A Comparison of Long- vs. Short-Term Recall of Substance Use and HIV Risk Behaviors

TIM JANSSEN, PH.D.,^{a,b,*} JORDAN M. BRACISZEWSKI, PH.D.,^c ADAM VOSE-O'NEAL, M.S.W.,^a & ROBERT L. STOUT, PH.D.^a

^a*Decision Sciences Institute/Pacific Institute for Research and Evaluation, Pawtucket, Rhode Island*

^b*Center for Alcohol and Addiction Studies, Brown University School of Public Health, Providence, Rhode Island*

^c*Center for Health Policy & Health Services Research, Henry Ford Health System, Detroit, Michigan*

ABSTRACT. Objective: The Timeline Follow-back (TLFB) questionnaire has become a pre-eminent tool in substance use and human immunodeficiency virus (HIV) risk research, allowing researchers to assess fine-grained changes in risk behavior over long periods. However, data on accuracy of recall over long (12-month) periods are sparse, especially combined data on HIV risk and substance use from post-treatment samples. Studies on the development of substance use and HIV risk stand to benefit from data on the accurate recall of such behavior over longer retroactive spans of time. **Method:** The present study offers data

on the test–retest reliability of current TLFB assessment versus 6- and 12-month delayed TLFB assessment, using a post-treatment sample ($n = 50$). **Results:** Long-term reliability of TLFB data on HIV risk was predominantly good to excellent, with 13 of 20 assessed variables in that range. TLFB data on substance use was similar, with 22 of 26 variables resulting in good/excellent reliability. **Conclusions:** Our findings support the notion that, notable exceptions aside, the TLFB may be effectively used to assess retroactive HIV risk and substance use in periods of 12 months. (*J. Stud. Alcohol Drugs*, 78, 463–467, 2017)

THE USE OF LONGITUDINAL DATA has allowed researchers in the field of substance use to advance knowledge on time-based relations to escalation and relapse of substance use. However, gathering longitudinal data often comes with great expense and difficulty, as participants must be retained and re-tested over the studied timeframe. For this reason, researchers have often used retroactive assessment of substance use. The window of retroactive assessment is typically limited by the recall of the participant, which may be particularly challenging when the participant is recovering from a substance use disorder. In addition, assessment of human immunodeficiency virus (HIV) risk is often complex because of the stigmatized nature of the assessment subject. This uncertainty has led researchers to highlight the need for more data collection addressing the reliability of different recall periods (see Napper et al., 2010, for a review). To date, few studies have examined whether long-term recall yields reliable estimates of substance use and HIV risk. This requires comparing the accuracy of immediate recall to extended recall assessed at a later date. The current study offers data on the accuracy of long-term recall assessed at 6- and 12-month intervals using Timeline Follow-back methods (TLFB; Sobell & Sobell, 1992).

In posttreatment populations, noted limitations of ac-

curately assessing substance use and HIV risk behavior include the sensitivity of the subject to addicted persons (Harrison, 1997), the complication of sexual intercourse in exchange for drugs, desirability of abstinent outcomes (Yoshino & Kato, 1995), impairment of recall resulting from substance use (Wetherill & Fromme, 2011), mental deficits associated with chronic substance use (Rogers & Robbins, 2001), and the rapid natural change of substance use behavior (McLellan et al., 2000). Despite these limitations, 30-day, 90-day, 3-month, and 6-month recall of substance use using self-report methodology has proven fairly reliable according to a recent meta-analysis (Napper et al., 2010). This meta-analysis compared reliability using in-treatment populations, homeless adults, and psychiatric outpatients. Limitations of these reliability estimates include the lack of substance-specific reliability estimation (such as the specific reliability of assessments of opioid use relative to alcohol use), inconsistency in self-report methods, and inconsistency in joint assessment of HIV risk and substance use outcomes. A recent report of the 1-week test–retest reliability of 90-day retrospective substance use assessment (Wray et al., 2016) demonstrated sizeable differences in the reliability of different substances, with recall of non-heroin opioid use and sedative use being noted as particularly poor compared with others. Similarly, in terms of HIV risk assessment using the same method, it was found that definition of relationships (particularly defining relationships as casual) led to lack of reliability, whereas reliability for other types of HIV risk assessment was excellent.

