



Published in final edited form as:

J Subst Use. 2016 ; 21(3): 294–297. doi:10.3109/14659891.2015.1018974.

Examining the reliability of alcohol/drug use and HIV-risk behaviors using Timeline Follow-Back in a pilot sample

T.B. Wray¹, J.M. Braciszewski², W.H. Zywiak², and R.L. Stout²

¹Center for Alcohol and Addictions Studies, Brown University School of Public Health, Providence, RI

²Decision Sciences Institute, Pacific Institute for Research and Evaluation, Pawtucket, RI

Abstract

Research on the course of substance use disorders (SUDs) faces challenges in assessing behavior over lengthy time periods. Calendar-based methods, like the Timeline Followback (TLFB), may overcome these challenges. This study assessed the reliability of self-reported weekly alcohol use, drug use, and HIV-risk behaviors over the past 90 days using an interview TLFB. Individuals with SUD in outpatient treatment (N = 26) completed the TLFB at baseline and then a week later with separate interviewers. Weekly ratings were aggregated across 4 week intervals for each administration. Intra-class correlations were used to compare agreement between the two administrations. Reliabilities for alcohol and drug use ratings ranged from good to excellent for most drug categories (ICCs = 0.76 – 1.00), except opioid use (other than heroin) and sedative use produced sub-standard reliabilities (ICCs = 0.29 – 0.74). HIV-risk behavior reliabilities also ranged from good to excellent (ICCs = 0.70 – 0.97), but were substandard for the number of casual sex partners for some intervals (ICCs = 0.29, 0.63). Findings extend support for the use of TLFB to produce reliable assessments of many drugs and HIV-risk behaviors across longitudinal intervals.

Keywords

Alcohol; drug use; sexual behavior; HIV risk; measurement; assessment

INTRODUCTION

Longitudinal research on the course of substance-related disorders (SUDs) faces a number of significant challenges in assessing relevant outcomes over years. These challenges include the sensitive nature of SUD-related outcomes, the complicated course of SUDs and intricacies endemic to assessment in those with SUD. Use of alcohol and drugs, for example, is often the central outcome in longitudinal studies, but changes rapidly over time and across treatment episodes (McLellan, Lewis, O'Brien, & Kleber, 2000). Sensitivity about reporting substance use to an interviewer after being in treatment may also prevent consistent reports of drug use behavior (Harrison, 1997). Finally, limitations in memory and communication

skills among those with SUD may present problems in assessing drug use (Rogers & Robbins, 2001).

Similarly, assessing sexual behaviors that may increase HIV risk among those with SUDs is critical, but involves both common and unique challenges. Substantial variability in these behaviors over time and high rates of involvement may be substantial barriers to accurate recall. For example, those with SUD are more likely to report having engaged in sex in exchange for money or drugs (Bobashev, Zule, Osilla, Kline, & Wechsberg, 2009), resulting in high numbers of sex partners and events over short periods of time. HIV-risk behaviors also frequently change along with drug use involvement (e.g., Marsch, 1998), varying substantially over short periods of time. These challenges affirm the importance of developing and testing assessment instruments that are capable of reliably assessing these complex behaviors over long periods of time.

