Psychology of Addictive Behaviors
2014, Vol. 28, No. 1, 154–162

0893-164X/14/$12.00 DOI: 10.1037/a0030992

Reliability of the Timeline Followback for Cocaine, Cannabis, and Cigarette Use

Sean M. Robinson, Linda Carter Sobell, and Mark B. Sobell
Nova Southeastern University

Gloria I. Leo
Centre for Addiction and Mental Health, Toronto, Canada

The Timeline Followback (TLFB), a retrospective calendar-based measure of daily substance use, was initially developed to obtain self-reports of alcohol use. Since its inception it has undergone extensive evaluation across diverse populations and is considered the most psychometrically sound self-report measure of drinking. Although the TLFB has been extended to other behaviors, its psychometric evaluation with other addictive behaviors has not been as extensive as for alcohol use. The present study evaluated the test–retest reliability of the TLFB for cocaine, cannabis, and cigarette use for participants recruited from outpatient alcohol and drug treatment programs and the general community across intervals ranging from 30 to 360 days prior to the interview. The dependent measure for cigarette smokers and cannabis users was daily use of cigarettes and joints, respectively, and for cocaine users it was a "Yes" or "No" regarding cocaine use for each day. The TLFB was administered in different formats for different drug types. Different interviewers conducted the two interviews. The TLFB collected highly reliable information about participants' daily use of cocaine, cannabis, and cigarettes from 30, 90, to 360 days prior to the interview. Findings from this study not only suggest that shorter time intervals (e.g., 90 days) can be used with little loss of accuracy, but also add to the growing literature that the TLFB can be used with confidence to collect psychometrically sound information about substance use (i.e., cocaine, cannabis, cigarettes) other than alcohol in treatment- and nontreatment-seeking populations for intervals from ranging up to 12 months prior to the interview.

Keywords: Timeline Followback, TLFB, test–retest reliability, cocaine use, cannabis use, cigarette smokers

The Timeline Followback (TLFB) was developed as a self-report measure to gather retrospective estimates of daily drinking using a calendar-based format (Sobell, Maisto, Sobell, & Cooper, 1979; Sobell & Sobell, 1992). The TLFB, from the time of its inception in the 1970s, has been shown to be a psychometrically sound instrument for assessing daily alcohol use (Agrawal, Sobell, & Sobell, 2008; Sobell & Sobell, 2008), for providing detailed daily drinking information across a number of variables (e.g., number of days of low- and high-risk drinking, number of days abstinent, days to relapse, mean drinks per drinking day, mean drinks per week, longest continuous abstinent period) for intervals ranging from 1 to 12 months prior to the interview. The TLFB has been found to be a scientifically and clinically useful measure for assessing alcohol use in clinical settings for aiding diagnosis, treatment planning, and selecting goals, as well as for evaluating client progress and treatment outcomes (Agrawal et al., 2008; Sobell, Toneatto, & Sobell, 1994).

Retrospective calendar-based self-report instruments such as the TLFB provide a minimally invasive low-cost psychometrically sound alternative to biological measures (e.g., blood, urine, hair samples) when assessing recent substance use (Babor, Steinberg, Anton, & Del Boca, 2000; Sobell, Agrawal, & Sobell, 1997, 1999). Because the perception of biological screenings as intrusive or adversarial is not uncommon, the use of self-report instruments such as the TLFB may be particularly advantageous in settings that focus on maintaining a strong therapeutic alliance.

Although developed as a paper-and-pencil measure, the TLFB has also been successfully adapted and administered in a variety of other formats including group, telephone, computer, web-based, and online (Sobell, Brown, Leo, & Sobell, 1996; Pedersen, Grow, Duncan, Neighbors, & Larimer, 2012; Pedersen & LaBrie, 2006; Rueger, Trela, Palmeri, & King, 2012).

Since its inception, the use and evaluation of the TLFB has been extended to diverse populations including (a) those with comorbid mental illness (Carey, Carey, Maisto, & Henson, 2004; DeMarce, Burden, Lash, Stephens, & Grambow, 2007), (b) polysubstance users (Staines, Magura, Foote, Deluca, & Kosanke, 2001), (c) adolescent/minority samples (Dillon, Turner, Robbins, & Szapocznik, 2005), (d) homeless individuals (Sacks, Drake, Williams, Banks, & Herrell, 2003), (e) normal drinking college students

This article was published Online First December 31, 2012.

Sean M. Robinson, Linda C. Sobell, and Mark B. Sobell, Center for Psychological Studies, Nova Southeastern University; Gloria I. Leo, Centre for Addiction and Mental Health, Toronto, Canada.

We thank Ingrid Fedoroff, Pat Cleland, Giao Buchan, and Doug Gavin for their early input into the design, conduct, and evaluation of this study.

Correspondence concerning this article should be addressed to Sean M. Robinson, Center for Psychological Studies, Nova Southeastern University, 3301 College Avenue, Fort Lauderdale, FL 33314. E-mail: sr988@nova.edu

(Sobell, Sobell, Klajner, Pavan, & Basian, 1986), (f) arthritic patients (Bradlow & Mowat, 1985), and (g) chronic liver disease patients (DiMartini et al., 2006).

Although initially designed and tested to capture the retrospective reporting of severely dependent adult alcohol abusers, the TLFB has been extended (and in some cases modified) to assess other addictive behaviors (reviewed in Agrawal et al., 2008), including health-related behaviors (e.g., Copersino, Meade, Bigelow, & Brooner, 2010; Panza, Weinstock, Ash, & Pescatello, 2012). It has been evaluated and used with the following addictive behaviors: nicotine (Shiffman, 2009), cannabis (Norberg, Mackenzie, & Copeland, 2011), methamphetamine (Halkitis et al., 2009), opiates (Raistrick et al., 1994), prescription drugs (Sellers, Somer, Sobell, & Sobell, 1990), cocaine (Ehrman & Robbins, 1994), gambling (Weinstock, Whelan, & Meyers, 2004), and sexual behaviors (e.g., HIV risk behavior, condom use; birth control pills; Carey, Carey, Maisto, Gordon, & Weinhardt, 2001; Floyd et al., 2007; Weinhardt et al., 1998).