By using TLFB methods, researchers hope to reduce error by enumeration (listing occurrences of a certain risk behavior over time) and mental calculation (relying on subject calculations for the product of typical behavior given a set

Received: July 28, 2016. Revision: December 20, 2016.

This research was supported by National Institute on Drug Abuse Grant R01DA031154 (principal investigator: Robert L. Stout) and National Institute on Alcohol Abuse and Alcoholism Grant T32-AA007459 (principal investigator: Peter Monti).

*Correspondence may be sent to Tim Janssen at the Center for Alcohol and Addiction Studies, Brown University School of Public Health, Box G-S121-5, Providence, RI, 02912, or via email at: tim_janssen@brown.edu.

period). Instead of this, TLFB methods rely on a calendar approach, where trained interviewers ask participants to recall risk behavior on specific dates. We examined whether the lack of reliability for events with lower base rates may be exacerbated when the window of recall is extended, whereas reliability for higher base rate events may be less strongly affected. In addition, the use of TLFB methods to assess HIV risk has often been limited to the assessment of the number of sex partners or the frequency of engaging in specific sexual activities (such as vaginal, anal, or oral). This approach ignores much of the context of sexual relations as they pertain to HIV risk, such as condom use, simultaneous substance use, and the degree to which partners are casual. Therefore, we examine how long-term recall affects recall of different partner types, serostatuses, partner genders, and sex under influence.

The present study seeks to further knowledge on the reliability of long-term retrospective assessment of substance use and HIV risk using the TLFB method. To examine this reliability, frequencies of different types of substance use and the number of different types of sexual activity are assessed both as a 90-day retrospective (i.e., behavior in the last 90 days) and with a 6- or 12-month delay to assess that same 90-day period.

Method

Participants

Fifty participants recruited from substance abuse treatment centers were selected from a longitudinal study examining the relation between substance misuse and HIV risk relapse and the clients' social network, particularly the start, termination, and supportive nature of important social relations. (Copies of the study are available from the first author on request.) Participants were eligible for the main study if they (a) were enrolled in substance abuse treatment, (b) were at least 18 years old, (c) had used their primary substance within the past 6 months, (d) met *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition* (American Psychiatric Association, 1994), criteria for substance abuse and dependence on their primary substance within the last 12 months, (e) did not have legal issues that would limit their participation, and (f) were able to read and write English. Participants were recruited from intensive outpatient programs where they would attend therapeutic groups for 4 hours per day, at least 3 days per week. Treatment lengths rarely exceeded 4 weeks. Clinicians from intensive outpatient programs encouraged clients to seek aftercare. In this recall sub-study, some participants were in outpatient treatment and some were not. All procedures were in accordance with the ethical standards of the institutional research committee. Informed consent was obtained from all individual participants included in the study.

Procedure

For the current study, a subsample of participants who had been enrolled in the main study for at least 12 months was selected and offered the opportunity to re-assess their reported substance use from 6 months and 12 months previous. This was a sequential sample; the first 50 consenting subjects were enrolled. Compensation for this extra interview was a \$25 gift card. This opportunity was presented as optional, with performance and attendance not affecting their ongoing participation in the main study. If participants agreed, the reliability sub-study questionnaires were completed during regularly scheduled interviews for the main study. The reliability data were always collected first. These additional surveys were completed without attrition by all 50 participants. They were representative of the main sample on race, ethnicity, sexual preference, use of technology, social networking, cell phone use, marital status, health insurance, living arrangements, yearly income, religion, parole status, legal problems, primary drug of choice, age, gender, schooling, convictions, and incarceration but not health insurance.

Measures

At each interview in the main study, the TLFB (Sobell & Sobell, 1992) was used to assess participants' alcohol use, drug use, and engagement in HIV risk behaviors over the past 90 days. Alcohol was assessed on a daily basis, whereas the other variables were assessed on a weekly basis, with "Week 1" representing the week most proximal to the present and "Week 12" being the most distal. Alcohol use was assessed as the number of drinking days and "heavy drinking" days over a given week, with "heavy drinking" referring to days on which five or more standard drink units for men, or four or more for women, were consumed. Participants were also asked to indicate their use of any marijuana, cocaine, heroin, (other) opiates, sedatives, and/or "other" drugs on a given week. To aid in accuracy, participants were shown cards with comprehensive lists of drugs included in each category. HIV risk behaviors were assessed by first providing detailed definitions of "sexual activity," "partner," and "HIV status" for participants.