One approach to maximizing accurate recall could involve utilizing measurement methods that assess behavior close to its occurrence (e.g., daily diary). However, over long intervals, such methods are often burdensome to participants (Del Boca & Darkes, 2003). Calendar-based recall methods have been used in order to address some of these challenges, and involve presenting a visual calendar to aid individuals in recalling behaviors. One such instrument, the Timeline Follow Back interview (TLFB; Sobell & Sobell, 1992) has been shown to be reliable across a variety of behaviors, recall windows, and populations (Napper, Fisher, Reynolds, & Johnson, 2010), and may produce similar results to more intensive assessment methods (Carney, Tennen, Affleck, del Boca, & Kranzler, 1998; Wray, Reed, Hunsaker, Finn, & Simons, 2010). In a recent meta-analysis, Napper and colleagues (2010) reported that reliabilities for TLFBs assessing drug and alcohol use, administered face-to-face with test-retest intervals of 1–2 weeks, were generally acceptable at 30-day, 3-month, and 6-month follow-up windows, and with populations as diverse individuals with SUD in-treatment (Ehrman & Robbins, 1994; Fals-Stewart, O'Farrell, Freitas, McFarlin, & Rutigliano, 2000), homeless adults (Sacks, Drake, Williams, Banks, & Herrell, 2003), and psychiatric outpatients (Carey, Carey, Maisto, & Henson, 2004). The meta-analysis also suggested that TLFBs used to measure HIV-risk behaviors were also generally reliable, but only two of the 14 studies included in this portion of the review utilized a TLFB (Carey, Carey, Maisto, Gordon, & Weinhardt, 2001; Weinhardt et al., 1998). Still, at least one other study offers further support to the reliability of TLFB assessments of HIV-risk behavior (Midanik et al., 1998).

Overall, these findings suggest that interview-based TLFBs are a useful tool for longitudinal assessment of drug use and HIV-risk behaviors and are reliable across behaviors assessed and follow-up periods from 30-days to 6-months. However, past studies of TLFB reliabilities have often aggregated substance use outcomes (e.g., “substance use” or “alcohol vs. drug use”). Although there are exceptions (e.g., Ehrman & Robbins, 1994), reports of TLFB reliabilities for many specific substance use categories are rare. Similarly, although few exceptions exist (Carey et al., 2001; Midanik et al., 1998), TLFB reliability studies of HIV-risk variables have generally been limited to the number of sex partners and/or frequency of engaging in specific sex acts (i.e., vaginal, anal, and oral sex).

The present study examines the reliability of the TLFB across test-retest intervals of approximately one week among participants enrolled in a larger longitudinal study on the influence of social network characteristics on SUD outcomes over time. Participants completed the TLFB instrument across test-retest intervals of approximately one week and completed them with different interviewers to ensure consistency. Addressing some prior limitations in TLFB reliability research, we examine a range of specific substances, as well as more detailed questions about participants' sexual activity.

METHOD

Participants

For this sub-study, substance abuse treatment clients were recruited from two different day treatment programs. Participants were included in the study if they: 1) were 18+ years old, 2) met criteria for alcohol abuse or dependence prior to intake (within past year, with use in the past 6 months), 3) had stable residences, and 4) were willing to submit urine and breath samples. Exclusion criteria included: 1) active suicidal or homicidal ideation, 2) being on parole, and 4) having any pending legal charges that could result in incarceration. Individuals with other Axis I disorders were still eligible for the studies.

Procedure

Twenty-six individuals were recruited for participation in the reliability study, which involved completing an assessment battery at baseline ("test") and then the same battery a week later ("retest"). Upon recruitment, participants were first consented for this portion of the study and then completed the "test" administration of the assessment battery. Participants were then contacted by a different interviewer to schedule an appointment to complete the "retest" battery a week later. Both test and retest interviews took an average of 2 hours to complete. Participants were provided with \$25 in gift cards for completing the test battery, and \$40 in gift cards for the retest battery. Each of the two interviewers completed 13 tests, and for the other half of the sample 13 retests. If eligible, participants were invited to participate in the broader longitudinal study after completing the retest assessment. All study procedures were reviewed and approved by relevant agency institutional review boards (IRBs) and ethics committees.

Interviewer training

All assessments were administered via face-to-face interview. To ensure data quality, interviewers were trained using an extensive protocol: Rating mock and videotaped interviews, conducting mock interviews, observing and rating interviews conducted by experienced staff, and conducting their initial few interviews under supervision.