For other addictive behaviors, psychometric evaluations of the TLFB, particularly test–retest studies, are fewer in number than for alcohol. This may be because the TLFB was developed for measuring alcohol use as well as the fact that there are unique challenges associated with quantifying unregulated substances. Unlike the familiar "standard drink" used in the alcohol field, most other substances (with the exception of illicit nontherapeutic use of prescribed medication) lack standardized dosage units. Further, although alcohol manufacturers are required by law to label the percentage concentration of alcohol in their products, illicit substances may vary greatly in purity and potency (DuPont & Selavka, 2011) and are often characterized by multiple delivery methods (Raistrick et al., 1994). Finally, the complexities inherent in measuring illicit substance use are often compounded when polysubstance use is present.

Table 1 presents a brief overview of different study characteristics (e.g., number of participants, TLFB measures, test–retest interval) for test–retest studies for the three substances that are the focus of this article (i.e., cocaine, cannabis, nicotine). For comparison purposes, the current study is listed under each of the three drug types. Not counting the current study, Table 1 lists six cannabis test–retest studies, five for cocaine, and two for nicotine. In comparison, there are over two dozen published test–retest studies for alcohol (Agrawal et al., 2008). Studies in Table 1 had to use the conventional TLFB, not a modified version (e.g., Day, Collins, Degenhardt, Thetford, & Maher, 2004). Although there are 10 test–retest studies listed in the reference section, one study had cannabis and cocaine samples, and one had samples for all three substances.

Because it would take considerably more space to review the specific reliabilities (i.e., *r*'s and intraclass correlation coefficients [ICCs]) for each of the test–retest studies in Table 1, suffice it to say that collectively the studies reported high to excellent test–retest results, irrespective of the drug type. Although a recent review suggests that TLFB reports of cannabis and other illicit substance use are generally psychometrically sound (Hjorthøj, Hjorthøj, & Nordentoft, 2012), most of the studies for substances other than alcohol involve clinical populations. Further, some of the cannabis and cocaine studies in Table 1 involved participants whose primary problem was not the drug assessed by the study (e.g., Tonigan, Miller, & Brown, 1997; for alcohol users, illicit drug use and nicotine use were assessed by the TLFB).

The current study evaluated the test–retest reliability of the TLFB for retrospective daily reports of cannabis, cigarettes, and cocaine use for a relatively large sample of treatment- and nontreatment-seeking participants for 30, 90, and 360 day time intervals and for participants whose primary drug used was the drug type assessed in the study.

Method

Participants

To obtain a diverse group of respondents in terms of primary drug use, severity of dependence, and treatment-seeking status, participants were recruited from different alcohol and drug treatment programs at the Addiction Research Foundation (Toronto, Canada), and from advertisements posted throughout the community. Respondents were told the study was evaluating different methods of gathering information about a person's use of drugs, cannabis, or cigarettes and that they would be interviewed on two separate occasions. Participants' group designation and assignment was based on their self-report of primary substance used (cocaine, cannabis, or cigarettes). The ad read:

> If you used/smoked [cocaine/ marijuana/cigarettes] on a regular basis, we need your help: The Clinical Research and Treatment Institute, a fully affiliated teaching hospital at the University of Toronto and the Addiction Research Foundation, is conducting a study concerning different ways to gather information about a person's [substance] use, both in the past year as well as lifetime use. You will be interviewed on two occasions, approximately 7 to 14 days apart. Each interview will take about 30 minutes to complete. You will be paid $10 for each interview.

To be eligible for the study, participants had to meet the following criteria: (a) primary drug problem, including nicotine other than alcohol; (b) ≥15 years of age; (c) signed an informed consent; and (d) agreed to be interviewed twice within a 1- to 2-week interval. A total of 292 respondents met the initial interview criteria and were initially interviewed for the study: 154 were identified as primary cigarette smokers, 77 as primary cannabis users, and 60 as primary cocaine users. The majority of respondents returned for their second interview: (a) 100.00% (60/60) of cocaine participants; (b) 83.77% (129/154) of cigarette smokers; (c) 81.81% (63/77) of cannabis users [1 participant was considered an outlier due to extreme discrepancies (>3 *SD*) and excluded from all data analysis because the percentage of days used in the 360 interval at the first interview was 92% or 331 days and for the second interview it was only 33% or 119 days].

The 25 dropouts for the tobacco group did not differ significantly ($p > .05$) from the completers on any variable across the 360-day interval. The 14 cannabis participants who dropped out, however, had significantly lower quantity/frequency indices of use across the 360-day interval, including fewer total number of joints smoked, $t(76) = 2.40$, $p < .001$; lower mean number of joints smoked in any one day, $t(76) = 2.35$, $p = .021$; lower percent of days smoked four or more joints, $t(76) = 2.44$, $p = .02$; and lower mean number of joints smoked per use day, $t(76) = 2.35$, $p = .02$.

Table 1
Brief Overview of Timeline Followback Test-Retest Reliability Studies for Cannabis, Cocaine, and Nicotine Users

	Cannabis				
	Test-retest interval (days)	Assessment intervals (days)	Measures	# of participants	Other (unless noted all participants in treatment and adults)
Study (author, yr)					
Fals-Stewart et al. (2000)	14	30, 90, 360	# days used	113	
Carey et al. (2004)	5	30, 90	# days used	132	Psychiatric patients
Levy et al. (2004)	7	30, 60, 90	# days used, total joints, joints/occasion	93	Adolescents
Dennis et al. (2004)	7	101	# days used, Max # joints/day, joints/day	101	Adolescents
Norberg et al. (2011)	NR	90	# days used, grams used/day	92	Marijuanilla assessed
Tonigan et al. (1997)	NR	90	# days used	108	Different interviewers Time 1/Time 2 for 54 participants
Current study	9.1 (mean)	30, 90, 360	# days used, # joints used per using day, mean days using 1 to 3 joints and ≥4 joints per day	63	Nontreatment and treatment, different interviewers Time 1/Time 2
Cocaine					
Ehrman (1994)	42	30, 180	# days used, $ spent/day	59	
Fals-Stewart (2000)	14	30, 90, 360	# days used	113	
Sacks (2003)	10.5 (mean)	180	# days used	158	Homeless, psychiatric problems
Westerberg et al. (1998)	2–3	90	# days used	14	
Tonigan et al. (1997)	NR	90	# days used	108	Different interviewers Time 1/Time 2 for 54 participants
Current study	10.4 (mean)	30, 90, 360	# days used, longest consecutive days abstinent and using	60	Nontreatment and treatment, different interviewers Time 1/Time 2
Nicotine (Cigarettes)					
Brown (1998)	21, 140	6, 26, and 29 weeks post quit date	# days used, mean cigarettes/day, light and heavy smoking days	35	Major mood disorder
Tonigan et al. (1997)	NR	90	# days used	108	Different interviewers Time 1/Time 2 for 54 participants
Current study	9.7 (mean)	30, 90, 360	# days used, # cigarette smoked per smoking day, mean days smoking 1 to 10 and ≥11 cigarettes	129	Nontreatment and treatment, different interviewers Time 1/Time 2

Note. NR = not reported.