Interviewers inquired about the total number of sex partners that participants had over the previous 90 days, asking participants to subset these by those with known HIV-negative, HIV-positive, and unknown serostatuses. Participants were then asked the total number of sex partners they had for each week during the 90-day recall period, as well as the number who were male, female, steady, and casual. Finally, if participants indicated they had had partners with unknown or seropositive status, they were asked to indicate the frequency with which they used condoms and engaged in sex under the influence for each week on a 5-point scale, ranging from 0 (*never*) to 4 (*always*). For this long-term recall reli-

ability study, participants were asked to re-create their TLFB data from 6 and 12 months earlier. For example, participants who were interviewed on May 1, 2016, would be asked to recall their TLFB data from their December 1, 2015, interview (covering, roughly, September 1 to November 30) and their May 1, 2015, interview (covering, roughly, February 1 to April 30).

Analysis strategy

Within the 90-day monitoring window, mean monthly scores on the weekly frequency of substance use were calculated. To represent HIV risk behavior, mean monthly scores on the weekly number of partners were calculated. To accurately assess the long-term recall reliability of events with low base rate (see the supplemental materials that accompany the online version of this article for frequencies), events were also re-coded into dichotomous variables. We then calculated test–retest reliability using intra-class correlations (Shrout & Fleiss, 1979) and Cohen's kappa for consistency (Cohen, 1960). Paired-sample *t* tests estimated changes in level of substance use across reporting periods to test for drift. Tests were run using IBM SPSS Statistics for Windows, Version 24 (IBM Corp., Armonk, NY), and SAS Version 9.4 (SAS Institute, Inc., Cary, NC).

In addition, for condomless sex, a “not appropriate” category was created when participants had previously indicated not having sex with non-HIV-negative partners ($n = 39$ at 6 months, $n = 41$ at 12 months). For sex under influence, this category was created when not having sex with non-HIV-negative partners and/or not using substances. Cohen's kappa for consistency was then calculated assuming three answer categories (*not appropriate*; *yes, occurred*; *no, did not occur*).

Results

Demographics

Participants' mean age at the study's baseline was 39.79 years ($SD = 10.59$). Eighteen participants were female, and 32 were male. Six participants indicated a race other than White, whereas 14 indicated Hispanic ethnicity. Four participants indicated that their primary sexual preference was for their own gender. Of the 50 participants, 20 indicated that their primary drug of choice was alcohol, 14 indicated cocaine, 7 indicated heroin, 7 indicated opiates, 1 indicated marijuana, and 1 indicated sedatives.

Test–retest reliabilities

Table 1 offers intra-class correlations and Cohen's kappa estimates for the test–retest pairings of initial assessments and 6- and 12-month assessments. For zero rates over 80%,

interpretation of agreement scores rather than intra-class correlations is recommended (and thus emphasized in table as bold). When interpreting bold scores, results indicate that the accuracy of reassessments appears to be very good, with 37 of 46 reliability indices ranging from good to excellent ($<.40$ [poor], $.40-.59$ [fair], $.60-.74$ [good], and $.75-1.00$ [excellent]; Cicchetti et al., 1994).

Nearly all 6-month reliability estimates for substance use were excellent; exceptions were for heavy drinking (good), heroin (fair), and non-heroin opiates (fair). Similarly, most 6-month indices for risky sexual behaviors were good to excellent, with the number of steady (fair) and casual (poor) partners as the only exceptions. Results of 12-month reliability analyses, although lower, were also very good. For substance use variables, only sedatives (fair) and heroin (fair) were outside of the good to excellent range. Finally, 12-month reliability of sexual health indicators remained quite good, with steady partners (fair), casual partners (poor), and sex under the influence (fair) as the worst performers. At both 6- and 12-month recall assessments, participants overestimated their substance use and underestimated their sexual behavior relative to the initial assessment. Paired-sample *t* tests revealed significant mean increases from initial report occurred at 6 months for cocaine and hard drugs and at 12 months for drinking days, sedatives, hard drugs, all drug use, and drug of choice. These findings may indicate either a revealing of previously undisclosed use or a bias, such as remembering past use as worse than it was.

