Changes in Substance Use-Related Health Risk Behaviors on the Timeline Follow-Back Interview as a Function of Length of Recall Period

Anne Buu[1], **Runze Li**[2], **Maureen A. Walton**[1], **Hanyu Yang**[2], **Marc A Zimmerman**[3], and **Rebecca M Cunningham**[3,4,5]

[1]Psychiatry, University of Michigan, Ann Arbor, Michigan, USA

[2]Statistics, Pennsylvania State University, University Park, Pennsylvania, USA

[3]University of Michigan School of Public Health, Ann Arbor, Michigan, USA

[4]University of Michigan Injury Center, Ann Arbor, Michigan, USA

[5]Emergency Medicine, University of Michigan, Ann Arbor, Michigan, USA

Abstract

The timeline follow-back (TLFB) interview was adopted to collect retrospective data on daily substance use and violence from 598 youth seeking care in an urban Emergency Department in Flint, Michigan during 2009–2011. Generalized linear mixed models with flexible smooth functions of time were employed to characterize the change in risk behaviors as a function of the length of recall period. Our results suggest that the 1-week recall period may be more effective for capturing atypical or variable patterns of risk behaviors, whereas a recall period longer than 2 weeks may result in a more stable estimation of a typical pattern.

Keywords

timeline follow-back; psychometric; alcohol use; marijuana use; violence

INTRODUCTION

The timeline follow-back (TLFB, Sobell & Sobell, 1992) is a technique that uses a calendar and structured interview to assist retrospective recall of daily alcohol consumption over a specified time period (e.g., 1 month). It has also been adopted to assess a variety of health risk behaviors other than alcohol use. We conducted a survey of the literature in the last 3 decades using PubMed, Web of Science, and PsycINFO with the following key words: time-line followback (follow-back), TLFB, retrospective recall, calendar method, drinking,

alcohol, substance, drug, violence (violent), aggression, sexual behavior, sex. Our survey identified 238 empirical studies that employed this technique to collect data on alcohol use (86%), drug use (30%), violence (8%), and sexual risk behavior (3%). Its popularity stems from its well-known strengths. In comparison to daily data collection using new technology such as an interactive voice response (IVR) computer-based telephone system, it is less costly and demanding (Searles, Helzer, & Walter, 2000); it also does not have the issue of measurement reactivity that is defined as reducing health risk behaviors due to constant self-monitoring (Tucker, Blum, Xie, Roth, & Simpson, 2012). Further, it is more sensitive to atypical heavy drinking and variable patterns of drinking than traditional quantity-frequency measures that require estimation of an average or typical amount of consumption over a period of time (Sobell & Sobell, 1992). This strength is particularly crucial for studying young people who have not yet developed regular patterns of drinking (Collins, Kashdan, Koutsky, Morsheimer, & Vetter, 2008).

Although several researchers have found excellent psychometric properties of TLFB, the majority of them have focused on establishing the test–retest reliability that demonstrated people's ability to *consistently* recall their risk behavior in two interviews with a lag of a couple of weeks but provided little evidence about the *accuracy* of the recall (Schroder, Carey, & Vanable, 2003). Some researchers reported high criterion-related validity indicated by strong correlations between TLFB data and prospective daily reports (as gold standards) through an IVR system or mobile phone text-messaging (Bardone, Krahn, Goodman, & Searles, 2000; Kranzler, Abu-Hasaballah, Tennen, Feinn, & Young, 2004; Searles, Helzer, Rose, & Badger, 2002; Simpson et al., 2010; Simpson, Xie, Blum, & Tucker, 2011; Suffoletto, Callaway, Kristan, Kraemer, & Clark, 2012; Tucker, Foushee, Black, & Roth, 2007). Both the test–retest reliability and criterion-related validity have been established based on *summary measures* rather than *daily patterns*. Yet, summary measures have been known to leave out clinically meaningful information (Wang, Winchell, McCormick, Nevius, & O'Neill, 2002). For example, average drinks per day may not differentiate between one person who drinks moderately every day and another person who only binges on the weekend. In fact, a handful of studies showed that the day-to-day correspondence measured by within-subject correlations between TLFB and IVR data varied from -1 to 1 across individuals (Perrine, Mundt, Searles, & Lester, 1995; Searles et al., 2000; Simpson et al., 2011; Tucker et al., 2007). Furthermore, Simpson et al. (2010) used the graphs of two study participants' number of drinks along the 28 monitoring days to demonstrate that although the TLFB and IVR data presented the same general pattern of use, the retrospective TLFB tended to capture less day-to-day variability. Thus, sophisticated psychometric analysis that can characterize the daily patterns as well as intra- and inter-individual variability of TLFB data can enhance our knowledge of its strengths as well as limitations.

Different lengths of recall intervals have been adopted for administering TLFB interviews with the 90-day and 30-day being the most common ones (adopted by 47% and 32% of existing studies, respectively). Yet, our knowledge about how people's ability to recall their drinking behavior varies with the length of recall period is limited. The test–retest reliability of the 90-day TLFB as measured by Pearson's *r* on summary measures (e.g., total number of drinks, number of heavy drinking days) was shown to be unchanged across three different

time windows (1–30 days, 31–60 days, and 61–90 days prior to the assessment date) in multiple studies (Carey, Carey, Maisto, & Henson, 2004; Sobell, Sobell, Leo, & Cancilla, 1988; Sobell, Sobell, Klajner, Pavan, & Basian, 1986). The day-to-day agreement between test and retest upon abstinence or drinking, however, varied in a wide range across the interval of 90 days (Rice, 2007). Rice (2007) also found that the odds of reporting a drinking day decreased from the time window closest to the assessment (1–30 days) to the furthest window (61–90 days). Hoeppner, Stout, Jackson, and Barnett (2010) compared the data of 30-day TLFB and repeated 7-day TLFB from the same participants and found that more drinking was reported on the repeated 7-day TLFB than on the 30-day TLFB and this discrepancy increased as the length of recall period (measured in the number of days between interview date and day of drinking) increased. This result should be interpreted with caution though as the 30-day TLFB was interviewer-administered in person or by phone whereas the 7-day TLFB was self-administered online. Thus, the observed differences could also be contributed to type of administration or presentation medium (Brener, Billy, & Grady, 2003; Tourangeau & Yan, 2007).

In the past 10 years, the amount of studies using TLFB to assess risk behaviors highly comorbid with alcohol use such as drug use and violence has increased by over 5 times. These risk behaviors tend to have different levels of sensitivity and frequency from alcohol use (Brener et al., 2003; Tourangeau et al., 2007) so the TLFB technique may work differently for them. For example, self-report data from TLFB may be more accurate for those risk behaviors that are less sensitive or frequent. Yet, our knowledge about the psychometric properties of TLFB has primarily relied on alcohol use data. Particularly, psychometric studies that examine people's ability to recall different risk behaviors as a function of the length of recall period are needed. Furthermore, the occurrence or quantity/ frequency of health risk behaviors has been shown to vary across days during a week with peaks around weekends and holidays (Del Boca, Darkes, Greenbaum, & Goldman, 2004; Dierker et al., 2008; Woodyard & Hallam, 2010). Notably, such weekly variation has not been taken into account in previous studies that investigated people's ability to recall their drinking behavior in various lengths of the recall period and, therefore, could have potentially confounded the results.