Demographic and substance use history variables are shown in Table 2 for participants classified by their primary drug problem (cocaine, cannabis, cigarettes). Differences between the three groups of participants were not examined as they were selected to be three independent groups. Across the three groups, the mean age and education level was very similar ranging from 29.87 to 32.73 years and 12.01 to 12.36 years, respectively. Three quarters or more of all participants were male, not employed full time, blue collar (Hollingshead, 1958), and not married. Both drug groups reported having a problem with their primary drug for about 8 1/2 years, whereas the smokers reported having a problem with cigarettes for about 16 years. As shown in Table 2, several variables (i.e., 20-item Drug Abuse Screening Test [DAST-20; Skinner, 1982], *DSM* dependence diagnosis [American Psychiatric Association, 1987, 1994], number of drug-related arrests and drug-related hospitalizations) suggest that participants in both drug groups had serious substance-related problems. The fact that smokers reported waiting a mean of 47.30 min upon waking to have their first cigarette, and 54.26% of all smokers reported "always" smoking their first cigarette of the day within 30 min of waking is consistent with a diagnosis of nicotine dependence. As shown in Table 2, for all three drug categories at least half (53% to 72%) of participants were recruited from community ads, not from outpatient alcohol and drug treatment programs.

Procedures

The Addiction Research Foundation/University of Toronto Institutional Review Board approved the study. Participants signed an informed consent at the start of each interview and were paid $10 after the completion of each interview. At the first interview, all participants individually completed demographic and substance use history questions; cocaine and cannabis users completed the DAST, whereas cigarette smokers completed the 10-item version of the Fagerström Tolerance Questionnaire (FTQ; also referred to as the Revised Tolerance Questionnaire; Tate & Schmitz, 1993). All groups completed the version of the TLFB appropriate for their group (e.g., cocaine, cannabis, cigarette smokers). Participants were scheduled for a second interview within 7 to 14 days of their first interview. The mean (*SD*) test–retest interval in days between the two interviews for cocaine, cannabis, and cigarette users was 10.35 (3.98), 9.11 (8.07), and 9.57 (2.93), respectively.

For the second session, the same procedures were followed as for the first session, except that the DAST, Fagerström, demographic, and substance use history questions were not administered.

Consent forms for the second session indicated that the participants would be completing a second TLFB covering the same period as the first, and participants were informed that this was a standard procedure in determining a questionnaire's usefulness. Upon completion of the second interview, participants were informed that the specific objective of the study was to evaluate the test–retest reliability of the questionnaire.

Interviewers had a master's degree or higher. All interviewers received training in the administration of the TLFB and presented the instructions for completing the TLFB in a standardized manner at each interview. There were different interviewers for the first and second interview, and the second interviewer did not have knowledge of the participants' first interview answers.

Table 2
Demographic and Substance Use Variables for Primary Users of Cocaine, Cannabis, and Cigarette Smokers

Variable	Cocaine participants (n = 60)		Cannabis participants (n = 63)		Cigarette participants (n = 129)	
	%	*M* (*SD*)	%	*M* (*SD*)	%	*M* (*SD*)
Age (years) (range)		32.70 (7.65) (16–58)		29.87 (8.59) (16–51)		32.73 (10.45) (16–58)
Education (years)		12.03 (2.26)		12.01 (2.32)		12.36 (2.63)
Male	85.00		85.72		76.74	
Married/common-law	21.67		17.46		19.37	
Blue collar[a]	76.67		78.57[b]		87.29[c]	
Employed fulltime	21.67		19.04		20.93	
DSM drug dependence[d]	96.67		77.77		NA	
Years used primary drug/smoked cigarettes daily		8.83 (7.63)		8.56 (7.57)		16.03 (10.06)
Drug-related arrests		1.60 (3.18)		1.32 (2.55)		NA
Drug-related hospitalizations		5.05 (19.90)		1.10 (2.69)		NA
DAST score[e]		13.67 (3.90)		10.81 (4.92)		NA
Fagerström score[f]		NA		NA		35.08 (7.44)
Minutes upon waking to the first cigarette		NA		NA		47.30 (98.11)
Number of cigarettes smoked per day		NA		NA		22.43 (11.17)
Recruitment source						
Treatment program	28.33		36.51		47.29[g]	
Community poster	71.67		63.49		52.71	

Note. NA = not applicable.
[a] Hollingshead Scale. [b] n = 56. [c] n = 118. [d] *DSM* drug dependence either *DSM-IIIR* or *DSM-IV* as the study changed during the introduction of the *DSM-IV*. [e] DAST = 20-item version Drug Abuse Screening Test (scores range from 1 to 20). [f] Fagerström = 10-item version (scores range from 1 to 50). [g] These smokers were in treatment programs for alcohol or drug problems not tobacco cessation.

Measures

The demographic (e.g., age, education, gender, marital status) and substance abuse history questionnaire (e.g., years problem, consequences, arrests) was administered at the first session only.

DAST-20. The DAST-20 is a psychometrically sound self-administered questionnaire that requires less than 5 min to complete. The 20-item version of the DAST was used, with items requiring dichotomous ("yes/no") answers relating to drug use consequences over the past 12 months (Skinner, 1982). Scores range from 0 to 20, and a score of six or above is considered a positive screen for drug problems. This questionnaire was administered at the first session only.

FTQ. The FTQ is a self-administered psychometrically sound questionnaire designed to measure dependence on nicotine. The items making up the scale are comprised of questions, including where and how much one smokes. Scores are determined by summing the item responses and fall on a single dimension of severity with higher scores representing more severe dependence. The FTQ-10 was used in this study (Tate & Schmitz, 1993) with scores ranging from 1 to 50.