Discussion

The purpose of the current study was to compare the quality of substance use and HIV risk recall at 6- and 12-month delays. We estimated the correlation and agreement between long-term recall and initial recall. We found that recall at 6- and 12-month delays was of generally very good quality, with slightly better correlations and agreements at the 6-month delay than at the 12-month delay. Notable exceptions to this were recall of heroin and the number of both steady and casual partners, all of which coincided with low base rates in the current sample.

Results from the current study are in line with Napper et al.'s (2010) findings that reliability of 6-month recall of substance use and HIV risk events is generally acceptable. Of note, because the current study uses additional distinctions into types and circumstances of HIV risk, we demonstrated that recall of casual partners may be particularly poor. We extend findings from Napper et al. by reporting on long-term (12-month) recall, finding that, overall, reliability remained acceptable for both HIV risk and substance use variables. A single previous study reported on 12-month recall (Catania et al., 1990), concluding conservatively on a range of studies that 1- or 2-month estimates would be the optimal approach in assessing frequencies of sexual behavior. However, this

TABLE 1. Timeline Follow-back intra-class correlations and agreements at 6- or 12-month reassessments

Variable	6-month				12-month			
	Corr.	[95% CI]	Agree. κ	SE	Corr.	[95% CI]	Agree. κ	SE
Substance use								
Drinking days	.82	[.70, .89]	.68	.11	.78	[.63, .87]	.72	.10
Heavy drinking	.65	[.46, .78]	.42	.12	.72	[.56, .83]	.61	.11
Any alcohol use	.79	[.65, .87]	.68	.11	.77	[.62, .86]	.72	.10
Marijuana	.84	[.74, .91]	.67	.10	.81	[.68, .89]	.77	.10
Cocaine	.92	[.85, .96]	.91	.06	.66	[.47, .79]	.67	.11
Sedatives	.84	[.74, .91]	.37	.16	.55	[.33, .72]	.47	.22
All opiates combined	.83	[.72, .90]	.56	.14	.55	[.33, .72]	.81	.11
Hard drugs combined	.93	[.86, .96]	.80	.08	.62	[.40, .77]	.76	.09
All drugs combined	.90	[.83, .94]	.63	.14	.78	[.64, .87]	.75	.12
Drug of choice	.82	[.71, .90]	.61	.12	.64	[.43, .79]	.71	.10
Heroin	.84	[.74, .91]	.59	.15	.82	[.71, .89]	.45	.17
Opiates	.70	[.52, .82]	.56	.19	.32	[.06, .54]	.73	.15
Other drugs	.96	[.93, .98]	.78	.15	.59	[.38, .74]	.90	.10
HIV risk								
HIV+ partners ^a	*	*	*	*	*	*	*	*
HIV- partners	.77	[.63, .86]	.79	.09	.50	[.26, .68]	.83	.08
HIV-unknown partners	.84	[.73, .90]	.70	.12	.09	[-.20, .35]	.65	.14
Total partners	.74	[.59, .84]	.77	.11	.49	[.25, .68]	.65	.15
Male partners	.74	[.59, .85]	.85	.10	.45	[.21, .65]	.77	.13
Female partners	.43	[.19, .63]	.73	.18	.63	[.43, .77]	1.00	.00
Steady partners	.74	[.58, .84]	.56	.16	.35	[.08, .57]	.56	.17
Casual partners	.71	[.54, .83]	.33	.21	.12	[-.17, .38]	.38	.27
Condomless sex	.67	[.04, .92]	.82	.09	.75	[.08, .96]	.66	.14
Sex under the influence	*	*	.76	.10	*	*	.50	.15

Notes: Mean monthly scores based on 90-day assessment period. Interpretation of agreement scores is recommended for zero rates over 80% (recommended interpretations in bold). Paired-sample *t* tests reveal that significant mean differences occurred at 6 months for cocaine and hard drugs, and at 12 months for drinking days, sedatives, hard drugs, all drug use, and drug of choice. Corr. = correlation; agree. = agreement. ^aAll participants indicated having had no known seropositive partners in the assessed timeframe. Hence, reliability could not be computed.

study predates popularity of the TLFB method as an assessment strategy. In our study, we demonstrate using TLFB methodology that Catania et al.'s conclusion may have been overly conservative.

