An Experimental Study of the Agreement of Self-Administration and Telephone Administration of the Timeline Followback Interview*

STEPHEN A. MAISTO, PH.D.[†], JOSEPH C. CONIGLIARO, M.D., M.P.H[†], ADAM J. GORDON, M.D., M.P.H.[†], KATHLEEN A. McGINNIS, M.S.[†], and **AMY C. JUSTICE, M.D., PH.D.[†]**
Department of Psychology, Center for Health and Behavior, Syracuse University, 430 Huntington Hall, Syracuse, New York 13244

Abstract

Objective—The Timeline Followback (TLFB) interview has become state-of-the-science for the collection of retrospective self-reports of daily alcohol consumption. Such data are especially useful for addressing questions of the co-occurrence of quantity of alcohol consumption and other behaviors, such as HIV-related risky sex, on the event level. The purpose of this study was to determine if the TLFB could be used effectively by self-administration compared with the more costly telephone interview in a large, multisite observational study of HIV-positive and HIV-negative adults.

Method—An experimental design was used to compare self-administered and telephone-administered TLFB modes in a subsample ($N = 70$) of the Veterans Aging Cohort Study, an ongoing longitudinal study of more than 6,000 HIV-positive and HIV-negative men and women presenting for treatment at eight Department of Veterans Affairs Infectious Disease or General Medicine clinics. Participants were randomly assigned to one of four experimental groups defined by mode and sequence of a TLFB administration on two occasions occurring within 1 week: telephone-telephone, telephone-self, self-telephone, and self-self.

Results—Analyses showed no differences in median total number of drinks reported between modes of TLFB administration or sequence of mode of administration. The same findings held for classification of participants as "hazardous" drinkers. Additional analyses showed good-to-excellent test-retest reliability of self-reports for both modes of TLFB administration.

Conclusions—The data derived from this study provide strong experimental evidence for the utility of the self-administered, 30-day TLFB in collecting daily alcohol consumption in large observational studies of HIV-positive and HIV-negative individuals.

THE TIMELINE FOLLOWBACK (TLFB) interview was designed as a method of obtaining retrospective self-report data on daily alcohol consumption by personal interview (Sobell et al., 1979; Sobell and Sobell, 2003). One area in which application of the TLFB can be extremely useful is in the collection of retrospective data relevant to hypotheses about the co-occurrence of health-related behaviors among participants in large observational samples. One such application involves the use of the TLFB in research on the relationship between alcohol use on the event level and behavior relevant to HIV/AIDS and

its treatment (Irwin et al., 2006; Stein et al., 2002; Weinhardt et al., 2001). Braithwaite et al. (2006) used the 30-day telephone interview TLFB to test the relationship between alcohol consumption and adherence to prescribed medication regimens among participants in the Veterans Aging Cohort Study (VACS; Justice et al., 2006). VACS is an eight-site longitudinal study of HIV-positive and HIV-negative individuals presenting for treatment in the U.S. Department of Veterans Affairs Infectious Disease and General Medicine clinics, respectively. VACS has enrolled more than 6,000 patients to date. As part of the baseline assessment, a 30-day TLFB was administered to 50% of the participants to measure alcohol consumption and other HIV-related behaviors at study enrollment.

Braithwaite et al. (2006) used these data from the VACS to show that adherence was temporally related to alcohol consumption in a dose-response pattern. They showed that the TLFB has utility in large, multisite observational studies of clinical samples. Despite the utility of the 30-day telephone interview TLFB, these data were collected at considerable cost, which led to the decision to administer the TLFB to only 50% of the VACS baseline sample. Therefore, it would be a clinical research advance if it could be demonstrated that event-level information on the co-occurrence of alcohol use and HIV-related behaviors could be collected reliably by use of a less expensive method, such as participant self-administration.