TLFB. The TLFB is a self-report method for assessing alcohol use, cigarette use, and illicit substance use (Agrawal et al., 2008; Sobell & Sobell, 1992, 2008). Using a calendar format, the TLFB obtains retrospective estimates of daily drug use over a specified interval prior to the interview date. Memory aids can be used to enhance recall (Sobell & Sobell, 1992). As reviewed earlier, the TLFB is a psychometrically sound measure of a variety of addictive behaviors.

In this study, all participants completed TLFB interviews spanning 360 days for both the first and second interviews. It typically takes respondents about 10 min to complete a 90-day interval and 20 to 30 min to complete a 360-day interval. Similar to the alcohol TLFB, which uses daily estimates of drinking based on a metric of standard drinks, the cigarette and cannabis versions of the TLFB allowed for respondents to record daily use on a continuous scale (i.e., number of cigarettes smoked or number of joints smoked, respectively) whereas the cocaine TLFB used a dichotomous rating ("yes" or "no" on a given day). Dichotomous ratings were used because the concentration and dosage of cocaine is quite variable and often unknown due to it being cut with other substances (e.g., baking soda, arsenic). Although assessing the potency of cannabis can be difficult, the tendency of those who use marijuana is to smoke a certain number of "joints" or "blunts" per day, which allows a rough estimate of the relative amount used versus a "yes" or "no" on a given day. The TLFB for cocaine was administered in a paper-and-pencil version, and for cannabis it was by computer (Sobell et al., 1996). The TLFB for cigarette smokers was administered using a computer and paper-and-pencil format, with the administration format counterbalanced over interviews (order effects were not significant, $p > .05$).

Data Analysis

Descriptive and inferential statistics were performed using SPSS version 19. For comparison of mean scores on continuous data, independent sample *t* tests were used, and for categorical variables, chi-squared analyses were used.

Self-reports of substance use. The TLFB variables were calculated for time intervals spanning 30, 90, and 360 days prior to the interview day for each group. The days were calculated from the day before the interview (i.e., 30 days were Days 1 to 30 from the date of the interview; 90 days were Days 1 to 90 from the interview). Several variables were computed from the available data including: (a) frequency of use (e.g., percent days abstinent, longest number of consecutive days abstinent, longest number of consecutive days used); (b) quantity of use (e.g., mean number of cigarettes/joints smoked, total number of cigarettes/joints smoked, greatest number of cigarettes/joints smoked in one day); and (c) quantity/frequency of use (e.g., percent of days smoked one to three joints, percent of days smoked ≥ four joints, percent of days smoking ≥11 cigarettes).

Test–retest reliability. Two measures of reliability were used: (a) ICCs and (b) Pearson correlations (*r*), both unadjusted and adjusted (i.e., zero pairs deleted). *ICCs* are defined as the proportion of total variance of an observation that is associated with the class to which it belongs (Winer, 1971), and thus represent an unbiased estimate and are more sensitive to changes in

Table 3
*Mean (*SD*) Test-Retest Scores and Reliability Indices for Cocaine Users (*N = 60*) Across Three Time Intervals Prior to the Date of the First Interview*

Variable	Interval(days)	Session		Reliability indices		
		Session 1 *M* (*SD*)	Session 2 *M* (*SD*)	r[a]	ICC	Adjusted *r* (n)[a,b]
Percent days abstinent	30	71.86 (27.62)	70.00 (27.50)	.75	.75	.68 (58)
	90	66.91 (26.80)	65.67 (27.49)	.90	.90	.88 (58)
	360	60.94 (27.03)	60.48 (26.58)	.91	.91	.89 (58)
Longest consecutive days abstinent	30	13.35 (9.42)	13.15 (9.90)	.75	.74	.73 (58)
	90	24.48 (22.66)	24.75 (25.04)	.91	.90	.91 (58)
	360	48.33 (41.80)	43.38 (40.64)	.82	.81	.81 (58)
Longest consecutive days using	30	4.42 (6.47)	4.37 (5.90)	.65[c]	.63[c]	.62 (52)[c]
	90	13.80 (23.39)	10.98 (19.74)	.78	.76	.78 (57)
	360	57.63 (104.70)	47.40 (97.28)	.86	.85	[d]

Note. ICC = intraclass correlation coefficients.
[a] All *p* values <. 0001. [b] Correlations are based on all nonzero pairs (i.e., zero pairs deleted); numbers of cases upon which the correlation is based are in parentheses. [c] Studentized Deleted Residual detected three outliers for the variable mean longest consecutive days using in 30 days. When these outliers were removed all values for this variable increased and were high: unadjusted $r = .89$ ($n = 57$), ICC = .89 (n = 57), adjusted $r = .88$ (n = 49). [d] No zero pairs in data set.

Table 4
Mean (SD) Test-Retest Scores and Reliability Indices for Cannabis Users (N = 63) Across Three Time Intervals Prior to the Date of the First Interview

Variable	Interval (days)	Session		Reliability indices		
		Session 1 *M* (*SD*)	Session 2 *M* (*SD*)	r[a]	ICC	Adjusted *r* (*n*)[a,b]
Percent days abstinent	30	37.62 (35.48)	37.09 (37.01)	.93	.92	.87 (44)
	90	31.18 (30.59)	30.48 (31.47)	.89	.89	.83 (46)
	360	27.32 (29.44)	26.18 (29.97)	.96	.96	.96 (55)
Longest consecutive days abstinent	30	6.48 (7.74)	6.67 (8.40)	.82	.81	.75 (44)
	90	10.60 (13.57)	10.32 (14.94)	.81	.80	.76 (46)
	360	24.44 (30.61)	16.37 (22.70)	.87	.81	.86 (55)
Longest consecutive days using	30	14.79 (12.23)	15.63 (12.13)	.94	.94	[c]
	90	43.97 (37.50)	44.81 (36.68)	.93	.92	[c]
	360	151.64 (134.93)	153.29 (143.41)	.91	.91	[c]
Greatest number of joints on any day	30	7.18 (6.07)	6.68 (5.51)	.71[d]	.69[d]	[c]
	90	8.78 (6.81)	8.44 (7.21)	.84	.83	[c]
	360	12.02 (9.75)	11.05 (8.40)	.93	.91	[c]
Total number of joints used	30	81.22 (75.86)	83.41 (84.32)	.83	.82	[c]
	90	259.48 (218.52)	275.90 (242.80)	.79	.78	[c]
	360	1199.97 (1012.82)	1198.48 (943.98)	.95	.94	[c]
Number of joints used per using day	30	3.70 (2.35)	3.56 (2.34)	.70[d]	.69[d]	[c]
	90	3.89 (2.41)	4.07 (2.53)	.80	.79	[c]
	360	4.26 (2.71)	4.26 (2.57)	.94	.93	[c]
Percent days using 1 to 3 joints	30	33.07 (30.59)	35.61 (32.36)	.82	.81	.78 (56)
	90	39.75 (31.55)	39.19 (30.92)	.79	.79	.77 (59)
	360	40.40 (30.69)	40.04 (31.43)	.85	.85	.84 (61)
Percent days using ≥4 joints	30	29.31 (32.50)	27.30 (31.37)	.86	.86	.81 (47)
	90	29.07 (31.21)	30.34 (31.61)	.82	.82	.80 (55)
	360	32.28(31.50)	33.77 (32.25)	.87	.86	.86 (60)

Note. ICC = intraclass correlation coefficients.
[a] All p values $<.0001$. [b] Correlations are based on all nonzero pairs (i.e., zero pairs deleted); numbers of cases upon which the correlation is based is in parentheses. [c] No zero pairs in data set. [d] Studentized deleted residual detected outliers for two variables (mean greatest number of joint on any day in the 30-day interval; mean number of joints used per using day in the 30-day interval). When the one and two outliers were removed, the subsequent unadjusted *r*'s and ICCs increased and were high for both variables (mean greatest number of joint on any day [$n = 62$]: unadjusted $r = .81$, ICC = .80; and mean number of joints used per using day [$n = 61$]: unadjusted $r = .85$, ICC = .85).

test–retest means, and correct for test–retest agreement expected on the basis of chance alone (Wastell & Barker, 1988). Use of ICCs are preferred to Pearson correlations because they are unbiased and are sensitive to changes in test–retest means (Maisto, McKay, & Connors, 1990).

Although Pearson product-moment correlations express the degree to which paired measures are systematically related within their distributions, they are not able to account for absolute differences (Westerberg et al., 1998). Despite this limitation, Pearson correlations were included for comparison with past evaluations of the TLFB. Finally, as has been done in past evaluations of the TLFB (e.g., Sobell, Sobell, Leo, & Cancilla, 1988) to reduce the potential for artificial inflation of correlation coefficients resulting from a high prevalence of zero data pairs across time intervals (e.g., abstinent at both time points), adjusted correlation coefficients were also calculated with zero pairs removed from the dataset.

Results

Table 3 presents the mean (*SD*) test–retest scores for both interviews and reliability indices for 60 primary cocaine users over three time intervals (30, 90, and 360 days prior to the date of the first interview) for three cocaine use variables derived from the TLFB data. Student deleted residuals (Kutner, Nachtsheim, & Neter, 2004) detected outliers for one variable (mean longest consecutive days using cocaine) in the 30-day interval for three participants. As shown in the footnotes to Table 3, when data for the three participants with extreme scores were removed, the unadjusted *r* and ICC increased and were high. With this one exception, the remaining unadjusted correlations (.75 to .91) and ICCs (.75 to .91) were high and in the excellent range (i.e., .75–1.00; Cicchetti, 1994). As discussed above, to control for a high prevalence of abstinent days, zero-pairs were excluded from the analyses, and with two exceptions (i.e., mean percent days abstinent in 30 days, $r = .68$; mean longest consecutive days using in 30 days, $r = .62$), the adjusted reliabilities were similarly high and in the excellent range (.73 to .91). All adjusted and unadjusted correlations were statistically significant ($p \leq .0001$).

Table 4 shows the mean (*SD*) test–retest scores for both interviews and reliability indices for 63 primary cannabis users over three time intervals (30, 90, and 360 days prior to the date of the first interview) for several cannabis use variables derived from the TLFB data. Student deleted residuals (Kutner et al., 2004) detected outliers for two variables (mean greatest number of joints on any day in the 30-day interval; mean number of joints used per using day). As shown in the footnotes to Table 4, when data for these few participants with extreme scores were removed, the unadjusted *r*'s and ICCs increased and were high. Excluding these two ex-

Table 5
Mean (SD) Test-Retest Scores and Reliability Indices for Cigarette Smokers (N = 129) Across Three Time Intervals Prior to the Date of the First Interview

Variable	Interval (days)	Session		Reliability indices		
		Session 1 *M* (*SD*)	Session 2 *M* (*SD*)	*r* [a]	ICC	Adjusted *r* (*n*)[b]
Percent days abstinent	30	1.99 (7.77)	1.71 (6.81)	.76	.76	.65 (21)
	90	2.45 (8.57	2.47 (8.63)	.88	.88	.84 (31)
	360	2.72 (8.39)	2.08 (6.98)	.84	.82	.80 (49)
Longest consecutive days abstinent	30	0.50 (2.15)	0.41 (1.85)	.76	.75	.68 (21)
	90	1.26 (4.70)	1.40 (5.72)	.83	.81	.78 (31)
	360	4.57 (16.18)	3.07 (11.70)	.87	.82	.86 (49)
Greatest number of cigarettes on any day	30	29.03 (13.19)	27.78 (12.68)	.75	.75	[c]
	90	31.70 (15.70)	30.2 (14.09)	.81	.80	[c]
	360	34.49 (16.13)	33.98 (14.66)	.80	.79	[c]
Total number of cigarettes smoked per interval	30	657.81 (329.66)	657.47 (321.67)	.92	.92	[c]
	90	2006.13 (1001.46)	1988.99 (953.18)	.95	.95	[c]
	360	8007.16 (4152.24)	8004.3 (3893.61)	.94	.93	[c]
Number of cigarettes smoked per smoking day	30	22.19 (10.80)	22.18 (10.60)	.92	.92	[c]
	90	22.64 (10.88)	22.48 (10.40)	.95	.94	[c]
	360	22.63 (11.29)	22.55 (10.67)	.94	.93	[c]
Percent days 1 to 10 cigarettes smoked	30	18.61 (32.74)	18.76 (34.23)	.85	.85	.75 (58)
	90	17.03 (30.87)	16.19 (31.55)	.94	.94	.93 (69)
	360	17.46 (30.74)	16.57 (31.60)	.95	.95	.94 (82)
Days smoked ≥11 cigarettes per day	30	79.41 (34.32)	79.53 (35.69)	.85	.84	.73 (119)
	90	80.52 (32.95)	81.34 (33.71)	.95	.94	.89 (119)
	360	79.81 (32.74)	81.35 (33.11)	.96	.96	.95 (125)

Note. ICC = intraclass correlation coefficients.
[a] All p values $<$. 0001. [b] Restricted to non-zero pairs; numbers of cases upon which the correlation is based is in parentheses. [c] No zero pairs in data set.

ceptions, the remaining unadjusted correlations and ICCs were high and in the excellent range (.79 to .96), and (.78 to .96), respectively. Again, to control for a high prevalence of abstinent days, when zero-pairs were excluded the adjusted reliabilities with the same two exceptions as for unadjusted reliabilities were also high and in the excellent range (.75 to .96). All adjusted and unadjusted correlations were statistically significant ($p \leq .0001$).

Table 5 shows the mean (*SD*) test–retest scores for both interviews and reliability indices for 129 primary cigarette smokers over three time intervals (30, 90, and 360 days prior to the date of the first interview) for several variables related to use of cigarettes as derived from the TLFB data. All Pearson unadjusted correlations and ICCs were high and in the excellent range: .75 to .96 and .75 to .96, respectively. With two exceptions (30 days for mean percent days abstinent, $r =$. 65, and mean longest consecutive days abstinent, $r = .68$), the adjusted test–retest correlations (.73–.93) for cigarette smokers were high and in the excellent range. All adjusted and unadjusted correlations were statistically significant ($p \leq .0001$).

Discussion

The results of the present study demonstrate that self-reports of cigarette smokers, and cocaine and cannabis users obtained using the TLFB have high test–retest reliability for intervals ranging from 30 to 90 to 360 days prior to the interview date. Because the present study included clinical (in treatment) and nonclinical participants with varying degrees of dependence, the generalizability of the TLFB to a broader population of drug abusers and smokers can be expected.

The present findings are consistent with studies that evaluated the TLFB with alcohol users (Sobell et al., 1996; Sobell & Sobell, 1992), and for studies that involved cigarette smokers and illicit substance users (Ehrman & Robbins, 1994; Fals-Stewart, O'Farrell, Freitas, McFarlin, & Rutigliano, 2000; Hersh et al., 1999; Hjorthøj et al., 2012). In summary, the present study adds to the growing literature supporting the TLFB as a psychometrically sound measure across illicit and licit substance users in both clinical and nonclinical populations.

Because the TLFB data had good temporal stability over all three intervals the use of a shorter target interval (e.g., 30 to 90 days prior to the interview) should be adequate for most purposes for assessing cocaine, cannabis, and cigarette use. These results are consistent with a study with alcohol abusers that examined different time windows for estimating annual drinking behavior and found that a 1- to 3-month time window is recommended depending on the purpose of data collection (Vakili, Sobell, Sobell, Simco, & Agrawal, 2008). The results of the present study further suggest that shorter time windows, which are more time and resource efficient, can be used with little to no loss in the reliability of the data.

Although participants' reports were gathered using procedures known to enhance the accuracy of reports (e.g., informing participants of confidentiality, using clinically trained interviewers; Maisto et al., 1990), this study did not include an evaluation of the validity of reported use. However, Hjorthøj, Hjorthøj, and Nordentoft (2012) recently concluded that "use of TLFB for detection of illicit substances in populations with substance use disorders appears to give highly valid estimates of substance use both in trials and in prospective studies" (p. 231). It should be noted,

however, that exceptions to valid self-reporting exist (e.g., persons coerced into treatment, those who expect negative consequences for reports of substance use; Cook, Bernstein, & Andrews, 1997).

In summary, this study had a number of advantages over previous TLFB studies: large participant samples, participants with three primary substance use problems, interviewers were blinded to the participants' first or second interview data, use of the metric of number of "joints" per day for cannabis users, multiple temporal intervals were evaluated up to 360 days from the interview, more precise variables derived for cigarette and cannabis users (e.g., mean days using one to four joints and four or more joints per day; mean days smoking 1 to 10 cigarettes and 11 or more per day), and the findings that a shorter time frame can be used to collect reliable TLFB substance use data.

References

Agrawal, S., Sobell, M. B., & Sobell, L. C. (2008). The Timeline Followback: A scientifically and clinically useful tool for assessing substance use. In R. F. Belli, F. P. Stafford & D. F. Alwin (Eds.), *Calendar and time diary methods in life course research* (pp. 57–68). Beverly Hills, CA: Sage. doi:10.4135/9781412990295.d8

American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders* (3rd ed. Revised). Washington, DC: Author.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

Babor, T. F., Steinberg, K., Anton, R., & Del Boca, F. (2000). Talk is cheap: Measuring drinking outcomes in clinical trials. *Journal of Studies on Alcohol, 61,* 55–63.

Bradlow, A., & Mowat, A. G. (1985). Alcohol consumption in arthritic patients: Clinical and laboratory studies. *Annals of the Rheumatic Diseases, 44,* 163–168. doi:10.1136/ard.44.3.163

Brown, R. A., Burgess, E. S., Sales, S. D., Whiteley, J. A., Evans, D. M., & Miller, I. W. (1998). Reliability and validity of a smoking timeline follow-back interview. *Psychology of Addictive Behaviors, 12*(2), 101–112. doi:10.1037/0893-164X.12.2.101

Carey, K. B., Carey, M. P., Maisto, S. A., & Henson, J. M. (2004). Temporal stability of the Timeline Followback interview for alcohol and drug use with psychiatric outpatients. *Journal of Studies on Alcohol, 65,* 774–781.

Carey, M. P., Carey, K. B., Maisto, S. A., Gordon, C. M., & Weinhardt, S. L. (2001). Assessing sexual risk behaviour with the Timeline Followback (TLFB) approach: Continued development and psychometric evaluation with psychiatric outpatients. *International Journal of STD & AIDS, 12,* 365–375. doi:10.1258/0956462011923309

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.