Participant recall of number of steady and casual partners at 6- and 12-month delays ranged from poor to fair. Potentially, these findings may be explained due to the nature of developing relationships, where certain relationships that were important or seemed lasting to the participant at initial assessment are reflected on as casual after the fact. Conversely, relationships that started on a casual basis may have evolved into more steady ones. In this sense, the lack of agreement would not reflect an inability to recall HIV risk but a redefinition of what the HIV risk event initially constituted.

The current study's strengths were that we used two reassessments, one at 6 months and one at 12 months, based on two different initial time frames. We included TLFB assessments of both substance use and HIV risk, allowing us to investigate whether TLFB methodology was more reliable at allowing recall of one type of event over the other. Finally, all participants completed all baseline and recall assessments. Limitations of the current study included sample size and skew (which may increase variability in reliability, based on bias for some types of substance use and HIV risk outcomes).

In addition, only the TLFB was used to assess recall accuracy. Aftercare differences may have biased some results.

Our results indicate that TLFB may be reliably used to assess substance use and many aspects of HIV risk at both 6- and 12-month delays, especially if these behaviors occur with sufficient frequency. Regarding HIV risk, although not tested in the current study, we feel it is important to interpret HIV risk events with caution, as relationships shift and certain HIV risks are not applicable to all people. Overall, however, there was excellent recall on instances of condomless sex and sex under the influence at the 6-month follow up but fair to good recall on these events at 12 months. The TLFB is a valuable research instrument for capturing HIV risk as well as substance use over extended periods. Accurately recording the trajectory of post-treatment behaviors facilitates mechanisms research and provides insight into time-varying treatment effects.

References

- American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.
- Catania, J. A., Gibson, D. R., Marin, B., Coates, T. J., & Greenblatt, R. M. (1990). Response bias in assessing sexual behaviors relevant to HIV transmission. *Evaluation and Program Planning*, 13, 19-29. doi:10.1016/0149-7189(90)90005-H

- Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6*, 284–290. doi:10.1037/1040-3590.6.4.284
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*, 37–46. doi:10.1177/001316446002000104
- Harrison, L. (1997). The validity of self-reported drug use in survey research: An overview and critique of research methods. *NIDA Research Monograph, 167*, 17–36.
- McLellan, A. T., Lewis, D. C., O'Brien, C. P., & Kleber, H. D. (2000). Drug dependence, a chronic medical illness: Implications for treatment, insurance, and outcomes evaluation. *JAMA, 284*, 1689–1695. doi:10.1001/jama.284.13.1689
- Napper, L. E., Fisher, D. G., Reynolds, G. L., & Johnson, M. E. (2010). HIV risk behavior self-report reliability at different recall periods. *AIDS and Behavior, 14*, 152–161. doi:10.1007/s10461-009-9575-5
- Rogers, R. D., & Robbins, T. W. (2001). Investigating the neurocognitive deficits associated with chronic drug misuse. *Current Opinion in Neurobiology, 11*, 250–257. doi:10.1016/S0959-4388(00)00204-X
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*, 420–428. doi:10.1037/0033-2909.86.2.420
- Sobell, L. C., & Sobell, M. B. (1992). Timeline follow-back. In R. Z. Litten & J. P. Allen (Eds.), *Measuring alcohol consumption* (pp. 41–72). New York, NY: Humana Press.
- Wetherill, R. R., & Fromme, K. (2011). Acute alcohol effects on narrative recall and contextual memory: An examination of fragmentary blackouts. *Addictive Behaviors, 36*, 886–889. doi:10.1016/j.addbeh.2011.03.012
- Wray, T. B., Braciszewski, J. M., Zywiak, W. H., & Stout, R. L. (2016). Examining the reliability of alcohol/drug use and HIV-risk behaviors using timeline follow-back in a pilot sample. *Journal of Substance Use, 21*, 294–297.
- Yoshino, A., & Kato, M. (1995). Influence of social desirability response set on self-report for assessing the outcome of treated alcoholics. *Alcoholism: Clinical and Experimental Research, 19*, 1517–1519. doi:10.1111/j.1530-0277.1995.tb01016.x