The purpose of this study was to conduct an experiment to compare the use and agreement of the 30-day TLFB self-administered or telephone interview-administered survey with VACS participants. Based on previous findings in the literature testing different TLFB administration modes to collect data on several health-related behaviors in diverse populations and contexts, it was hypothesized that there would be good reliability of administering the TLFB by use of both self-administered and telephone interview modes in this veteran patient population and that there would be no differences between modes of TLFB administration in the quantity of alcohol reported in the previous 30 days.

Method

Participants

The participants were recruited from the ongoing VACS. A total of 72 VACS participants were recruited into the experiment from August 2005 to August 2006; the main criterion of eligibility to participate in the TLFB substudy was that the participant must have reported consuming at least one beverage containing alcohol in the previous 30 days. In addition, participants were excluded from participating in the TLFB substudy if they refused to allow investigators to link their VACS data to this substudy. Participants were recruited from four of the VACS sites: Bronx Veterans Affairs (VA) Medical Center, Houston VA Medical Center, Los Angeles VA Medical Center, and VA Pittsburgh Healthcare System.

Measures

The self-administered TLFB assessment consisted of the participant completing a printed TLFB form. The telephone-administered TLFB assessment consisted of an interviewer contacting the participant and administering the TLFB via a telephone interview. The only assessment made in either the self-administered or telephone interview-administered assessment was the TLFB; measures of other variables were obtained in the VACS study.

The TLFB was used to collect retrospective self-report data on the number of "standard" drinks consumed on each of the previous 30 days. A standard drink was defined as the alcohol equivalent of 12 oz of domestic (U.S.) beer, 5 oz of 12%-alcohol wine, or 1.5 oz 80-proof distilled spirits. The TLFB data were used to define two measures: (1) total number of drinks in the previous 30 days and (2) classification as "hazardous drinker" in the previous

30 days; a hazardous drinker was defined as someone who consumed more than 60 total drinks or 4 drinks on one occasion for men, and more than 30 total or 3 drinks on one occasion for women (National Institute on Alcohol Abuse and Alcoholism, 2007).

Procedures

Each participant completed two TLFB assessments. After informed consent was obtained from the participants, they were randomly assigned to one of the four experimental conditions that varied on the order/sequence and mode of TLFB assessments (self-self, self-telephone, telephone-self, and telephone-telephone). The two interviews (hereafter referred to as administered at Time 1 and Time 2, respectively) were held not more than 1 week apart. The self-administered TLFB assessment condition was consistent with how self-administered questionnaire data are typically collected in an in-person clinical study. In this regard, participants reported their alcohol use on a printed TLFB form; the research assistant's role was to explain the study to the participant, obtain informed consent, and hand the participant a blank calendar for the purpose of recording the TLFB data. On providing informed consent, the participant completed the 30-day calendar. While the participant completed the calendar, the research assistant was available to answer any questions that the participant may have and ensure that the data were complete. In the telephone conditions, interviewers contacted the participants and administered the TLFB by telephone interview. Interviewers used a computer-assisted telephone interviewing protocol to administer the TLFB. Research assistants were trained in conducting TLFB assessments according to procedures specified in Sobell and Sobell (1996). Participants received $20 after completion of the second TLFB assessment.

Results

Sample description and missing data

Among the 72 participants, 3 were missing 1 day of data in the first TLFB administration; for the second TLFB, 5 participants were missing 1 day of data, 1 participant was missing 4 days of data, 1 participant was missing 6 days of data, and 1 participant was missing 12 days of data. For these participants, the missing days were filled in on the TLFB calendar with the mean number of drinks for the nonmissing days in the respective assessment session.

Two participants were missing data for an entire interview and were dropped from all analyses, to give a final number of 70 participants for this study. There were no statistical differences among the four experimental conditions in demographic variables. However, there was a significant difference in median number of days between TLFB assessments, as the telephone-telephone groups (6 days) was higher (Kruskal-Wallis test, $p = .05$) than that of the other three groups (3 or 4 days).