Cook, R. F., Bernstein, A. D., & Andrews, C. M. (1997). Assessing drug use in the workplace: A comparison of self-report, urinalysis, and hair analysis. In L. Harrison & A. Hughes (Eds.), *The validity of self-reported drug use: Improving the accuracy of survey estimates (NIDA Research Monograph* 167), (pp. 247–272). Washington, DC: NIDA. doi:10.3109/10826089509048734

Copersino, M. L., Meade, C. S., Bigelow, G. E., & Brooner, R. K. (2010). Measurement of self-reported HIV risk behaviors in injection drug users: Comparison of standard versus timeline follow-back administration procedures. *Journal of Substance Abuse Treatment, 38,* 60–65. doi:10.1016/j.jsat.2009.06.004

Day, C., Collins, L., Degenhardt, L., Thetford, C., & Maher, L. (2004). 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, 12,* 433–443. doi:10.1080/16066350410001713231

DeMarce, J. M., Burden, J. L., Lash, S. J., Stephens, R. S., & Grambow, S. C. (2007). Convergent validity of the Timeline Followback for persons with comorbid psychiatric disorders engaged in residential substance use treatment. *Addictive Behaviors, 32,* 1582–1592.

Dennis, M. L., Funk, R., Godley, S. H., Godley, M. D., & Waldron, H. (2004). Cross-validation of the alcohol and cannabis use measures in the global appraisal of individual needs (gain) and Timeline Followback (TLFB; Form 90) among adolescents in substance abuse treatment. *Addiction, 99,* 120–128. doi:10.1111/j.1360-0443.2004.00859.x

Dillon, F. R., Turner, C. W., Robbins, M. S., & Szapocznik, J. (2005). Concordance among biological, interview, and self-report measures of drug use among African-American and Hispanic adolescents referred for drug abuse treatment. *Psychology of Addictive Behaviors, 19,* 404–413. doi:10.1037/0893-164X.19.4.404

DiMartini, A., Day, N., Dew, M. A., Javed, L., Fitzgerald, M. G., Jain, A., & Fontes, P. (2006). Alcohol consumption patterns and predictors of use following liver transplantation for alcoholic liver disease. *Liver transplantation: Official publication of the American Association for the Study of Liver Diseases and the International Liver Transplantation Society, 12,* 813–820. doi:10.1002/lt.20688

DuPont, R. L., & Selavka, C. M. (2011). *Testing to identify recent drug use*. Arlington, VA: American Psychiatric Publishing, Inc.

Ehrman, R. N., & Robbins, S. J. (1994). Reliability and validity of 6-month timeline reports of cocaine and heroin use in a methadone population. *Journal of Consulting and Clinical Psychology, 62,* 843–850. doi:10.1037/0022-006X.62.4.843

Fals-Stewart, W., O'Farrell, T. J., Freitas, T. T., McFarlin, S. K., & Rutigliano, P. (2000). The Timeline Followback reports of psychoactive substance use by drug-abusing patients: Psychometric properties. *Journal of Consulting and Clinical Psychology, 68,* 134–144. doi:10.1037/0022-006X.68.1.134

Floyd, R. L., Sobell, M., Velasquez, M. M., Ingersoll, K., Nettleman, M., Sobell, L., . . . Project CHOICES Efficacy Study Group. (2007). Preventing alcohol-exposed pregnancies: A randomized controlled trial. *American Journal of Preventive Medicine, 32,* 1–10. doi:10.1016/j.amepre.2006.08.028

Halkitis, P. N., Solomon, T. M., Moeller, R. W., Doig, S. A., Espinosa, L. S., Siconolfi, D., & Homer, B. D. (2009). Methamphetamine use among gay, bisexual and non-identified men-who-have-sex-with-men: An analysis of daily patterns. *Journal of Health Psychology, 14,* 222–231. doi:10.1177/1359105308100206

Hersh, D., Mulgrew, C. L., VanKirk, J., & Kranzler, H. R. (1999). The validity of self-reported cocaine use in two groups of cocaine abusers. *Journal of Consulting and Clinical Psychology, 67,* 37–42. doi:10.1037/0022-006X.67.1.37

Hjorthøj, C. R., Hjorthøj, A. R., & Nordentoft, M. (2012). Validity of Timeline Follow-Back for self-reported use of cannabis and other illicit substances—systematic review and meta-analysis. *Addictive Behaviors, 37,* 225–233. doi:10.1016/j.addbeh.2011.11.025

Hollingshead, A. B., & Redlich, F. C. (1958). *Social class and mental illness: A community study*. New York, NY: Wiley. doi:10.1037/10645-000

Kutner, M. H., Nachtsheim, C., & Neter, J. (2004). *Applied linear regression models* (4th ed.). New York, NY: McGraw-Hill/Irwin.

Levy, S., Sherritt, L., Harris, S., Gates, E., Holder, D., Kulig, J., & Knight, J. R. (2004). Test-retest reliability of adolescent' self-report of substance use. *Alcoholism: Clinical and Experimental Research, 28*(8), 1236–1241. doi:10.1097/01.ALC.0000134216.22162.A5

Maisto, S. A., McKay, J. R., & Connors, G. J. (1990). Self-report issues in substance abuse: State of the art and future directions. *Behavioral Assessment, 12,* 117–134.

Norberg, M. M., Mackenzie, J., & Copeland, J. (2011). Quantifying cannabis use with the Timeline Followback approach: A psychometric evaluation. *Drug and AlcoholDependence*. doi:10.1016/j.drugalcdep.2011.09.007

Panza, G. A., Weinstock, J., Ash, G. I., & Pescatello, L. S. (2012). Psychometric evaluation of the timeline followback for exercise among college students. *Psychology of Sport and Exercise*. doi:10.1016/j.psychsport.2012.06.002

Pedersen, E. R., Grow, J., Duncan, S., Neighbors, C., & Larimer, M. E. (2012). Concurrent validity of an online version of the Timeline Followback assessment. *Psychology of Addictive Behaviors, 26,* 672–677. doi:10.1037/a0027945

Pedersen, E. R., & LaBrie, J. W. (2006). A within-subjects validation of a group-administered Timeline Followback for alcohol use. *Journal of Studies on Alcohol, 67*(2), 332–335.

Raistrick, D., Braddshaw, J., Tober, G., Weiner, J., Allison, J., & Healey, C. (1994). Development of the Leeds Dependence Questionnaire (LDQ): A questionnaire to measure alcohol and opiate dependence in the context of a treatment evaluation package. *Addiction, 89,* 563–572. doi:10.1111/j.1360-0443.1994.tb03332.x

Rueger, S. Y., Trela, C. J., Palmeri, M., & King, A. C. (2012). Self-administered web-based timeline followback procedure for drinking and smoking behaviors in young adults. *Journal of Studies on Alcohol and Drugs, 73*(5), 829–833.

Sacks, J. A. Y., Drake, R. E., Williams, V. F., Banks, S. M., & Herrell, J. M. (2003). Utility of the Time-Line Follow-Back to assess substance use among homeless adults. *Journal of Nervous and Mental Disorders, 191,* 145–153. doi:10.1097/01.NMD.0000054930.03048.64

Sellers, E. M., Somer, G. R., Sobell, L., & Sobell, M. (1990). Discontinuation symptoms among persistent alprazolam users. *Clinical Pharmacology and Therapeutics, 47,* 168.

Shiffman, S. (2009). How many cigarettes did you smoke? Assessing cigarette consumption by global report, Timeline Followback, and ecological momentary assessment. *Health Psychology, 28,* 519–526. doi:10.1037/a0015197

Skinner, H. A. (1982). The Drug Abuse Screening Test. *Addictive Behaviors, 7,* 363–371. doi:10.1016/0306-4603(82)90005-3

Sobell, L. C., Agrawal, S., & Sobell, M. B. (1997). Factors affecting agreement between alcohol abusers' and their collaterals' reports. *Journal of Studies on Alcohol, 58,* 405–413.

Sobell, L. C., Agrawal, S., & Sobell, M. B. (1999). Utility of liver function tests for screening "alcohol abusers" who are not severely dependent on alcohol. *Substance Use & Misuse, 34,* 1723–1732. doi:10.3109/10826089909039424

Sobell, L. C., Brown, J., Leo, G. I., & Sobell, M. B. (1996). The reliability of the Alcohol Timeline Followback when administered by telephone and by computer. *Drug and Alcohol Dependence, 42,* 49–54. doi:10.1016/0376-8716(96)01263-X

Sobell, L. C., Maisto, S. A., Sobell, M. B., & Cooper, A. M. (1979). Reliability of alcohol abusers' self-reports of drinking behavior. *Behaviour Research and Therapy, 17,* 157–160. doi:10.1016/0005-7967(79)90025-1

Sobell, L. C., & Sobell, M. B. (1992). Timeline Follow-back: A technique for assessing self-reported alcohol consumption. In R. Z. Litten & J. P. Allen (Eds.), *Measuring alcohol consumption: Psychosocial and biological methods* (pp. 41–72). Towota, NJ: Humana Press. doi:10.1007/978-1-4612-0357-5_3

Sobell, L. C., & Sobell, M. B. (2008). Alcohol Timeline Followback (TLFB). In J. Rush, M. B. First, & D. Blacker (Eds.), *Handbook of psychiatric measures* (2nd ed., pp. 466–468). Washington, DC: American Psychiatric Association.

Sobell, L. C., Sobell, M. B., Leo, G. I., & Cancilla, A. (1988). Reliability of a timeline method: Assessing normal drinkers' reports of recent drinking and a comparative evaluation across several populations. *British Journal of Addiction, 83,* 393–402. doi:10.1111/j.1360-0443.1988.tb00485.x

Sobell, L. C., Toneatto, T., & Sobell, M. B. (1994). Behavioral assessment and treatment planning for alcohol, tobacco, and other drug problems: Current status with an emphasis on clinical applications. *Behavior Therapy, 25,* 533–580. doi:10.1016/S0005-7894(05)80199-6

Sobell, M. B., Sobell, L. C., Klajner, F., Pavan, D., & Basian, E. (1986). The reliability of a timeline method for assessing normal drinker college students' recent drinking history: Utility for alcohol research. *Addictive Behaviors, 11,* 149–161. doi:10.1016/0306-4603(86)90040-7

Staines, G. L., Magura, S., Foote, J., Deluca, A., & Kosanke, N. (2001). Polysubstance use among alcoholics. *Journal of Addictive Diseases, 20*(4), 53–69. doi:10.1300/J069v20n04_06

Tate, J. C., & Schmitz, J. M. (1993). A proposed revision of the Fagerström Tolerance Questionnaire. *Addictive Behaviors, 18,* 135–143.

Tonigan, J. S., Miller, W. R., & Brown, J. M. (1997). The reliability of Form 90: An instrument for assessing alcohol treatment outcome. *Journal of Studies on Alcohol, 58,* 358–364.

Vakili, S., Sobell, L. C., Sobell, M. B., Simco, E. R., & Agrawal, S. (2008). Using the Timeline Followback to determine time windows representative of annual alcohol consumption with problem drinkers. *Addictive Behaviors, 33,* 1123–1130. doi:10.1016/j.addbeh.2008.03.009

Wastell, D. G., & Barker, G. R. (1988). Intraclass correlations: A two-facet case study and some comments on the concept of reliability. *Bulletin of the Psychonomic Society, 26,* 583–586.

Weinhardt, L. S., Carey, M. P., Maisto, S. A., Carey, K. B., Cohen, M. M., & Wickramasinghe, S. M. (1998). Reliability of the timeline followback sexual behavior interview. *Annals of Behavioral Medicine, 20*(1), 25–30. doi:10.1007/BF02893805

Weinstock, J., Whelan, J. P., & Meyers, A. W. (2004). Behavioral assessment of gambling: An application of the Timeline Followback method. *Psychological Assessment, 16*(1), 72–80. doi:10.1037/1040-3590.16.1.72

Westerberg, V. S., Tonigan, J. S., & Miller, W. R. (1998). Reliability of Form 90D: An instrument for quantifying drug use. *Substance Abuse, 19,* 179–189. doi:10.1080/08897079809511386

Winer, B. J. (1971). *Statistical principles in experimental design*. New York, NY: McGraw Hill.

Received September 3, 2012
Revision received October 18, 2012
Accepted October 19, 2012 ■