Measures

The *Timeline Follow-Back* (Sobell & Sobell, 1996) was used to assess participants' weekly alcohol use, drug use, and engagement in HIV-risk behaviors over the past 90 days. All 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 5 or more standard drink units for men, or 4 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. Finally, HIV-risk behaviors were assessed by first providing detailed definitions of “sexual activity,” “partner types,” and “HIV status” for participants. Then, interviewers inquired about the *total* number of sex partners they had over the 90-day period, 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, participants 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*). Both the test and retest administrations assessed the same time interval.

Statistical analysis

Individual weeks of the 90-day assessment period were pooled into three, 4-week-long intervals, with values of “4” for a given window (e.g., Weeks 1–4) representing having used a given drug on all four weeks during this period. Intra-class correlations (ICCs) comparing these 4-week intervals at test to those at retest were calculated for weekly-reported alcohol, drug use, and sex behaviors. Percent agreement and Cohen’s *kappa* were calculated for sex partner items reported globally for the whole 90-day period.

RESULTS

Twenty-four participants was recruited for this sub-study, and all participants who completed the baseline assessments also completed the follow-up battery, with an average of 8.9 days (SD = 2.8, mode = 7, range: 6 to 18 days) between the two interviews. Participants were an average of 40.5 years old (SD = 18.6, range 24 to 57), and 61.50% were female. Thirty nine percent of the sample was Latino/a persons, 4% were American Indian/Alaskan Native, 19% Black/African American, and 77% White. Eight percent reported their sexual orientation as bisexual, with the remainder identifying their orientation as heterosexual. Nineteen percent of the sample had been charged with prostitution (for those charged at least once, charges: M = 4.8, SD = 4.6, mode = 1, median = 4, range 1 to 12). Primary “drugs of choice” were cocaine (34.6%), alcohol (26.9%), heroin (26.9%), and other opiates (11.5%).

Drug and alcohol TLFB reliabilities

For the vast majority of drugs assessed in the TLFB, reliability across the three, 4-week periods assessed was acceptable to excellent, with ICCs ranging from 0.70 (sedative use for weeks 9–12, most distal) to 1.00 (heroin use for weeks 1–4, most proximal; See Table 1). However, ratings for opioid use for Weeks 1–4 and Weeks 5–8 were poor (*ICCs* = 0.39, 0.29, respectively). In addition, ratings were markedly lower (though generally acceptable) for sedative use when compared with other drug categories. However, these weak results may be due to both difficulties in rating certain drug categories, as well as the specific characteristics of a sample that is early in recovery.

HIV-risk behavior TLFB reliabilities

Reliabilities were also generally strong across a majority of the HIV-risk behaviors assessed (see Table 1). Global assessment of the number of sex partners across the 90-day period were good, with 84.62% agreement from test to retest, $k = 0.75$, $SE = 0.11$, $p < .001$. Agreement for the number of partners with HIV-negative (80.77%, $k = 0.64$, $SE = 0.17$, $p = .001$) and HIV-status unknown (76.92%, $k = 0.64$, $SE = 0.12$, $p < .001$) were acceptable, and agreement on the number of HIV positive partners was excellent (100%, but all respondents reported “0”).

Although the true range of reliabilities for the weekly HIV-risk behaviors ranged from poor to excellent ($ICCs = 0.29 - 0.97$), most key variables exhibited good to excellent agreement, including the total number of weekly partners, number of male partners, partners who were steady, condom use frequency, and sex under the influence. Among these variables, ICCs ranged from 0.71 to 0.97. However, reliabilities were fairly poor for the number of casual partners, surprisingly in the most proximal weeks of the assessment. Like drug use behaviors, these results may have been due to confusion about what constitutes a “casual” sexual partnership. The reliability of ratings of the number of female partners during the middle 4-week period were also lower than expected ($ICC = 0.56$).

