview among psychiatric outpatients: A preliminary report. *Psychology of Addictive Behaviors*. 1997, 11:26–33.
- (10) Sobell LC, Sobell MB, Leo GI: Reliability of the alcohol timeline followback when administered by telephone and by computer. Twenty-Ninth Annual Convention of the Association for the Advancement of Behavior Therapy. Washington, DC: November 1995.
- (11) Ehrman RN, Robbins SJ: Reliability and validity of six-month timeline reports of cocaine and heroin use in a methadone population. *Journal of Consulting and Clinical Psychology*. 1994, 62:843–850.
- (12) Siegel K, Mesagno FP, Chen JY, Christ G: Factors distinguishing homosexual males practicing risky and safer sex. *Social Science and Medicine*. 1989, 28:561–569.
- (13) Valdiserri RO, Lyter D, Leviton LC, et al: Variables influencing condom use in a cohort of gay and bisexual men. *American Journal of Public Health*. 1988, 78:801–805.
- (14) Crosby GM, Stall RD, Paul JP, Barrett DC, Midanik LT: Condom use among gay/bisexual male substance abusers using the timeline follow-back method. *Addictive Behaviors*. 1996, 21:249–257.
- (15) Kalichman SC, Carey MP, Johnson BT: Prevention of sexually transmitted HIV infection: A meta-analytic review of the behavioral outcome literature. *Annals of Behavioral Medicine*. 1996, 18:4–13.
- (16) Gayle HD, Keeling RP, Garcia-Tunon M, et al: Prevalence of HIV infection among college and university students. *New England Journal of Medicine*. 1990, 323:1538–1541.
- (17) Carey MP: Assessing and reducing risk of infection with the human immunodeficiency virus (HIV). In Koocher GP, Norcross JC, Hills SS (eds), *Psychologist's Desk Reference*. New York: Oxford University Press (in press, 1998).
- (18) Wincze JP, Carey MP: *Sexual Dysfunction: A Guide for Assessment and Treatment*. New York: Guilford, 1991.
- (19) Kelly JA: *Changing HIV Risk Behavior: Practical Strategies*. New York: Guilford, 1995.
- (20) 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 Mental Health Journal*. 1997, 33:133–142.
- (21) Carey MP, Maisto SA, Kalichman SC, et al: Motivational enhancement, behavioral skills training, and education to reduce risk of HIV infection in urban women. *Journal of Consulting and Clinical Psychology*. 1997, 65:531–541.
- (22) Kalichman SC, Sikkema K, Kelly JA, Bulto M: Use of a brief behavioral skills intervention to prevent HIV infection among chronic mentally ill adults. *Psychiatric Services*. 1995, 46:275–280.
- (23) Kelly JA, Murphy DA, Washington CD, et al: The effects of HIV/AIDS intervention groups for high-risk women in urban clinics. *American Journal of Public Health*. 1994, 84:1918–1922.
- (24) Bruning JL, Kintz BL: *Computational Handbook of Statistics* (2nd Ed.). Glenview, IL: Scott, Foresman, 1977.
- (25) Efron B, Tibshirani RJ: *An Introduction to the Bootstrap*. New York: Chapman & Hall, 1993.
- (26) StataCorp: *Stata Statistical Software: Release 5.0*. College Station, TX: Stata Corporation, 1996.
- (27) Lunneborg CE: Estimating the correlation coefficient: The bootstrap approach. *Psychological Bulletin*. 1985, 98:209–215.
- (28) Nunnally JC, Bernstein IH: *Psychometric Theory* (3rd Ed.). New York: McGraw-Hill, 1994.
- (29) 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 Education and Prevention*. 1991, 3:207–214.
- (30) Dowling-Guyer S, Johnson ME, Fisher DC, et al: Reliability of drug users' self-reported HIV risk behaviors and validity of self-reported drug use. *Assessment*. 1994, 1:383–392.
- (31) McKinnon K, Courmos F, Meyer-Bahlburg HFL, et al: Reliability of sexual risk behavior interviews with psychiatric patients. *American Journal of Psychiatry*. 1993, 150:972–974.

- (32) Needle R, Fisher DG, Weatherbee N, et al: Reliability of self-reported HIV risk behaviors of drug users. *Psychology of Addictive Behaviors*. 1995, 9:242–250.
- (33) Temple MT, Leigh BC: Alcohol consumption and unsafe sexual behavior in discrete events. *Journal of Sex Research*. 1992, 29:207–219.
- (34) Carey MP, Carey KB, Kalichman SC: Risk for human immunodeficiency virus (HIV) infection among persons with severe mental illness. *Clinical Psychology Review*. 1997, 17:271–291.
- (35) Fisher JD, Fisher WA: Changing AIDS risk behavior. *Psychological Bulletin*. 1992, 111:455–474.
- (36) DeLamater J, MacCorquodale P: The effects of interview schedule variations on reported sexual behavior. *Sociological Methods Research*. 1975, 4:215–236.