This study builds upon previous research in some important ways. First, we fit flexible smooth functions of time to characterize the *daily patterns* of risk behaviors reported on TLFB across time in the recall period, instead of calculating summary measures for different time segments that may lose important information. Second, because great *individual differences* have been found in previous studies (Perrine et al., 1995; Searles et al., 2000; Simpson et al., 2011; Tucker et al., 2007), we provide a statistical modeling approach that employs random effects to take account of such variability. Third, our statistical models also reflect fluctuations in risk behaviors due to weekends and holidays. Fourth, our analysis involves TLFB data on not only alcohol use but also marijuana use and violence so we can examine if people's recall ability as a function of the length of recall period varies across risk behaviors. Fifth, because of the longitudinal design of our study, we are able to examine if the patterns observed at baseline can be replicated at the 6 month follow up.

METHOD

Design and Sample

The present study is a secondary analysis of data collected as part of the Flint Youth Injury (FYI) study, which is an ongoing prospective two-year longitudinal study of 14 to 24 year-old youth with recent drug use who sought care in an urban Emergency Department (ED). FYI was conducted at Hurley Medical Center that is a large urban level-1 trauma center located in Flint, Michigan. The Center is the only public hospital located in Flint, where poverty and crime rates are high and comparable to other urban areas, including Detroit, Michigan; Hartford, Connecticut; Camden, New Jersey; St. Louis, Missouri; and Oakland, California (Federal Bureau of Investigation, 2007). Study procedures were approved and conducted in compliance with the Institutional Review Boards for the University of Michigan and Hurley Medical Center. A Certificate of Confidentiality was obtained from the National Institutes of Health.

Potential participants (ages 14–24) were approached in ED treatment and waiting areas by trained research assistants. Recruitment occurred from December 2009 to September 2011 with coverage of 24 hours per day on Thursday through Sunday and of 21 hours per day on Monday to Wednesday. Exclusion criteria included presenting to the ED for acute sexual assault, child abuse, or suicidal ideation/attempt, altered mental status or psychotic symptoms which would preclude informed consent, non-English speaking, or absence of a parent/guardian (if a patient was under 18 years old). After written consent was obtained, the patient self-administered a computerized screening survey (20–40 minutes) on a laptop computer with touch screen and audio capability and received a $1.00 gift (e.g., pens, notebooks) for participation. Using a standardized instrument, the ASSIST (WHO ASSIST Working Group, 2002), the screening survey asked patients to indicate their use in the past 6 months among 9 drug categories including cannabis, cocaine, methamphetamine, inhalants, hallucinogens, street opioids, prescription stimulants, sedatives/sleeping pills, and prescription opioids. All participants who reported past six-month drug use on the screening survey were invited to participate in a baseline survey (70–90 minutes; $20 remuneration) that included self-administered and research assistant administered portions (including the TLFB interview). Both surveys were privately administered and, to ensure privacy, family or friends accompanying the patient in the ED were not allowed in the same location as the study participant during survey administration. All the participants who completed the baseline survey were followed up in 6, 12, 18, and 24 months after the ED visits.

It is important to note that the FYI study was designed to oversample youth presenting to the ED with violent injury in order to inform youth violence prevention. Thus, attempts were made to screen all youth presenting to the ED for violent injury. Patients with violent injuries who were too unstable to recruit in the ED and who were admitted to the hospital were approached in the hospital if they stabilized within 72 hours. Based on the age and gender of youth presenting with violent injury who reported past 6 month drug use, we proportionally sampled a comparison group who sought care for reasons other than violent injuries (e.g., abdominal pain, fever). The youth in this comparison group was also screened and completed the baseline and follow-up assessment following the same procedure. The

sample used in this study consisted of 598 youth (58% with violent injuries) who have completed TLFB assessment at the baseline. Fifty-nine percent of the participants were male, 61% were African-American, and 69% received public assistance. At baseline, 67% of the study sample reported alcohol use, 97% reported marijuana use, and 93% reported violent behaviors in the past 6 months. In addition, the proportions of users of other drugs were much lower (we only report the top three here: 12% for sedatives/sleeping pills, 11% for prescription opioids, and 6% for cocaine). About 85% of these participants (i.e. 510) also participated in the 6 month follow-up assessment and their data were used to verify the results derived from the baseline data.

Timeline Follow-Back Interview

Standard administration instructions were used for TLFB interviews. The interviewer presented a calendar to the participant and asked him/her to recall any alcohol or drug use behaviors in the past month. They first obtained an overall picture of the participant's past month by identifying holidays and significant events (e.g., birthdays, parties, or accidents) as well as routines related to school, work, or sports. The participant was next asked to recall his/her daily alcohol or drug use during the past month using the calendar as a temporal framework. For alcohol use, the amount of drinks was inquired and converted to the quantity of standard drinks using an alcohol equivalents chart with the participant's input, whereas the participant was not asked to quantify his/her drug use (i.e., only yes/no was recorded for each drug category). Once the information about substance use behaviors was gathered, the interviewer switched the focus to any fighting or violence the participant has been involved in during the past month. For every fight, we collected detailed information such as the reason for the fight and whether alcohol or drug was used before or during the fight. After the TLFB interview, the completed calendar was transcribed into a coding sheet by the interviewer. A second research assistant transcribed the calendar into another coding sheet. Discrepancy was later discussed and resolved. The final coding sheet was double entered to ensure data quality.

Statistical Models

In this study, we analyzed the 30-day TLFB data on alcohol use, marijuana use, and violence using generalized linear mixed models (GLMM) implemented in the MCMCglmm R package (Hadfield, 2010). This analysis chose to focus on marijuana use because it was the most frequently mentioned drug and also the number of users for the other drugs was small. Our models characterize the change of self-reported risk behaviors along the length of recall period as flexible smooth functions while adjusting for individual differences as well as fluctuations with weekdays and holidays.

Suppose Y_{ij} is subject i's alcohol consumption (measured in the number of standard drinks) at the t_{ij}th day prior to the TLFB interview ($t_{ij} = 1, \ldots, 29$). The data corresponding to the interview date (i.e., $t_{ij} = 0$) were not used in the analysis for two reasons. First, the risk behavior on the interview date was very unusual due to the prohibition of substance use and violence in the ED setting. Second, the data corresponding to the interview date tended to include behaviors occurring in only a portion of the day (e.g., the participant who was interviewed at 9 am would report behaviors occurring in a shorter period of time than the

one interviewed at 5 pm). We created two binary variables in order to model the peaks of substance use related to weekend W_{ij} and holiday H_{ij} ("yes" is coded as 1; "no" is coded as 0). Because Y_{ij} is a count variable with a high frequency of zero values, the zero-inflated Poisson distribution (Lambert, 1992) was adopted to model such a highly skewed discrete distribution as a finite mixture of the constant 0 and a regular Poisson distribution. The probability of abstinence $\hat{\pi}_{ij}$ (in the zero component) and the mean of alcohol consumption among drinkers $\hat{\lambda}_{ij}$ (in the Poisson component) were modeled as GLMM:

$$\text{logit}(\pi_{ij}) = f_z(t_{ij}) + X_{ij}\beta_z + b_{zi}, \quad (1)$$

$$\log(\lambda_{ij}) = f_p(t_{ij}) + X_{ij}\beta_p + b_{pi}, \quad (2)$$

where $X_{ij} = (W_{ij}, H_{ij})$ contains the covariates with corresponding fixed effects β_z, β_p for the zero and Poisson components, respectively; the random effects $b_{zi} \sim N(0, \sigma_z^2), b_{pi} \sim N(0, \sigma_p^2)$ were employed to account for individual differences. Further, the baseline functions $f_z(t_{ij})$, $f_p(t_{ij})$ are flexible smooth functions of the length of recall period that were approximated using the truncated power basis (de Boor, 2001) with the degree 3 (recommended by Ruppert, Wand, & Carroll, 2003) and knots chosen according to the deviance information criterion (DIC; Spiegelhalter, Best, Carlin, & van der Linde, 2002).

Because the estimation of coefficients in GLMM cannot be expressed in an analytical form, we adopted the Bayesian approach implemented in the MCMCglmm R package to simulate the posterior distributions of the co-efficients. The simulation was run 100,000 times with the first 50,000 times as the burn-in period to ensure convergence. Sensitivity analysis was also conducted to confirm that the results were not sensitive to the choice of priors. Given the resulting estimates, we can calculate the probability of abstinence $\hat{\pi}_{ij}$ and the mean of alcohol consumption among drinkers $\hat{\lambda}_{ij}$ on an ordinary day (i.e., it was not a weekend day or holiday) based on Equations (1) and (2), respectively. The mean of alcohol consumption on each day in the recall period can then be estimated as $(1 - \hat{\pi}_{ij})\hat{\lambda}_{ij}$. The variance of alcohol consumption on each day within the recall period can also be estimated using the deviance divided by the sample size (see the formula corresponding to the zero-inflated Poisson model in PROC GENMOD; SAS, 2008).

Unlike the alcohol use data that contain information of quantity, the data for marijuana use and violence are both binary variables ("yes" was coded as 1; "no" was coded as 0) which were assumed to follow the Bernoulli distribution with the mean μ_{ij} indicating the probability of marijuana use (or violence). This parameter was modeled as GLMM similar to Equation (1). The same procedure described for the alcohol model can be applied to estimate the coefficients in GLMM and the probability of marijuana use (or violence) on an ordinary day. The corresponding variance on each day in the recall period can also be estimated similarly (see the formula corresponding to the binomial model in PROC GENMOD; SAS, 2008).

RESULTS

Daily Average of Substance Use and Violence

We first conducted conventional analysis involving the most commonly adopted summary measure, the average substance use or violence per day, within 4 recall windows: Week 1 (covering the 7 days closest to the TLFB interview date) to Week 4 (furthest from the TLFB). Table 1 lists the means and standard deviations of this summary measure by risk behavior (alcohol use, marijuana use, and violence) and by week. We also conducted pairwise comparison on the means between weeks using Bonferroni adjustment to control the overall p–value at the .05 level. The standard deviations are large in general. For alcohol use, the average consumption at Week 1 was significantly higher than the ones at later weeks. In terms of marijuana use, no difference was found between any two weeks. Regarding violence, the level at Week 1 tended to be higher than the ones at later weeks; the level at Week 2 was also higher than the one at Week 4.

The GLMM Model for Alcohol Use

Because of the excess zero's and discrete scale of the alcohol use data, we adopted the zero-inflated Poisson model to characterize the distribution as a mixture of the constant 0 and a regular Poisson distribution. A major strength of this model is that we can test the effects of weekend and holiday on the probability of abstinence (i.e., the zero component) and the average amount of alcohol consumption among those who drank (i.e., the Poisson component) separately and simultaneously. Table 2 summaries the results of these two sets of hypothesis testing. The Bayesian approach simulated the posterior distribution for each fixed effect. The median of the distribution was adopted as the point estimate. In order to conduct the hypothesis testing at the significance level of .05, we also employed the 2.5th and 97.5th percentiles to form a 95% confidence band. The results on the left panel show that the probability of abstinence was significantly lower on weekend days and holidays. In addition, the magnitude of the effect due to holidays was greater. The right panel of Table 2 also shows that the average amount of alcohol consumption among drinkers was significantly higher on weekend days and holidays. Again, the holiday effect was more salient. Furthermore, the large standard errors of random effects shown on the bottom of Table 2 demonstrate great individual differences in both the zero and Poisson components.

Figure 1 shows the estimates of the probability of abstinence (the left panel) and the average amount of alcohol consumption among drinkers (the right panel) on each day in the recall period, controlling for the effects of weekend and holiday as well as individual difference. Although the probability of abstinence was generally high during ordinary days (i.e., it was not a weekend day or holiday), it increased from 0.93 to 0.98 in the first week and then stayed high for the rest of the time in the recall window. Among drinkers, the average amount of alcohol consumption on an ordinary day was in general high (3–4 drinks per day); we found an overall downward trend from the first week to the fourth week in spite of a slight rebound during the second week. The changes in probability of abstinence and average alcohol consumption among drinkers observed in Figure 1 both contributed to the pattern of change in the overall mean that was observed among the entire sample.

Figure 2 delineates the change in the overall mean and variance of daily alcohol consumption across the 30-day recall period. The left panel shows that the average number of alcohol consumed on an ordinary day was in general low and yet it started at an elevated level and decreased during the first week and then stayed low in the rest of the recall period. The right panel of Figure 2 displays the estimated variance of alcohol use on each day with a LOESS smoother (Cleveland, 1979) fitted to facilitate visualization of the pattern of change during the recall period. In comparison to the pattern of change in the mean, the decreasing trend in the variance was more drastic for each day going back in time, especially during the first two weeks.

The GLMM Models for Marijuana Use and Violence

Table 3 depicts the estimates and 95% confidence bands for the effects of weekend and holiday on marijuana use (the left panel) as well as violence behavior (the right panel). Like the finding of alcohol use, the participants were more likely to use marijuana during weekends and holidays; the holiday also has a more salient effect among the two. The right panel of Table 3, however, shows that for violent behavior, the likelihood of occurrence was not affected by weekend or holiday. Moreover, the large standard errors of random effects shown on the bottom of Table 3 demonstrate great individual differences in both marijuana use and violent behavior.

Figure 3 delineates the change in the mean and variance of daily marijuana use across time during the 30-day recall period. The left panel shows that the probability of marijuana use on an ordinary day was generally high (>0.6). Starting at an elevated level, the probability of marijuana use decreased during the first week and then stayed at an almost constant level for the rest of the recall period. The right panel of Figure 3 demonstrates a more drastic decreasing trend in the variance, especially during the first half of the recall period. Figure 4 characterizes the change in the mean and variance of violent behavior for each day going back in time during the recall period. The left panel shows a similar trend in the probability of violent behavior to the one in the probability of marijuana use, except that the probability of violent behavior was low (<0.2) throughout the entire recall period. Like alcohol use and marijuana use, the variance for violent behavior had a more drastic decreasing trend than the mean, especially in the first half of the recall window.

The GLMM Models Based on The Six Months Follow-up Data

One of the possible explanations for the small scale decreasing trend in the mean during the first week of the recall period commonly found across all the three types of risk behaviors is that the participants of this study were recruited from their ED visits that may have resulted from a short period of heavier involvement in substance use or violence. To verify this hypothesis, we fit the same set of models on the 30-day TLFB data collected from 510 participants (out of the original 598 participants who provided TLFB data at baseline) who participated in the follow-up assessment conducted in 6 months after the ED visits. Both the alcohol and marijuana use at the follow-up had similar decreasing trends in the mean during the first week while the mean for violence was consistently low during the entire 30-day recall period. Thus, this possible explanation may only be applicable to violence but not to alcohol and marijuana use. In fact, decay in self-reported alcohol consumption has been

found to begin with the second day of recall and persistent over a week in other studies using TLFB in ED settings (Gmel & Daeppen, 2007; Vinson, Reidinger, & Wilcosky, 2003).

DISCUSSION

This is the first study examining the performance of TLFB as a function of the length of recall period across multiple risk behaviors including alcohol use, marijuana use, and violence. Particularly, the psychometric properties of TLFB for violent behaviors have only been investigated in a study on intimate partner violence (Fals-Stewart, Birchler, & Kelley, 2003), although this interview technique has also been adopted to study other types of violence in recent years (e.g., Epstein-Ngo et al., 2013; Parrott, Galllagher, Vincent, & Bakeman, 2010). We found that both weekend and holiday contributed to a higher probability of alcohol and marijuana use as well as a larger amount of alcohol consumption among drinkers. The magnitude of the effect due to holiday was consistently higher than the one for weekend. The likelihood of occurrence for violence, however, did not increase during weekends and holidays. In general, individual differences were high across the three types of risk behaviors. Controlling for the effects of weekend and holiday as well as individual differences, we found that the probability of abstinence from alcohol was high while the one from marijuana was low. Moreover, for all the three risk behaviors, the mean of daily involvement decreased during the first week and then stayed at the same level throughout the rest of the recall period. In comparison to the mean, the variance had a more drastic decreasing trend as the reported date moved further away from the interview date.

The finding that the probability of abstinence from alcohol was high while the one from marijuana was low during ordinary days may be partly related to the fact that the majority of our study participants are African Americans. According to national data, alcohol use disorder was more prevalent among Whites and Hispanics compared to African Americans, whereas marijuana use disorder was greatest among African Americans compared to other race/ethnicities (Pacek, Malcolm, & Martins, 2012). Furthermore, the literature has shown that although young marijuana users tended to report most events on weekdays (Shrier, Walls, Rhoads, & Blood, 2013), young alcohol users were most likely involved in drinking on weekend days when they spent more time socializing (Finlay, Ram, Maggs, & Caldwell, 2012). Thus, within a relapse prevention framework, these findings may stress the importance of identification and avoidance of triggers associated with isolative behaviors (marijuana use) versus social behaviors (alcohol use).

Researchers have distinguished two types of memories: episodic memory and semantic memory (Jaccard & Wan, 1995). While episodic memory refers to the retrieval of information about specific episodes of a behavior, semantic memory refers to generalization about behavior that is stored in memory. Some question formats and short recall periods may encourage individuals to access episodic versus semantic memory when retrospectively reporting risk behaviors. Although the way TLFB is administered aims to encourage the access of episodic memory, participants may tend to switch to semantic memory as the reported date goes further back in time. The finding that the mean became almost constant after the first week and the finding that the variance decreased for each day going back in

time both seem to support such a hypothesis. Our results are also consistent with a recent study showing that 30-day TLFB reports of alcohol consumption were more highly correlated with 14-day reports than with 7-day reports (Fiellin, McGinnis, Maisto, Justice, & Bryant, 2013).

A major limitation of this study is that we do not have a gold standard (i.e., criterion) such as daily substance use or violence data collected through an IVR computer-based telephone system or mobile phone text-messaging to validate the accuracy of the TLFB data across the recall period. Although prospective daily data collection can reduce recall bias or decay, employing it as a gold standard in a criterion-related validation study is technically challenging. The number one issue is that the prospective daily data collection may potentially serve as a mnemonic for the retrospective TLFB report so the correlation between the two measures may be inflated. Another issue is that prospective daily data collection may facilitate participants' self-monitoring and thus alter their risk behaviors. Future studies with more sophisticated designs and advanced statistical methods are needed to deal with these issues. Another important limitation of this study is that our findings were based on a sample of youth presenting to an urban emergency department. The levels of drinking, marijuana use, and violence among this sample tended to be below what one would expect from most clinical trial samples. It is possible that the deterioration of memory found in our sample may occur to a lesser extent among more regular or heavier users. In fact, a previous study has found greater agreement between TLFB and real-time electronic interviews for those who reported more alcohol consumption (Carney, Tennen, Affleck, Del Boca, & Kranzler, 1998). Since TLFB is one of the most commonly used measures in clinical trials, it is important to verify if our findings can be replicated in those settings in future studies. Moreover, some researchers adopt an alternative administration approach that uses past week data to generate a "standard week" as an estimate for preceding weeks and then uses the calendar to make adjustments up or down. The results of this study may not be generalizable to studies employing that approach.

This study has some important strengths. Although we did not have daily process data to validate our results, we were able to take advantage of the longitudinal design of the FYI study and used the 6 month follow-up data to verify our findings derived from the baseline data. Furthermore, this study makes a unique contribution to the literature by adopting a statistical modeling approach to characterize the change in retrospective report of alcohol use, marijuana use, and violent behavior in TLFB interviews as a function of the length of recall period, controlling for the weekend and holiday effects as well as individual differences. A major strength of our modeling approach is that the zero-inflated Poisson model has allowed us to examine quantitative information of substance use at a micro-level by modeling the probability of abstinence and the average consumption among users separately and simultaneously. This model has also been shown to estimate parameters more accurately than the conventional Poisson model that was not designed to handle the excess zeros commonly observed in discrete data collected from the substance abuse field such as substance use quantity/frequency and symptom count (Buu, Johnson, Li, & Tan, 2011; Buu, Li, Tan, & Zucker, 2012). Another important strength of our modeling approach is that we adopted the smoothing technique to characterize the change of self-reported risk behavior across time in the recall period as a flexible smooth function, without imposing any pre-

specified simple shapes such as the commonly adopted linear or quadratic function that hardly fits empirical data with many time points like the data typically collected from TLFB interviews.

Our study has resulted in some important implications for future use of TLFB to assess alcohol use, marijuana use, and violence. First, the effects of weekend and holiday on alcohol and marijuana use are significant and therefore have to be taken into account when the TLFB data are analyzed for studying either psychometric properties of TLFB or the association between risk behaviors and precursors/consequences. For example, a previous study has found that recall biases of alcohol consumption reported in TLFB interviews were apparent for every day of the week, but the bias was highest for Fridays and Saturdays (Gmel & Daeppen, 2007). Second, our analysis employing the zero-inflated Poisson model demonstrates that it is necessary to model the probability of abstinence and the amount of consumption among users separately and simultaneously when the data are discrete and have excess zeros. Failing to do so would lead to a biased picture that is dominated by the information from zero values. Our results show that although the probability of abstinence on an ordinary day was high, the people who did drink still consumed a considerable amount of alcohol in a day and their self-reported consumption decayed as the length of recall period became greater. This kind of micro-level information would not be available from conventional analysis that relies on simple statistics of summary measures like the one presented in Table 1. Third, our results replicate previous research that found individual differences in self-reported risk behaviors are great. We have also proposed a modeling approach that employs random effects to take account of this inter-individual variability as we study people's overall recall ability in TLFB interviews. Fourth, our findings indicate that although TLFB was designed to encourage participants' access of episodic memory that better captures atypical and variable patterns of risk behaviors, people may still switch to semantic memory as the reported date goes further back in time. Thus, if the purpose is to study atypical or variable patterns of risk behaviors, the one-week recall period may be more appropriate. On the other hand, if the purpose is to derive a typical pattern of risk behaviors, a recall period longer than 2 weeks may result in a more stable estimation.

Acknowledgments

Biographies

Anne Buu, Ph.D., is Research Assistant Professor in the Department of Psychiatry and Addiction Research Center at the University of Michigan. Her research interests include longitudinal data analysis, bioinformatics, daily patterns of substance use and related health risk behaviors, and substance abuse prevention/intervention. She is the Principal Investigator of two methodology projects (K01 & R01) funded by the National Institutes of Health.

Runze Li, Ph.D., is Distinguished Professor of Statistics and Professor of Public Health Sciences, a member of the Methodology Center, Pennsylvania State University, University Park, Pennsylvania, USA. Dr. Li's research interest includes analysis of intensive longitudinal data, variable selection for high dimensional data, and statistical methodology applications to substance use research, life science research, and engineering research. His work has been funded by the National Institute on Drug Abuse (NIDA) and the National Science Foundation. He has published in a broad assortment of methodological and substantive journals. Dr. Li is co-editor(in-chief) of Annals of Statistics, and served as Associate Editor of Annals of Statistics, Journal of American Statistical Association, and Statistica Sinica. Awards include NSF Career award, Fellow of Institute of Mathematical Statistics, Fellow of American Statistical Association and the United Nations' World Meteorological Organization Gerbier-Mumm International Award for 2012.

Maureen Walton, M.P.H., Ph.D., is Associate Professor in the Department of Psychiatry and Addiction Research Center at the University of Michigan. Her research interests include developing and testing the efficacy of interventions for alcohol, drug use, and violence in community health care settings, such as the emergency department, primary care, and substance use treatment. Her research focuses on the interrelationship among multiple risk behaviors such as alcohol, illicit drugs, and violence, particularly among traditionally understudied populations such as adolescents, women, and African-Americans.

Hanyu Yang, Ph.D. student in Department of Statistics, the Pennsylvania State University, University Park, Pennsylvania, USA. Mr. Yang has interests in the intensive longitudinal data analysis, and its applications to substance use research.

Marc A. Zimmerman, Ph.D., is Professor of Health Behavior and Health Education, and Psychology, at the University of Michigan. Dr. Zimmerman's research focuses on adolescent health and resiliency. He directs the CDC funded Prevention Research Center of Michigan and the CDC funded Youth Violence Prevention Center. His work includes both longitudinal studies of development and evaluation of community-based prevention programs. He is also editor of Youth & Society.

Dr. Cunningham, Associate Professor, is in the Department of Emergency Medicine, University of Michigan Medical School, and an Associate Professor, Health Behavior & Health Education, University of Michigan School of Public Health. She is also Director of the University of Michigan Injury Center, has a distinguished career in researching intentional injury and substance use prevention, particularly of youth and young adult populations. Her focus on brief interventions in the emergency room has helped position the emergency department as a critical location for public health interventions, specifically for violence. She is currently leading two NIH-funded studies on substance abuse: one focusing on the intersection of youth violence and drug use, and one focusing on underage alcohol misuse and associated injury. She concurrently continues her work as a practicing Emergency Department physician at the University of Michigan Health System.

GLOSSARY

|	
Generalized linear mixed model (GLMM)	This is a commonly adopted longitudinal data analysis method that employs random effects to handle dependence among repeated measures within the same subject/cluster. It can accommodate to different measurement scales of the outcomes using different link functions (e.g., the identify link for continuous outcomes, the logit link for binary outcomes).
Timeline follow-back interview (TLFB)	This is a technique that uses a calendar and structured interview to assist retrospective recall of daily alcohol consumption over a specified time period. It has also been adopted to assess a variety of other health risk behaviors such as drug use, violence, and HIV risk sexual behaviors.

REFERENCES

Bardone AM, Krahn DD, Goodman BM, Searles JS. Using interactive voice response technology and timeline follow-back methodology in studying binge eating and drinking behavior: Different answers to different forms of the same question? Addictive Behaviors. 2000; 25:1–11. [PubMed: 10708315]

Brener ND, Billy JOG, Grady WR. Assessment of factors affecting the validity of self-reported health-risk behavior among adolescents: Evidence from the scientific literature. Journal of Adolescent Health. 2003; 33:436–457. [PubMed: 14642706]

Buu A, Johnson NJ, Li R, Tan X. New variable selection methods for zero-inflated count data with applications to the substance abuse field. Statistics in Medicine. 2011; 30:2326–2340. [PubMed: 21563207]

Buu A, Li R, Tan X, Zucker RA. Statistical models for longitudinal zero-inflated count data with applications to the substance abuse field. Statistics in Medicine. 2012; 31:4074–4086. [PubMed: 22826194]

Carey KB, Carey MP, Maisto S, Henson JM. Temporal stability of the timeline followback interview for alcohol and drug use with psychiatric outpatients. Journal of Studies on Alcohol. 2004; 65:774–781. [PubMed: 15700516]

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 "electronic interviews". Journal of Studies on Alcohol. 1998; 59:447–454. [PubMed: 9647427]

Cleveland W. Robust locally weighted regression and smoothing scatterplots. Journal of the American Statistical Association. 1979; 78:829–836.

Collins LR, Kashdan TB, Koutsky JR, Morsheimer ET, Vetter CJ. A self-administered timeline follow-back to measure variations in underage drinkers' alcohol intake and binge drinking. Addictive Behaviors. 2008; 33:196–200. [PubMed: 17720324]

De Boor, C. A practical guide to splines. NY: Springer; 2001.

Del Boca FK, Darker J, Greenbaum PE, Goldman MS. Up close and personal: Temporal variability in the drinking of individual college students during the first year. Journal of Counseling and Clinical Psychology. 2004; 72:155–164.

Dierker L, Stolar M, Lloyd-Richardson E, Tiffany S, Flay B, Collins L, Nichter M, Bailey S, Clayton R. Tobacco, alcohol, and marijuana use among first-year U.S. college students: A time series analysis. Substance Use and Misuse. 2008; 43:681–699.

Epstein-Ngo QM, Cunningham RM, Whiteside LK, Chermack ST, Booth BM, Zimmerman MA, Walton MA. A daily calendar analysis of substance use and dating violence among high risk urban youth. Drug and Alcohol Dependence. 2013; 130:194–200. [PubMed: 23219602]

Fals-Stewart W, Birchler GR, Kelley ML. The timeline followback spousal violence interview to assess physical aggression between intimate partners: Reliability and validity. Journal of Family Violence. 2003; 18:131–142.

Federal Bureau of Investigation. Crime in the United States. 2007 Retrieved February 20, 2014, from http://www.fbi.gov/ucr/07cius.htm.

Fiellin D, McGinnis KA, Maisto SA, Justice AC, Bryant K. Measuring alcohol consumption using timeline followback in non-treatment-seeking medical clinic patients with and without HIV infection: 7-, 14-, or 30-day recall. Journal of Studies on Alcohol and Drugs. 2013; 74:500–504. [PubMed: 23490581]

Finlay AK, Ram N, Maggs JL, Caldwell LL. Leisure activities, the social weekend, and alcohol use: Evidence from a daily study of first-year college students. Journal of Studies on Alcohol and Drugs. 2012; 73:250–259. [PubMed: 22333332]

Gmel G, Daeppen J-B. Recall bias for seven-day recall measurement of alcohol consumption among emergency department patients: Implications for case-crossover designs. Journal of Studies on Alcohol and Drugs. 2007; 68:303–310. [PubMed: 17286350]

Hadfield JD. MCMC methods for multi-response generalized linear mixed models: The MCMCglmm R package. Journal of Statistical Software. 2010; 33:1–22. [PubMed: 20808728]

Hoeppner BB, Stout RL, Jackson KM, Barnett NP. How good is fine-grained timeline follow-back data? Comparing 30-day TLFB and repeated 7-day TLFB alcohol consumption reports on the person and daily level. Addictive Behaviors. 2010; 35:1138–1143. [PubMed: 20822852]

Jaccard J, Wan CK. A paradigm for studying the accuracy of self-reports of risk behavior relevant to AIDS: Empirical perspectives on stability, recall bias, and transitory influences. Journal of Applied Social Psychology. 1995; 25:1831–1858.

Kranzler HR, Abu-Hasaballah K, Tennen H, Feinn R, Young K. Using daily interactive voice response technology to measure drinking and related behaviors in a pharmacotherapy study. Alcoholism: Clinical and Experimental Research. 2004; 28:1060–1064.

Lambert D. Zero-inflated Poisson regression, with an application to defects in manufacturing. Technometrics. 1992; 34:1–13.

Pacek LR, Malcolm RJ, Martins SS. Race/ethnicity differences between alcohol, marijuana, and co-occurring alcohol and marijuana use disorders and their association with public health and social problems using a national sample. The American Journal on Addictions. 2012; 21:435–444. [PubMed: 22882394]

Parrott DJ, Galllagher KE, Vincent W, Bakeman R. The link between alcohol use and aggression toward sexual minorities: An event-based analysis. Psychology of Addictive Behaviors. 2010; 24:516–521. [PubMed: 20853937]

Perrine MW, Mundt JC, Searles JS, Lester LS. Validation of daily self-reported alcohol consumption using interactive voice response (IVR) technology. Journal of Studies on Alcohol. 1995; 56:486–490.

Rice C. Retest reliability of self-reported daily drinking: Form 90. Journal of Studies on Alcohol and Drugs. 2007; 68:615–618. [PubMed: 17568968]

Ruppert, D.; Wand, MP.; Carroll, RJ. Semiparametric Regression. New York, NY: Cambridge Press; 2003.

SAS Institute Inc.. SAS/STAT 9.2 User's Guide. Cary, NC: SAS Institute Inc.; 2008.

Schroder KEE, Carey MP, Vanable PA. Methodological challenges in research on sexual risk behavior. Part II. Accuracy of self-reports. Annals of Behavioral Medicine. 2003; 26:104–123. [PubMed: 14534028]

Searles JS, Helzer JE, Walter DE. Comparison of drinking patterns measured by daily reports and timeline follow back. Psychology of Addictive Behaviors. 2000; 14:277–286. [PubMed: 10998953]

Searles JS, Helzer JE, Rose GL, Badger GJ. Concurrent and retrospective reports of alcohol consumption across 30, 90, and 366 days: Interactive voice response compared with the timeline follow back. Journal of Studies on Alcohol. 2002; 63:352–362. [PubMed: 12086136]

Shrier LA, Walls C, Rhoads A, Blood EA. Individual and contextual predictors of severity of marijuana use events among young frequent users. Addictive Behaviors. 2013; 38:1448–1456. [PubMed: 23010685]

Simpson CA, Xie L, Blum ER, Tucker JA. Agreement between prospective interactive voice response telephone reporting and structured recall reports of risk behaviors in rural substance users living with HIV/AIDS. Psychology of Addictive Behaviors. 2011; 25:185–190. [PubMed: 21443312]

Simpson TL, Galloway C, Rosenthal CF, Bush KR, McBride B, Kivlahan DR. Daily telephone monitoring compared with retrospective recall of alcohol use among patients in early recovery. The American Journal on Addictions. 2010; 20:63–68. [PubMed: 21175922]

Sobell, LC.; Sobell, MB. Timeline follow-back: A technique for assessing self-reported ethanol consumption. In: Allen, J.; Litten, RZ., editors. Measuring alcohol consumption: Psychosocial and biological methods. Totowa, NJ: Humana Press; 1992. p. 41-72.

Sobell LC, Sobell MB, Leo GI, Cancilla A. Reliability of a timeline method: Assessing normal drinkers' reports of recent drinking and a comparative evaluation across several populations. British Journal of Addiction. 1988; 83:393–402. [PubMed: 3395719]

Sobell MB, Sobell LC, Klajner F, Pavan D, Basian E. The reliability of a timeline method for assessing normal drinker college students' recent drinking history: Utility for alcohol research. Addictive Behavior. 1986; 11:149–161.

Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society Series B. 2002; 64:583–639.

Suffoletto B, Callaway C, Kristan J, Kraemer K, Clark DB. Text-message-based drinking assessments and brief interventions for young adults discharged from the emergency department. Alcoholism: Clinical and Experimental Research. 2012; 36:552–560.

Tourangeau R, Yan T. Sensitive questions in surveys. Psychological Bulletin. 2007; 133:859–883. [PubMed: 17723033]

Tucker JA, Blum ER, Xie L, Roth DL, Simpson CA. Interactive voice response self-monitoring to assess risk behaviors in rural substance users living with HIV/AIDS. AIDS and Behavior. 2012; 16:432–440. [PubMed: 21311964]

Tucker JA, Foushee HR, Black BC, Roth DL. Agreement between prospective interactive voice response self-monitoring and structural retrospective reports of drinking and contextual variables during natural resolution attempts. Journal of Studies on Alcohol and Drugs. 2007; 68:538–542. [PubMed: 17568958]

Vinson DC, Reidinger C, Wilcosky T. Factors affecting the validity of a timeline follow-back interview. Journal of Studies on Alcohol. 2003; 64:733–740. [PubMed: 14572197]

Wang SJ, Winchell CJ, McCormick CG, Nevius SE, O'Neill RT. Short of complete abstinence: An analysis exploration of multiple drinking episodes in alcoholism treatment trials. Alcoholism: Clinical and Experimental Research. 2002; 26:1803–1809.

WHO ASSIST Working Group. The alcohol, smoking and substance involvement screening test (ASSIST): Development, reliability and feasibility. Addiction. 2002; 97:1183–1194. [PubMed: 12199834]

Woodyard CD, Hallam JS. Differences in college student typical drinking and celebration drinking. Journal of American College Health. 2010; 58:533–538. [PubMed: 20452929]
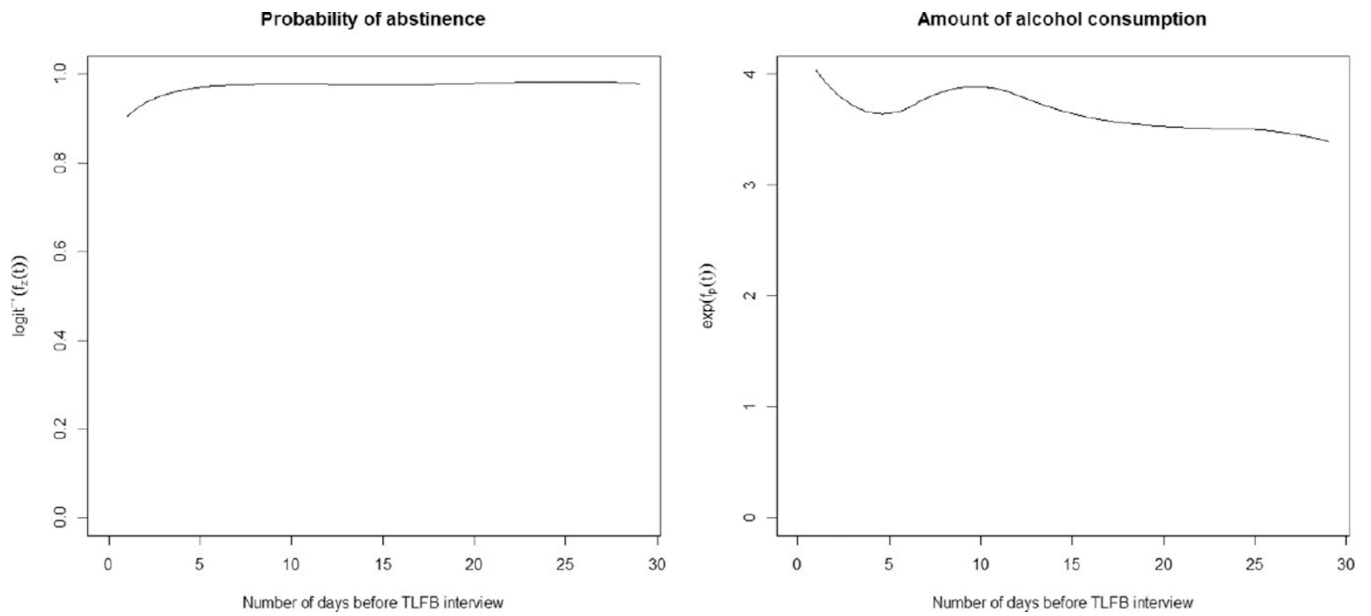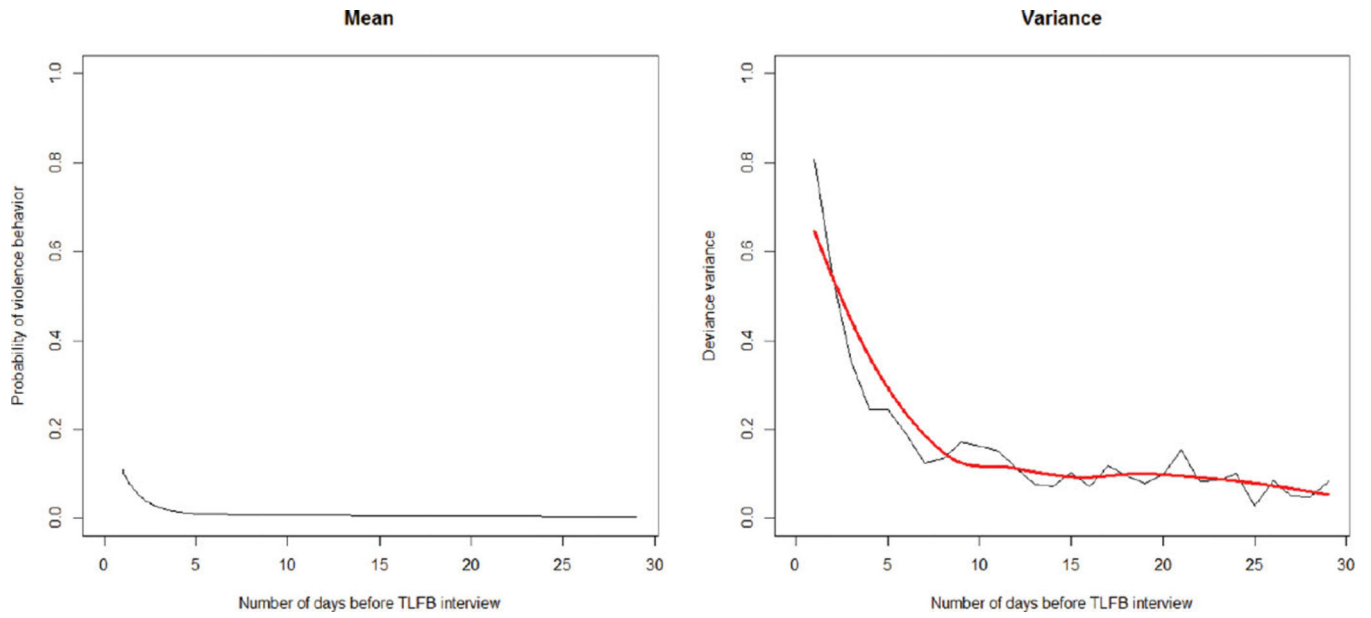
FIGURE 1.
The change in probability of abstinence and average amount of alcohol consumption among drinkers as a function of the length of recall period.
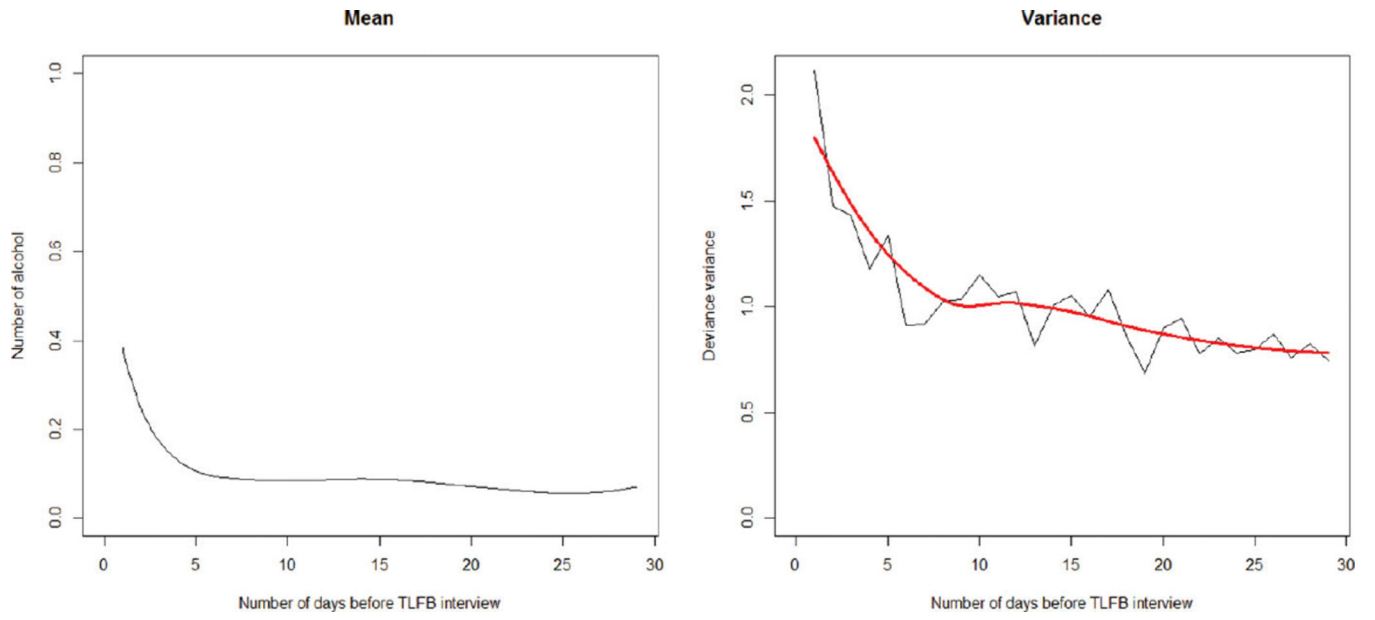
FIGURE 2.
The change in mean and variance of alcohol use as a function of the length of recall period.
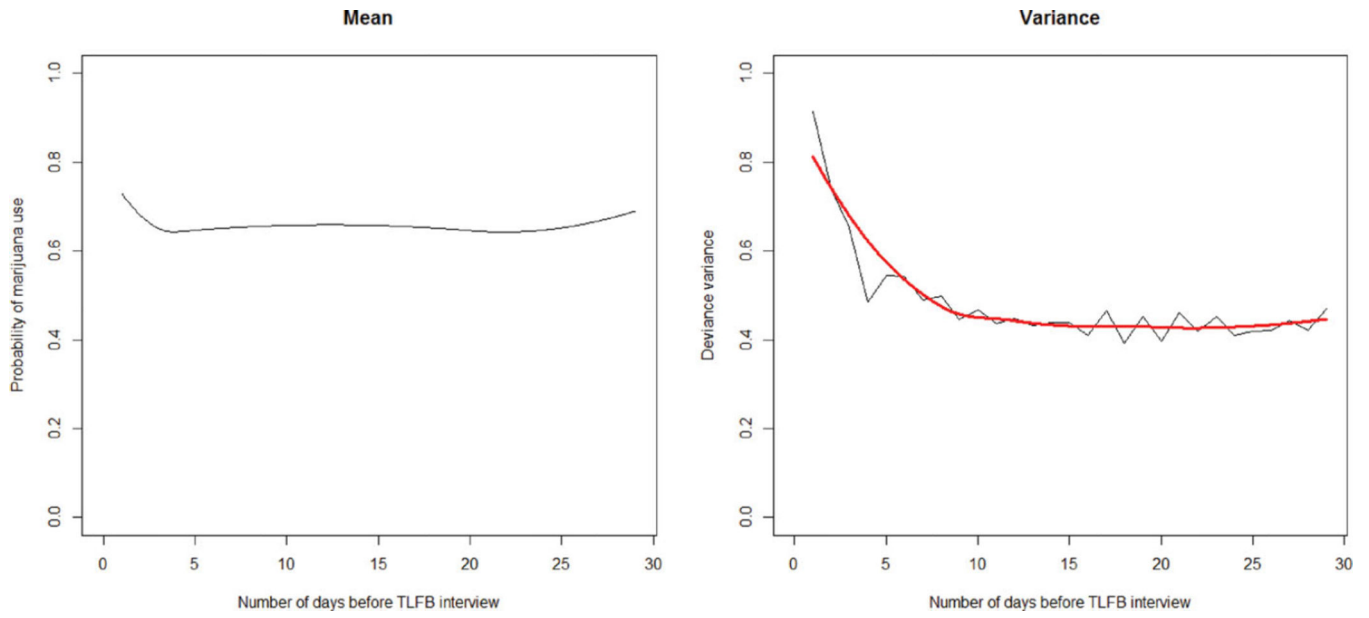
FIGURE 3.
The change in mean and variance of marijuana use as a function of the length of recall
period.

FIGURE 4.
The change in mean and variance of violent behavior as a function of the length of recall period.

TABLE 1

Means and standard deviations of daily averages of substance use and violence by week ($N = 598$)

	Week 1[a]		Week 2		Week 3		Week 4		Significantly different pairs[b]
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	
Alcohol use	1.01	2.28	0.79	2.06	0.76	2.24	0.69	2.16	1–2, 1–3, 1–4
Marijuana use	0.53	0.42	0.52	0.44	0.52	0.44	0.53	0.44	
Violence	0.06	0.11	0.02	0.08	0.02	0.08	0.01	0.08	1–2, 1–3, 1–4, 2–4

[a]Week 1 is the 7 days that were closest to the timeline follow-back interview date.

[b]The pairs of weeks that had significantly different means (pairwise comparison with Bonferroni adjustment controlling the overall *p*-value at the .05 level).

TABLE 2

Estimates of the parameters in the generalized linear mixed models of alcohol use data ($N = 598$)

	Coefficients of Zero Component β_z			Coefficients of Poisson Component β_p		
	Estimate	2.5th percentile	97.5th percentile	Estimate	2.5th percentile	97.5th percentile
Weekend	−1.27[*]	−1.29	−1.25	0.17[*]	0.11	0.24
Holiday	−1.60[*]	−1.77	−1.49	0.27[*]	0.12	0.44
S.E. of random effect	2.16	2.15	2.18	0.76	0.69	0.80

[*]
The coefficient is significantly different from 0 at the .05 level.

TABLE 3

Estimates of the parameters in the generalized linear mixed models of marijuana use and violent behavior data ($N = 598$)

	The model for Marijuana Use			The model for violent behavior		
	Estimate	2.5th percentile	97.5th percentile	Estimate	2.5th percentile	97.5th percentile
Weekend	0.71*	0.61	0.85	0.15	−0.02	0.41
Holiday	0.89*	0.52	1.24	−0.04	−0.64	0.60
S.E. of random effect	4.82	4.71	5.01	1.19	1.05	1.36

*
The coefficient is significantly different from 0 at the .05 level.