DISCUSSION

This study compared ratings of drug use and sexual behavior reported on an interview-based, 90-day TLFB at baseline and one week later. In general, reliabilities for alcohol and drug use ratings ranged from good to excellent and were comparable or higher when compared to those reported in past studies using similar follow-up windows (Carey et al., 2004; Day, Collins, Degenhardt, Thetford, & Maher, 2004; Fals-Stewart et al., 2000; Sacks et al., 2003). These results suggest that assessment of drug and alcohol use in the past 90-days using the TLFB can be reliably accomplished. Two important caveats to these conclusions are worthy of note: Opioid use (other than heroin) and sedative use. Lower reliabilities were observed for these two categories, and could be due to complexities in their use for recreational versus medical purposes, confusion about certain drugs belonging to these categories, or strong demand among some participants to report no use. However, it is important to note that use was endorsed infrequently for both drug classes (for example, $M_T = 0.04$, $SD_T = 0.20$, and $M_R = 0.06$, $SD_R = 0.32$, for opioid use in weeks 1–4, suggesting that low base rates may also have contributed to low reliability. To our knowledge, this is the first study reporting on reliabilities for these drug categories to date, and as such, future research would be useful for further clarifying these results.

This study is one of only a few to examine the reliability of HIV-related risk behavior using the TLFB. Like drug use ratings, reliabilities for HIV-risk behaviors were also generally good to excellent, again lending further support to past findings assessing these behaviors over similar recall periods (Carey et al., 2001; Weinhardt et al., 1998). One important limitation, however, might be in assessing respondents’ relationships with their sex partners (e.g., “casual” vs. “steady”) in those with SUD, since ratings for the number of “casual” partners within certain recall windows (more proximal to the assessment, in this case) were poor. As such, future research using the TLFB to assess HIV-risk behaviors should be

careful to explicitly define the type of relationships respondents may have with their sex partners. However, it is also possible that poor reliabilities reflect rapidly changing relationships among those in recovery from SUD. Finally, rapport with participants may also play a role, with participants being cautious when reporting occurrences of casual sex during the test phase.

Several limitations are important to note. First, this study involved a relatively small sample within a broader longitudinal project. Second, this study specifically enrolled a sample of those with SUD who were typically at the beginning phases of an episode of treatment. As such, use was a relatively low base rate event across each 4-week period assessed, and results may have been unduly affected by lack of concordance in only a few reports.

In summary, the current study showed that assessing drug use and HIV-risk behaviors in the past 90-days using an interview-based TLFB is feasible and generally reliable. These results extend past findings to new drug categories and sex behaviors, and lend additional support for the use of the TLFB as an effective tool for assessing these behaviors in longitudinal research. Although the TLFB has important limitations, such as the reliance on self-report and the possibility of recall bias, our findings highlight the TLFB as one tool that can produce reliable assessment when biological sampling is impossible or when more intensive methods are not feasible (e.g., in longitudinal research focusing on change over months and years). Overall, the integrity of the TLFB is strong, and these methods are well-suited for longitudinal protocols that often have long assessment periods.

Acknowledgments

Funding support:

This research was supported by NIDA grant R01DA031154.

References

- Bobashev GV, Zule WA, Osilla KC, Kline TL, Wechsberg WM. Transactional sex among men and women in the south at high risk for HIV and other STIs. *Journal of Urban Health*. 2009; 86(1):32–47. [PubMed: 19513853]
- Carey KB, Carey MP, Maisto SA, Henson JM. Temporal stability of the timeline followback interview for alcohol and drug use with psychiatric outpatients. *Journal of studies on alcohol*. 2004; 65(6): 774. [PubMed: 15700516]
- Carey MP, Carey K, Maisto S, Gordon C, Weinhardt L. Assessing sexual risk behaviour with the Timeline Followback (TLFB) approach: continued development and psychometric evaluation with psychiatric outpatients. *International journal of STD & AIDS*. 2001; 12(6):365–375. [PubMed: 11368817]
- Carney MA, Tennen H, Affleck G, del Boca FK, Kranzler HR. Levels and Patterns of Alcohol Consumption Using Timeline Follow-Back, Daily Diaries and Real-Time. *Journal of Studies on Alcohol and Drugs*. 1998; 59(4):447.
- Day C, Collins L, Degenhardt L, Thetford C, Maher L. Reliability of heroin users' reports of drug use behaviour using a 24 month timeline follow-back technique to assess the impact of the Australian heroin shortage. *Addiction Research & Theory*. 2004; 12(5):433–443.
- Del Boca FK, Darkes J. The validity of self-reports of alcohol consumption: state of the science and challenges for research. *Addiction*. 2003; 98(s2):1–12. [PubMed: 14984237]