Amount of alcohol consumed

Table 1 presents the median total number of standard drinks reported across the previous 30 days by the TLFB and the percentage of participants classified as hazardous drinker based on those reports for each of the experimental groups and for the total sample. Several tests were conducted on these data to investigate the prediction of no differences between modes of TLFB administration. First, a Kruskal-Wallis analysis of variance of ranks of total drinks reported across the 30 days showed no differences among the groups at Time 1 or at Time 2. In addition, a Wilcoxon signed ranks test showed no differences in the ranks of the number of drinks reported between the two assessments for the total sample or between times among the four groups.

The analyses of the hazardous drinker classification data first tested whether the four experimental conditions differed in the proportion of cases classified as "hazardous" at Time 1 and Time 2, respectively. Chi square tests showed no differences among the groups at Time 1 or at Time 2. McNemar's test of correlated proportions showed no differences in hazardous drinker classification rates between Time 1 and Time 2 overall or for any of the four groups.

Reliability

The reference point in these analyses was test-retest reliability, which was excellent for the total number of drinks data; the Spearman's ρ for the self-self group was .93, and it was .96 for the telephone-telephone group. The consistency in reports of total number of drinks was lower for the two switch-mode groups than for the same-mode groups but was nevertheless high and statistically significant (p's all < .01). Furthermore, percentage agreement in classification as hazardous drinker between the two times was 93.8 for the self-self group (κ = .87) and 100.0 (κ = 1.00) for the telephone-telephone group. However, switching modes caused considerably less agreement in hazardous drinker classifications between Times 1 and 2 for both the self-telephone and telephone-self groups.

A supplemental, descriptive analysis was run on self-reports of the amount of alcohol consumed on *each day* of the 30-day retrospective interval. Accordingly, each participant contributed 30 data points to each of the relevant computations These analyses showed that the range of Spearman's ρ was .71-.77, all in the good consistency range.

Multivariate model

The final analysis run tested a logistic regression model that included both whether participants switched mode of TLFB administration and time between TLFB administration (<4 days or 4-7 days) in the prediction of agreement of hazardous drinker classification. The results showed that the time between interviews was not significantly associated with agreement. However, consistent with the analyses described earlier, the odds that participants who did not switch TLFB administration modes had consistent classifications as hazardous drinker were 11.00 times higher than the odds for those who did switch modes.

Discussion

In this study, the self-administered and telephone interview modes of TLFB administration did not yield differential reports of total alcohol consumption in the previous 30 days, or in the rate of classification of participants as hazardous drinkers based on those reports. Furthermore, the test-retest reliability data for both modes of administration were excellent. Overall, this study provides strong empirical support for clinical researchers' use of the self-administered TLFB in in-person clinical research.

One finding was differences among groups in the effects of switching mode of TLFB administration on the consistency of number of drinks reported. In this regard, there was a lower correlation between reports for both switch-mode groups compared with the telephone-telephone group. The correlation between drink totals reported in Time 1 and 2 in the self-self group was also larger than that for either switch-mode group, but the differences were not statistically significant. These findings are not to argue that the test-retest data for number of drinks were poor in any group; no correlation between Time 1 and Time 2 data for any group was less than .84.

The groups did differ considerably in the magnitude of kappa for agreement of classification of patients as hazardous drinkers at Times 1 and 2. It makes sense that switching modes of TLFB administration would yield less consistent data compared with the test-retest data,

given the change in the assessment context and its associated cues that the participants in the switch-mode groups experienced.

It may be argued that the finding of no differences in total number of drinks over 30 days and rate of hazardous drinker classification among the four groups was the result of a possible lack of statistical power rather than because of a real absence of differences between the two modes of TLFB administration. In this regard, the sample size for each of the four groups was not large in the context of a major observational study such as VACS, although it was not atypical for an experimental study. Furthermore, this experiment was designed as a VACS feasibility substudy, so that recruiting a large number of participants would not have been consistent with the aims of this substudy. Despite the relatively small group sample sizes that resulted from these design considerations, it seems unlikely that a lack of statistical power can explain the pattern of findings. The data showed no trends that would suggest that increasing the sample size would have changed the outcomes of the relevant statistical tests showing no differences. In addition, the consistency of this experiment's data with the findings of a large volume of previous research on the reliability and accuracy of the TLFB suggests that this experiment's results are not an artifact of low statistical power.

In conclusion, this experiment demonstrated that using the self-administered TLFB to collect retrospective 30-day daily alcohol use data in large, multisite VA infectious disease and general medicine clinical samples may be a viable option for researchers. Future research may address questions of testing periods longer than 30 days and the reliability and accuracy of collecting co-occurring symptoms or behaviors on the event level in the context of large, multisite studies of clinical populations.

References

Braithwaite RS, McGinnis KA, Conigliaro J, Maisto SA, Crystal S, Day N, Cook RL, Gordon A, Bridges MW, Seiler JFS, Justice AC. A temporal and dose-response association between alcohol consumption and medication adherence among veterans in care. Alcsm Clin. Exp. Res. 2006; 29:1190–1197.

Irwin TW, Morgenstern J, Parsons JT, Wainberg M, Labouvie E. Alcohol and sexual risk behavior among problem drinking men who have sex with men: An event level analysis of timeline followback data. AIDS Behav. 2006; 10:299–307. [PubMed: 16482407]

Justice AC, Dombrowski E, Conigliaro J, Fultz SL, Gibson D, Madenwald T, Goulet J, Simberkoff M, Butt AA, Rimland D, Rodriguez-Barradas MC, Gilbert CL, Oursler KAK, Brown S, Leaf DA, Goetz MB, Bryant K. Veterans Aging Cohort Study (VACS): Overview and description. Med. Care. 2006; 44(8 Suppl. 2):S13–S24. [PubMed: 16849964]

National Institute on Alcohol Abuse and Alcoholism. Helping Patients Who Drink Too Much: A Clinician's Guide. Updated 2005 Edition. National Institute on Alcohol Abuse and Alcoholism; Bethesda, MD: 2007. NIH Publication No. 07-3769

Sobell LC, Maisto SA, Sobell MB, Cooper AM. Reliability of alcohol abusers' self-reports of drinking behavior. Behav. Res. Ther. 1979; 17:157–160. [PubMed: 426744]

Sobell, LC.; Sobell, MB. Alcohol Timeline FollowBack: User's Guide. Addiction Research Foundation; Toronto, Canada: 1996.

Sobell, LC.; Sobell, MB. Alcohol consumption measures.. In: Allen, JP.; Wilson, VB., editors. Assessing Alcohol Problems: A Guide for Clinicians and Researchers. 2nd Edition. National Institute on Alcohol Abuse and Alcoholism; Bethesda, MD: 2003. p. 75-99.NIH Publication No. 03-3745

Stein MD, Charuvastra A, Anderson B, Sobota M, Friedmann PD. Alcohol and HIV risk taking among intravenous drug users. Addict. Behav. 2002; 27:727–736. [PubMed: 12201380]

Weinhardt LS, Carey MP, Carey KB, Maisto SA, Gordon CM. The relation of alcohol use to HIV-risk sexual behavior among adults with a severe and persistent mental illness. J. Cons. Clin. Psychol. 2001; 69:77–84.

Table 1

Median total drinks past 30 days reported and percentage of group classified as hazardous drinker, by experimental group, for Assessment Times 1 and 2

Experimental group	n	Median no. of drinks		Percentage hazardous drinkers	
		Time 1	Time 2	Time 1	Time 2
Self-self	16	15.5	21.5	62.5	56.3
Self-telephone	19	35.0	30.0	57.9	57.9
Telephone-self	17	28.0	25.0	70.6	52.9
Telephone-telephone	18	26.5	24.5	55.6	55.6
All participants	70	28.0	25.0	61.4	55.7

Notes: *Hazardous drinker* is defined as, during the previous 30 days, having greater than 60 total drinks or 4 drinks on one occasion for men or greater than 30 total drinks or 3 drinks on one occasion for women.