- Ehrman RN, Robbins SJ. Reliability and validity of 6-month timeline reports of cocaine and heroin use in a methadone population. *Journal of Consulting and Clinical Psychology*. 1994; 62(4):843. [PubMed: 7962889]
- Fals-Stewart W, O'Farrell TJ, Freitas TT, McFarlin SK, Rutigliano P. The timeline followback reports of psychoactive substance use by drug-abusing patients: psychometric properties. *Journal of consulting and clinical psychology*. 2000; 68(1):134. [PubMed: 10710848]
- Harrison L. The validity of self-reported drug use in survey research: an overview and critique of research methods. *NIDA Res Monogr*. 1997; 167:17–36. [PubMed: 9243555]
- Marsch LA. The efficacy of methadone maintenance interventions in reducing illicit opiate use, HIV risk behavior and criminality: a meta-analysis. *Addiction*. 1998; 93(4):515–532. [PubMed: 9684390]
- McLellan AT, Lewis DC, O'Brien CP, Kleber HD. Drug dependence, a chronic medical illness. *JAMA: the journal of the American Medical Association*. 2000; 284(13):1689–1695. [PubMed: 11015800]
- 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 followback technique. *Journal of Studies on Alcohol and Drugs*. 1998; 59(6):681.
- Napper LE, Fisher DG, Reynolds GL, Johnson ME. HIV risk behavior selfreport reliability at different recall periods. *AIDS and Behavior*. 2010; 14(1):152–161. [PubMed: 19475504]
- Rogers RD, Robbins TW. Investigating the neurocognitive deficits associated with chronic drug misuse. *Current opinion in neurobiology*. 2001; 11(2):250–257. [PubMed: 11301247]
- Sacks JAY, Drake RE, Williams VF, Banks SM, Herrell JM. Utility of the time-line follow-back to assess substance use among homeless adults. *The Journal of nervous and mental disease*. 2003; 191(3):145–153. [PubMed: 12637840]
- Sobell, L.; Sobell, M. *Timeline followback user's guide: A calendar method for assessing alcohol and drug use*. Toronto: Addiction Research Foundation; 1996.
- Sobell, LC.; Sobell, MB. *Measuring alcohol consumption*. Springer; 1992. Timeline follow-back; p. 41-72.
- Weinhardt LS, Carey MP, Maisto SA, Carey KB, Cohen MM, Wickramasinghe SM. Reliability of the timeline follow-back sexual behavior interview. *Annals of Behavioral Medicine*. 1998; 20(1):25–30. [PubMed: 9755348]
- Wray, TB.; Reed, RN.; Hunsaker, R.; Finn, JR.; Simons, JS. "How much did you drink on Friday?" Comparisons of three self-report measures of alcohol use with transdermal alcohol assessment. San Diego, CA: Poster presented at the annual convention of the American Psychological Association; 2010.

TABLE 1

Pairwise ICCs of Sex Behavior Ratings at Test and Retest

	Weeks 1–4	Weeks 5–8 Weeks	Weeks 9–12
Total # partners	0.92	0.92	0.86
# Male partners	0.97	0.96	0.93
# Female partners	0.79	0.56	0.70
# Steady partners	0.88	0.87	0.77
# Casual partners	0.29	0.63	0.94
Condom use frequency	0.93	0.94	0.71
Sex under the influence	0.92	0.89	0.81

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript