HHS Public Access

Author manuscript

Addict Behav. Author manuscript; available in PMC 2021 March 01.

Published in final edited form as:

Addict Behav. 2020 March ; 102: 106198. doi:10.1016/j.addbeh.2019.106198.

Examining measurement reactivity in daily diary data on substance use: results from a randomized experiment

Anne Buu[a,*], **Songshan Yang**[b], **Runze Li**[b], **Marc A. Zimmerman**[c], **Rebecca M. Cunningham**[d], **Maureen A. Walton**[e]

[a]Department of Health Promotion and Behavioral Sciences, University of Texas Health Science Center at Houston, 7000 Fannin St., Suite 2516, Houston, TX 77030, USA

[b]Department of Statistics and The Methodology Center, Pennsylvania State University, 413 Thomas Building, University Park, PA 16802-2111, USA

[c]Department of Health Behavior and Health Education & Injury Center, University of Michigan, 1415 Washington Heights, Ann Arbor, MI 48109, USA

[d]Department of Emergency Medicine & Injury Center, University of Michigan, 2800 Plymouth Rd, Bldg 10-G080, Ann Arbor, MI 48109, USA

[e]Addiction Center & Injury Center, University of Michigan, 4250 Plymouth Road, Ann Arbor, MI 48109, USA

Abstract

The debate about whether measurement reactivity exists in daily diary research on substance use is still unsettled due to the issues of study design and statistical methodology. This study proposes a time-varying effect model (TVEM) that characterizes the trajectory of substance use behaviors with nonparametric functions determined by the data rather than imposes presumed parametric functions. It also allows researchers to investigate the effect of measurement reactivity on not only the likelihood of using substances but also the amount of substance use. The TVEM was applied to analyze diary data on alcohol and marijuana use collected from an experiment, which randomized 307 participants in Michigan into daily and weekly assessment schedules during 2014–2016. This study found short-term measurement reactivity on alcohol use, but did not find a significant reactivity effect on marijuana use. The daily group had smaller odds of abstinence from drinking but lower expected drinking quantity in the first week of assessment, which dissipated by the second week. The results indicate that although daily self-monitoring could have short-term reactivity on substance use behaviors that tend to fluctuate across days, such as alcohol use, it does not affect substance use behaviors that are quite consistent, such as marijuana use. Our findings imply that although daily monitoring of drinking may motivate people to reduce the quantity consumed once they start to drink, it may also arouse their desire to start drinking. Yet, both

*Corresponding author at: 7000 Fannin Street, Suite 2516, Houston, TX 77030, USA. Yuh-Pey.A.Buu@uth.tmc.edu (Anne Buu).

Conflict of Interest The authors declare no conflicts of interest.

effects tend to last only one week, as participants accommodate to the monitoring by the second week.

Keywords

reactivity; diary; experiment; alcohol; marijuana

Introduction

Measurement reactivity is a research question of common interest in behavioral science: could psychological measurement influence the self-reports of behavior by people being assessed? Daily self-reporting on health risk behaviors such as substance use could lead to more mindful self-monitoring which can be a useful form of therapeutic behavior change (Mitchum, Kelley, & Fox, 2016). Thus, measurement reactivity may potentially compromise the validity of studies collecting daily diary data. Alternatively, Reynolds, Robles, and Repetti (2016) pointed out that diary research typically asks participants to report on experiences that have already occurred without providing feedback or encouragement to review behavioral patterns over time, and therefore may not produce therapeutic benefits of self-monitoring.

Researchers investigating measurement reactivity effects on health risk behaviors in diary studies have mainly used an interactive voice response (IVR) system, which is a computer system that can automatically administer surveys with prerecorded audio and record participants' responses into databases (e.g., this technique was already commonly used for consumers to check their account information by interacting with a computer through a telephone keypad or speech recognition before smart phones became available). Although longitudinal observational studies involving daily IVR assessment showed declines in alcohol consumption (Helzer, Badger, Rose, Mongeon, & Searles, 2002) and risky sexual behaviors (Schroder, Johnson, & Wiebe, 2007), these findings were limited by the small sample size and the potential confounding effect due to an increasing amount of missing assessment in later days. Tucker et al. (2012) used IVR to collect daily use of alcohol, illicit drugs, and sexual activity from HIV/AIDS patients for up to 10 weeks and found reactivity effects on drug use and risky sex but not on alcohol use. Measurement reactivity was demonstrated by lower odds of reporting a target risk behavior associated with an increase in the cumulative number of IVR call days (indicating a higher dose), controlling for the frequency of the behavior at baseline. The study also acknowledged that the relationships observed were correlational and warranted further research by experimentally manipulating IVR access.

We know of only two studies that randomized participants to IVR or control groups and evaluated group differences in drinking outcomes after the IVR assessment was finished, adjusting for pre-randomization drinking levels (Helzer, Rose, Badger, Searles, Thomas, Lindberg, & Guth, 2008; Simpson, Kivlahan, Bush, & McFall 2005). Yet, the results were not consistent. Simpson et al. (2005) did not find significant group differences, whereas Helzer et al. (2008) found that the IVR group reported even higher consumption than the

control group (i.e., the opposite effect to measurement reactivity). One common issue of both studies is that drinking outcomes were measured by the timeline follow-back (TLFB; Sobell & Sobell, 1992), a commonly adopted assessment method that uses a calendar and structured interview to assist *retrospective* recall of daily alcohol consumption over a specified time period, rather than by the IVR data, because the latter were not available for the control group. Thus, a possible explanation of the resulting opposite effect is that the frequent IVR calls may have served as a mnemonic for the TLFB reports in the IVR group, making the outcomes from the two groups incomparable.

To address the dilemma over the dependence on TLFB reports due to inclusion of the control group that was not assessed at all during the experimental period, a recent study (Buu, Massey, Walton, Cranford, Zimmerman, & Cunningham, 2017) proposed a new design involving a different control group that was engaged in some assessment during the experimental period, and thus was more comparable with the experimental group. This new design was inspired by previous studies showing that (1) drinking behaviors assessed by daily IVR and weekly IVR in the same individuals were highly correlated (Tucker, Foushee, Black, & Roth, 2007); and (2) young adults complied well with a protocol of repeated weekend assessment that captured nearly prospective reports of the peak of substance use (Kuntsche & Robert, 2009). This recent study randomized participants into two groups: one group reported health risk behaviors *daily*, whereas the other group retrospectively reported about their behaviors in the previous 7 days on Sunday or Monday *weekly*. Comparing these groups regarding their trajectories of the target behavior may provide better scientific evidence for measurement reactivity, because this protocol allows researchers to make inferences based on the more prospective data collected *during* the experimental period rather than the conventional TLFB data collected *after* the experimental period.

Conventional analytical approaches adopted to study measurement reactivity have important limitations. Analysis of covariance (ANCOVA) can be conveniently implemented to test group differences in TLFB data collected after daily assessment, controlling for TLFB data collected before assessment. Yet, this method uses only retrospective data and summary measures (e.g., the average number of drinks consumed per day) which are known to leave out clinically meaningful information (Wang, Winchell, McCormick, Nevius, & O'Neill, 2002). Another important drawback of the ANCOVA approach is that it cannot characterize changes in behavior as a function of time. This drawback can be addressed by a more advanced conventional method, the generalized linear mixed model (GLMM), that employs a polynomial function of time (mostly a linear function). The GLMM approach, however, imposes a pre-specified shape on the trajectory of health risk behaviors that tend to fluctuate irregularly, especially in the beginning weeks of the self-monitoring process (Yang, Cranford, Li, & Buu, 2015).

Yang, Cranford, Li, Zucker, and Buu (2017) proposed a time-varying effect model (TVEM) to delineate the trajectories of distinct groups and conduct hypothesis testing for group differences. A major strength of the TVEM is that the trajectories are estimated through non-parametric regression functions that do not assume fixed shapes like the GLMM. This model, thus, provides a new lens to examine measurement reactivity by comparing the experimental and control groups in terms of their behavioral trajectories during the

experimental period. Furthermore, an extension of the TVEM (Yang, Cranford, Jester, Li, Zucker, & Buu, 2017) allows researchers to test whether measurement reactivity affects the likelihood of using substances or the amount of substance use, which is an important research question but has not been investigated simultaneously.

We aim to fill the current knowledge gaps by applying the TVEM to analyze daily diary data on alcohol and marijuana use collected from a randomized experiment (Buu et al., 2017), which randomized participants into daily and weekly assessment schedules. The daily group was hypothesized to report greater odds of abstinence and to report using a smaller amount of substances (once they were engaged in substance use) than the weekly group, because the daily group was engaged in a more intensive self-monitoring process. Group differences were also hypothesized to be only short-term because no feedback (e.g., graphical presentation of behavioral patterns over time) was provided.

Method

Study Sample and Procedures

The Measurement and Methodology (M&M) Study (Buu et al., 2017) is a randomized experiment that was designed to examine the psychometric properties of daily diary data such as measurement reactivity as a function of assessment schedules and methods. Study procedures were approved by the Institutional Review Board of University of Michigan (HUM00070757). Participants of the M&M Study were recruited by re-contacting a cohort of adolescents and emerging adults who enrolled in a previous observational study, the Flint Youth Injury (FYI) Study (Bohnert, Walton, Ranney, Bonar, Blow, Zimmerman, Booth, & Cunningham, 2015). The FYI Study recruited 600 youth (ages 14–24) when they sought care at the Hurley Medical Center Emergency Department (ED) in Flint, Michigan. The inclusion criterion was self-reported use of illicit drugs in the past year on a short computer survey (mostly marijuana). The exclusion criteria were: (1) youth who did not understand English; (2) youth deemed unable to provide informed assent/consent; and (3) prisoners at time of ED presentation.

Three hundred and seven participants aged 18–29 (mean=24) were recruited from the subject pool of the FYI study into the M&M Study during 2014–2016. They were randomized to four (2 × 2) assessment groups with different combinations of assessment schedules (daily or weekly) and assessment methods (IVR or short message service (SMS)). The resulting four groups did not differ on demographics and substance use at baseline (Buu et al., 2017). When IVR was assigned, the compliance rate for the weekly group was significantly higher than that for the daily group; such a group difference, however, did not exist between the two groups assessed by SMS (Buu et al., 2017). About 50% of the participants were male; 60% Black; 26% White; and 66% under public assistance. Based on self-report substance use in past 6 months at baseline, 69% used alcohol, 67% used nicotine, 73% used marijuana, 12% used other illicit drugs, and 18% misused prescription drugs.

At baseline, the participants self-administered a 30-min computerized assessment including demographic information and conventional measures of substance use related risk behaviors/problems in past six months. Afterwards, a 20–30 min staff-administered TLFB interview

was conducted to collect retrospective data on substance use related behaviors for each day in the past 90 days. Because substances are often used during late evening hours, participants in the *daily* groups reported daily by IVR/SMS about their behaviors on the previous day for 90 days, starting from the next day of the baseline assessment. The *weekly* groups retrospectively reported about their behaviors in the previous 7 days on every Sunday or Monday after the baseline. The daily and weekly groups both went through the experimental period of 90 days, after which another 90-day TLFB interview was conducted to collect retrospective data on relevant behaviors. We only include a brief description of the protocol that is directly related to the topic of this manuscript. Interested readers may refer to Buu et al. (2017) for other details of the M&M Study.

Because the present study focuses on measurement reactivity that only applies to those participants who were currently using the target substance, the statistical analysis on daily alcohol consumption only included 109 current alcohol users at baseline; and the analysis on marijuana use only included 153 current marijuana users. The current alcohol users were defined as those who reported using alcohol at least 2–4 times per month in past 6 months on the self-administered computerized assessment at baseline. The current marijuana users were defined as those who reported using marijuana at least once per week in past 6 months at baseline.

Daily Consumption Questions

The statistical analysis conducted in this study focused on the following questions (with response options) asked during the 90-day experimental period: “How many drinks containing alcohol did you have yesterday?” (0–60); “How many times did you use marijuana yesterday?” (0=“none”; 1=“once”; 2=“twice”; 3=“3–4 times”; 4=“5–6 times”; 5=“7–9 times”; 6=“10 or more times”). Unlike alcohol that can be quantified using the number of standard drinks, it is challenging to quantify marijuana with a short question in IVR/SMS assessment because it can be consumed in a variety of ways (e.g., joints, blunts, vaporizers etc.). Thus, we inquired about the frequency of marijuana use instead.

Analytic Approach

Although the experimental period of the M&M Study lasted for 90 days, we focused statistical analysis on data collected from the first 14 days because our previous study involving daily IVR data collection for 14 days from alcohol users showed that measurement reactivity was only observed in the first week (Yang et al., 2015). This strategy also avoided a potential confounding effect due to declining compliance rates in later weeks. Among the current alcohol users, the average compliance rate in the first 14 days was 0.81. The corresponding average compliance rate for the current marijuana users was 0.85.

In this study, we applied the TVEM proposed by Yang et al. (2017) to examine measurement reactivity by characterizing and testing the differences between the daily and weekly groups in alcohol and marijuana use trajectories. This method allowed us to model the effects of measurement reactivity on the probability of abstinence and the frequency/quantity of substance use simultaneously. The technical details below describe the models tested. Let $Y(t_{ij})$ be the j-th observed outcome from the i-th subject at time $t_{ij}(i = 1,\ldots,N; j = 1,\ldots,J_i)$ and

k be the group that Subject i belongs to ($k = 1$ for the daily group; $k = 2$ for the weekly group). The goal was to compare the difference in substance use trajectories between the daily (experimental) and weekly (control) groups, adjusting for two covariates: the assessment method ($X_1 = 1$ for IVR; $X_1 = 0$ for SMS) and the weekend effect ($X_2(t_{ij}) = 1$ for weekend; $X_2(t_{ij}) = 0$ for weekday).

The quantity of alcohol consumption with excess zeros was modeled as

$$Y(t_{ij}) = \begin{cases} 0 & \text{with probability } p_{ij} \\ Poisson & \text{with probability } (1 - p_{ij}). \end{cases}$$

The binary part (probability of abstinence) was modeled as

$$logit[p_{ij}] = \mu_1(t_{ij}) + \beta_1(t_{ij})I_{\{k=1\}} + \gamma_{11}X_1 + \gamma_{12}X_2(t_{ij}) + a_i;$$

The Poisson part (quantity when engaging in drinking) can be modeled by

$$log\left[\mathrm{E}\left(\mathrm{Y}\left(t_{ij}\right)\right)\right] = \mu_2(t_{ij}) + \beta_2(t_{ij})I_{\{k=1\}} + \gamma_{21}X_1 + \gamma_{22}X_2(t_{ij}) + b_i.$$

Here, $\mu_1(t_{ij})$ and $\mu_2(t_{ij})$ are the trajectories of the weekly group; $\beta_1(t_{ij})$ and $\beta_2(t_{ij})$ delineate the time-varying differences between the daily and weekly groups; γ_{11} to γ_{22} correspond to the time-invariant covariate effects; and a_i, b_i are random effects modeling within-subject correlation, which are assumed to be independent, and each of them follows a normal distribution with the mean 0 and an unknown variance parameter. A major strength of the TVEM is that it can model *time-varying* effects, which are assumed to change across time (such as the difference between the daily and weekly groups in drinking behaviors) as well as *time-invariant* effects, which are assumed to be constant at all time points (such as the effect of the assessment method) in the same model.

Furthermore, to analyze the data on frequency of marijuana use which was an ordinal scale with excess zeros, the above models can be modified as follows:

$$Y(t_{ij}) = \begin{cases} 0 & \text{with probability } p_{ij} \\ Proportional\ odds & \text{with probability } (1 - p_{ij}) \end{cases}$$

The binary part (probability of abstinence) remained the same, but the proportional odds part (frequency when engaging in marijuana use) was modeled as

$$logit\left[\Pr\left(Y\left(t_{ij}\right) > l\right)\right] = \mu_2(t_{ij}) + \beta_2(t_{ij})I_{\{k=1\}} + \gamma_{21}X_1 + \gamma_{22}X_2(t_{ij}) + b_i - \theta_l,$$

where $\theta_l = \theta_0 + ld\ (l = 1,\ldots,L-1)$; $\theta_1 < \ldots < \theta_{L-1}$. Here, the link function (i.e., the cumulative logits) modeled each ordinal response category with its own intercept: the

intercept term for the lowest ordinal category was $-\theta_0$; the intercept terms for higher ordinal categories were $-\theta_l = -(\theta_0 + ld)$, $l = 1,\ldots,5$.

SAS PROC NLMIXED was used to carried out the computation (the SAS program is included in the Appendix). Interested readers may refer to Yang et al. (2017) for detailed information about parameter estimation and hypothesis testing.

Results

Although the study sample was at high risk for substance use, descriptive statistics indicate that about 28% of the daily reports reflected abstinence from marijuana use, and the abstinence rate for alcohol use was 71% among all the daily reports. Thus, zero-inflation was observed in the study data and should be dealt with in the analysis. Furthermore, the odds of abstinence corresponding to the four randomization groups are: 3.55 (daily IVR), 2.40 (weekly IVR), 2.52 (daily SMS), and 1.64 (weekly SMS).

The TVEM analysis on daily alcohol consumption estimated both the time-invariant and time-varying parameters under the binary part (modeling the probability of abstinence) as well as the Poisson part (modeling the expected quantity). Figure 1 shows the trajectories of the difference between the daily and weekly groups. Panel (a) characterizes the time-varying ratio of the odds for abstinence in the daily group to the odds in the weekly group (calculated as exp (β_1)); and Panel (b) depicts the time-varying ratio of expected drinking quantity in the daily group to the expected quantity in the weekly group (calculated as exp (β_2)). When the criterion, 1, (the dotted line) falls out of the 95% confidence interval (the dash lines), the result indicates that the daily group was different from the weekly group. Panel (a) shows that the daily group had smaller odds of abstinence (the odds ratio <1) on Day 2 to Day 8 but such a group difference disappeared in the second week. Similarly, Panel (b) demonstrates that the daily group had lower expected drinking quantity only from Day 2 to Day 6. Table 1 shows the estimated time-invariant effects of the assessment method and weekend as well as the estimated variances of random effects under the binary and the Poisson parts of the TVEM. Based on this table, the assessment method did not have significant effects on either the probability of abstinence or expected drinking quantity, whereas the weekend was associated with not only smaller odds for abstinence [exp(−1.028)=0.36] but also higher quantity [exp(0.175)=1.19]. Moreover, the significant random effects reflect individual differences in alcohol consumption.

The time-varying and time-invariant effects of the TVEM on marijuana use are presented in Figure 2 and Table 2, respectively. In Figure 2, Panel (a) characterizes the time-varying ratio of the odds for abstinence in the daily group to the odds in the weekly group, whereas Panel (b) depicts the time-varying ratio of the odds for higher use frequency in the daily group to the odds in the weekly group. The 95% confidence intervals in both panels cover the criterion, 1, indicating that the daily and the weekly groups were not different in either the odds for abstinence or the tendency to report higher frequency of marijuana use. Table 2 lists the estimated time-invariant effects of the assessment method and weekend as well as the estimated variances of random effects under the binary and the proportional odds parts of the TVEM. According to Table 2, the IVR group had greater odds for abstinence

[exp(1.468)=4.34] and yet reported higher frequency [exp(1.653)=5.22] in those days involving marijuana use. The weekend did not have significant impact on the odds for abstinence or the frequency of marijuana use. The significant random effects again demonstrate individual differences in marijuana use.

Discussion

The results showed some evidence of short-term measurement reactivity in daily reports on alcohol use (defined as significant differences between the daily and weekly groups) during the first week of assessment, which remitted by the second week. Although the findings supported the hypothesis that the daily group would consume a lower level of alcohol once they were engaged in drinking, the findings contradicted the other hypothesis that the daily group would be more likely to abstain from alcohol. In terms of marijuana use, the daily and weekly groups did not differ on either the likelihood of abstinence or the consumption level. That is, the reactivity effect on marijuana use was not evident.

The TVEM method has important strengths that make it possible to conduct fine-grained analysis on daily diary data. First, it delineates the trajectory of behaviors with nonparametric functions determined by the data rather than imposes presumed parametric functions (e.g., linear). This approach is particularly useful for analyzing repeated measures collected intensively in a short period that are usually not represented well by a simple shape. Secondly, the TVEM method involving a mixture of zero and a Poisson model (or a mixture of zero and a proportional odds model) allows researchers to investigate the effect of measurement reactivity on not only the likelihood of using substances but also the amount of substance use (measured by quantity or frequency). This approach is particularly applicable to data collected from the substance abuse field that frequently contain excess zero’s (Buu, Li, Tan, & Zucker, 2012).

Although the TVEM model found short-term measurement reactivity on alcohol consumption during the first week, the reactivity effect on marijuana use was not significant. In fact, a recent study found that in general, alcohol use behaviors tend to fluctuate across days whereas the pattern of marijuana use is quite consistent over time (Liu, Li, Zimmerman, Walton, Cunningham, & Buu, 2019). Further, the probability of marijuana use on an ordinary day was estimated to be very high (>0.6) in comparison to that of alcohol use in the study sample (Buu, Li, Walton, Yang, Zimmerman, & Cunningham, 2014), suggesting that this more regular daily behavior may be more difficult to alter. Moreover, the unusually high odds ratio of abstinence from marijuana on the first day of data collection as demonstrated in Figure 2 Panel (a) resulted from the greater abstinence rate reported by the daily groups (41%) in comparison to the rate among the weekly group (14%). This group difference may be associated with the study protocol requiring that the daily group started reporting about their behavior from the next day of the baseline assessment, whereas the weekly group started on Sunday or Monday after the baseline. Because the state law during the data collection period only legalized medical marijuana, the participants in the daily group may become so sensitive in the beginning that they hesitated to admit their use on the day of the baseline assessment. This hypothesis, however, needs to be verified in a future study.

Based on the results of TVEM analysis on alcohol consumption, the daily group had smaller odds of abstinence but lower expected drinking quantity in the first week of assessment, which dissipated by the second week. Although the hypothesis of measurement reactivity predicts that self-monitoring of health risk behavior may reduce the behavior, the outcome measure adopted in the literature varies. Some researchers used a binary outcome indicating whether the participant was engaged in risk behavior in a day (e.g., Tucker et al., 2012). Other researchers measured the quantity of substance use in a day (e.g., Helzer et al., 2002). Thus, the inconsistent findings in the literature could be partly contributed to different outcome measures. The TVEM model, on the other hand, makes it possible to examine both types of outcomes simultaneously. Our findings imply that although daily monitoring of drinking may motivate people to reduce the quantity consumed once they start to drink, it may also arouse their desire to start drinking. Yet, both of these effects lasted only one week, as participants accommodated to the monitoring by the second week.

The finding that measurement reactivity was either absent (marijuana use) or short-term (alcohol consumption) may be disappointing to researchers who are interested in eliciting behavior changes through daily self-monitoring. Possible explanations of this result include that (1) the intensity of assessment was relatively low; and (2) feedback was not provided. For example, ecological momentary assessment (EMA) involving multiple self-reports a day may be a better alternative to foster behavior changes. In fact, McCarthy, Minami, Yeh, and Bold (2015) randomized adult daily smokers into low-frequency (once) or high-frequency (6 times) daily EMA for 4 weeks, and found that higher frequency EMA was associated with lower craving, anxiety, anger, hunger and positive affect, although it was not associated with abstinence.

Furthermore, feedback is another important factor for promoting behavior changes that is missing in most daily diary studies including the present study. Newcomb, Swann, Mohr, and Mustanski (2018) randomized a sample of young men who have sex with men to receive daily diaries, weekly diaries, or no diaries for 2 months. Half of the diary participants were also randomized to receive automated weekly feedback that provided graphical presentation of the risk behavior patterns but no concrete behavior change strategies (i.e., simulating most publicly available online or mobile app programs). The statistical analysis modeled the day-to-day changes on alcohol use, binge drinking, marijuana use, drug use, and condomless anal sex using an intercept and a slope for each behavior outcome. The assessment schedule only affected the slope of condomless anal sex but not the slopes of any of the substance use outcomes. Feedback, however, did not have significant effects on the slopes of any of the risk behaviors. Thus, even supplementing daily diaries with automated feedback on individual's risk behavior patterns may not be sufficient for creating meaningful behavior change. Indeed, it may be necessary to build just-in-time adaptive interventions (JITAIs) into daily assessments to address alternative motives for use (e.g., mood, social influences) in order to produce and sustain behavior change (Nahum-Shani et al., 2018).

Our study has some limitations that warrant attention. First, the findings were based on a community sample of drug users with high proportion of minorities and thus require validation with other samples, particularly lower risk samples. Although the proposed TVEM method would still be applicable to lower risk samples because we expect the zero-

inflation would be even greater among them, it is likely that they may react to daily assessment of substance use differently. Second, although the average compliance was high across experimental groups in the first two weeks (Buu et al., 2017), missing assessment over time (reflected in wider confidence intervals in Week 2) raises concerns about the validity of the results based on available data points from all the current substance users. Yet, our sensitivity analysis with only the users who had over 80% compliance rates (finding similar results) suggests that this may not be a significant problem in our study. Third, our finding that the daily group had smaller odds of abstinence from drinking but lower expected drinking quantity in the first week of assessment could potentially reflect different degrees of recall bias instead of different drinking behaviors between the daily and weekly groups. Although the literature showed that retrospective reports tend to underestimate alcohol consumption in comparison to prospective reports (Collins, Graham, Hansen, & Johnson, 1985; Searles, Helzer, Rose, & Badger, 2002), this recall bias was revealed by longer term retrospective reports with the range of 30 days to 2.5 years. Conversely, the weekly assessment involved short term retrospective reports that were shown to be highly correlated with daily reports (Tucker, Foushee, Black, & Roth, 2007). Nevertheless, objective measurement (e.g., passive sensors) may be used in future studies to confirm whether the short-term measurement reactivity observed in this study reflects differences in real behaviors or reporting patterns. Fourth, the observed difference in measurement reactivity between alcohol and marijuana use could possibly reflect the different aspects of consumption asked during the experimental period (number of drinks for alcohol vs. number of times for marijuana). Future studies are needed to overcome the challenge of quantifying marijuana use using IVR/SMS assessment (Buu, Hu, Pampati, Arterberry, & Lin, 2017).

In spite of the above limitations, our study makes unique contributions to the addiction literature in several ways. This was the first study to report results using the TVEM method to examine measurement reactivity in daily diary data in a randomized study. This new TVEM method allowed us to characterize the trajectories of group differences in not only the likelihood of using substances but also the amount of substance use. The results indicate that although daily self-monitoring could have short-term measurement reactivity on substance use behaviors that tend to fluctuate across days such as alcohol consumption, it does not affect substance use behaviors that are quite consistent such as marijuana use. For future daily diary studies designed to investigate the association between alcohol use behavior and its precursor or consequence, researchers may use the data collected in the first week with caution as it may be especially susceptible to measurement reactivity. Conversely, our results also suggest that daily self-monitoring may not be sufficient to promote sustainable behavior changes. In fact, the literature has suggested better alternative approaches such as increasing the intensity of assessment beyond once per day or providing individualized feedback with behavior change strategies.

Appendix

```
PROC NLMIXED DATA=final COV;
       PARAMETER  a1=0 a2=0 a3=0 a4=0 a5=0
                  b1=0 b2=0 b3=0 b4=0 b5=0
```

```
                    c1=0 c2=0 c3=0 c4=0 c5=0
                    d1=0 d2=0 d3=0 d4=0 d5=0
                     e1= 0 e2 =0
                    f1=0 f2=0
                    g1=0 g2=0
                    d = 1
                    c = 0
                VarBinary=1 Varordinal=1 ;
        bounds d>0;
        /* Define our TVEM model: */
        linkp = random_binary +
a1*surveydayr_0+a2*surveydayr_1+a3*surveydayr_2+a4*surveydayr_3+a5*surveydayr
_4 + /* basis
functions for mu_{female}(t) in binary part */

c1*assessmentschedule*surveydayr_0+c2*assessmentschedule*surveydayr_1+c3*asse
ssmentschedule
*surveydayr_2+c4*assessmentschedule*surveydayr_3+c5*assessmentschedule*survey
dayr_4
+ e1*assessmenttype +f1*weekd+g1*assessmentschedule*assessmenttype/* basis
functions for
beta_{male}(t) in binary part */;
                        /* Logit probability of being a nondrinker */
        p0=exp(linkp)/(1+exp(linkp));  /* Probability of being a nondrinker */
        mu=c+random_ordinal +
b1*surveydayr_0+b2*surveydayr_1+b3*surveydayr_2+b4*surveydayr_3+b5*surveydayr
_4 + /* basis
functions for mu_{female}(t) in Poisson part */

d1*assessmentschedule*surveydayr_0+d2*assessmentschedule*surveydayr_1+d3*asse
ssmentschedule
*surveydayr_2+d4*assessmentschedule*surveydayr_3+d5*assessmentschedule*survey
dayr_4
+ e2*assessmenttype +f2*weekd+g2*assessmentschedule*assessmenttype; /* basis
functions for
beta_{male}(t) in Poisson part */
         /* Define the ZIP distribution: */

    IF pot1d=0 THEN
              11=log(p0+(1-p0)*(1–1/(1+exp(−5*d - mu))));


       ELSE if (pot1d =1) then 11=log((1-p0))+ log(1/(1+exp(− mu)));
       ELSE if (pot1d =2) then 11=log((1-p0))+ log(1/(1+exp(−d - mu)) - 1/
```

```
(1+exp(− mu)));
       ELSE if (pot1d =3) then 11=log((1-p0))+ log(1/(1+exp(−2*d - mu))-1/
(1+exp(−d - mu)));
       ELSE if (pot1d =4) then 11=log((1-p0))+ log(1/(1+exp(−3*d - mu))-1/
(1+exp(−2*d - mu)));
       ELSE if (pot1d =5) then 11=log((1-p0))+ log(1/(1+exp(−4*d - mu))-1/
(1+exp(−3*d - mu)));
       ELSE if (pot1d =6) then 11=log((1-p0))+ log(1/(1+exp(−5*d - mu))-1/
(1+exp(−4*d - mu)));

    MODEL pot1d~GENERAL(11);
    ODS OUTPUT ParameterEstimates=MyParams FitStatistics=MyFitStatistics
CovMatParmEst=MyCov;
       RANDOM random binary random ordinal ~ NORMAL([0,0], [VarBinary,
0,Varordinal])
SUBJECT=subjid;
RUN;
```

References

Agresti A (2013). Categorical data analysis, 3rd edition. Hoboken, NJ: John Wiley & Sons.

Bandura A (1991). Social cognitive theory of self-regulation. Organizational Behavior and Human Decision Processes, 50, 248–287. 10.1016/0749-5978(91)90022-L

Bohnert KM, Walton MA, Ranney M, Bonar EE, Blow FC, Zimmerman MA, Booth BM, Cunningham RM (2015). Understanding the service needs of assault-injured, drug-using youth presenting for care in an urban Emergency Department. Addictive Behaviors, 41, 97–105. 10.1016/j.addbeh.2014.09.019 [PubMed: 25452051]

Buu A, Hu Y, Pampati S, Arterberry BJ, & Lin H (2017). Predictive validity of cannabis consumption measures: results from a national longitudinal study. Addictive Behaviors, 73, 36–40. [PubMed: 28463803]

Buu A, Li R, Tan X, & Zucker RA (2012). Statistical models for longitudinal zero-inflated count data with applications to the substance abuse field. Statistics in Medicine, 31, 4074–4086. 10.1002/sim.5510 [PubMed: 22826194]

Buu A, Li R, Walton MA, Yang H, Zimmerman MA, Cunningham RM (2014). Changes in Substance Use-Related Health Risk Behaviors on the Timeline Follow-Back Interview as a Function of Length of Recall Period. Substance Abuse and Misuse, 49(10), 1259–1269. 10.3109/10826084.2014.891621

Buu A, Massey LS, Walton MA, Cranford JA, Zimmerman MA, & Cunningham RM (2017). Assessment methods and schedules for collecting daily process data on substance use related health behaviors: a randomized control study. Drug and Alcohol Dependence, 178, 159–164. 10.1016/j.drugalcdep.2017.05.003 [PubMed: 28651151]

Collins LM, Graham JW, Hansen WB, & Johnson CA (1985). Agreement between retrospective accounts of substance use and earlier reported substance use. Applied Psychological Measurement, 9, 301–309.

French DP & Sutton S (2010). Reactivity of measurement in health psychology: how much of a problem is it? What can be done about it? British Journal of Health Psychology, 15, 453–468. 10.1348/135910710X492341 [PubMed: 20205982]

Gunthert KC, & Wenze SJ (2012). Daily diary methods In Mehl MR & Conner TS (Eds.), Handbook of research methods for studying daily life (pp. 144–159). New York: Guilford Press.

Helzer JE, Badger GJ, Rose GL, Mongeon JA, & Searles JS (2002). Decline in alcohol consumption during two years of daily reporting. Journal of Studies on Alcohol, 63, 551–558. 10.15288/jsa.2002.63.551 [PubMed: 12380851]

Helzer JE, Rose GL, Badger GJ, Searles JS, Thomas CS, Lindberg SA, Guth S (2008). Using interactive voice response to enhance brief alcohol intervention in primary care settings. Journal of Studies on Alcohol and Drugs, 69, 251–258. 10.15288/jsad.2008.69.251 [PubMed: 18299766]

Kuntsche E & Robert B (2009). Short Message Service (SMS) technology in alcohol research—A feasibility study. Alcohol and Alcoholism, 44, 423–248. 10.1093/alcalc/agp033 [PubMed: 19482879]

Liu W, Li R, Zimmerman MA, Walton MA, Cunningham RM, & Buu A (2019). Statistical methods for evaluating the correlation between timeline follow-back data and daily process data with applications to research on alcohol and marijuana use. Addictive Behaviors, 94, 147–155. 10.1016/j.addbeh.2018.12.024 [PubMed: 30611576]

Mankarious E & Kothe E (2015). A meta-analysis of the effects of measuring theory of planned behavior constructs on behavior within prospective studies. Health Psychology Review, 9, 190–204. 10.1080/17437199.2014.927722 [PubMed: 26209208]

McCarthy DE, Minami H, Yeh VM, & Bold KW (2015). An experimental investigation of reactivity to ecological momentary assessment frequency among adults trying to quit smoking. Addiction, 110, 1549–1560. 10.1111/add.12996 [PubMed: 26011583]

Mitchum AL, Kelley CM & Fox MC (2016). When asking the question changes the ultimate answer: metamemory judgements change memory. Journal of Experimental Psychology: General, 145, 200–219. 10.1037/a0039923 [PubMed: 27045282]

Nahum-Shani I, Smith SN, Spring BJ, Collins LM, Witkiewitz K, Tewari A, & Murphy SA (2018). Just-in-time adaptive interventions (JITAIs) in mobile health: key components and design principles for ongoing health behavior support. Annals of Behavioral Medicine, 52, 446–462. 10.1007/s12160-016-9830-8 [PubMed: 27663578]

Newcomb ME, Swann G, Mohr D, & Mustanski B (2018). Do diary studies cause behavior change? An examination of reactivity in sexual risk and substance use in young men who have sex with men. AIDS and Behavior, 22, 2284–2295. 10.1007/s10461-018-2027-3 [PubMed: 29332235]

Reis HT (2012). Why researchers should think “real-world”: A conceptual rationale In Mehl MR & Conner TS (Eds.), Handbook of research methods for studying daily life (pp. 321). New York: Guilford Press.

Reynolds BM, Robles TF, & Repetti RL (2016). Measurement reactivity and fatigue effects in daily diary research with families. Developmental Psychology, 52, 442–456. 10.1037/dev0000081 [PubMed: 26689757]

Schroder KEE, Johnson CJ, & Wiebe JS, (2007) Interactive Voice Response technology applied to sexual behavior self-reports: A comparison of three methods. AIDS and Behavior, 11, 313–323. 10.1007/s10461-006-9145-z [PubMed: 16841191]

Searles JS, Helzer JE, Rose GL, & Badger GJ (2002). 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, 63, 352–362. [PubMed: 12086136]

Simpson TL, Kivlahan DR, Bush KR, & McFall ME (2005). Telephone self-monitoring among alcohol use disorder patients in early recovery: A randomized study of feasibility and measurement reactivity. Drug and Alcohol Dependence, 79, 241–250. 10.1016/j.drugalcdep.2005.02.001 [PubMed: 16002033]

Sobell LC & Sobell MB (1992). Timeline Follow-back: A technique for assessing self-reported ethanol consumption In Allen J & Litten RZ (Eds.), Measuring alcohol consumption: Psychosocial and biological methods (pp. 41–72). Totowa, NJ: Humana Press.

Tan X, Shiyko M, Li R, Li Y & Dierker L (2012). A time-varying effect model for intensive longitudinal data. Psychological Methods, 17, 61–77. 10.1037/a0025814 [PubMed: 22103434]

Tucker JA, Blum ER, Xie L, Roth DL, & Simpson CA (2012). Interactive voice response self-monitoring to assess risk behaviors in rural substance users living with HIV/AIDS. AIDS and Behavior, 16, 432–440. 10.1007/s10461-011-9889y [PubMed: 21311964]

Tucker JA, Foushee HR, Black BC, & Roth DL (2007) 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, 68, 538–542. 10.15288/jsad.2007.68.538 [PubMed: 17568958]

Wang SJ, Winchell CJ, McCormick CG, Nevius SE, & O'Neill RT (2002). Short of complete abstinence: An analysis exploration of multiple drinking episodes in alcoholism treatment trials. Alcoholism: Clinical and Experimental Research, 26, 1803–1809. 10.1097/01.ALC.0000042009.07691.12

Yang S, Cranford JA, Jester JM, Li R, Zucker RA, & Buu A (2017). A time-varying effect model for examining group differences in trajectories of zero-inflated count outcomes with applications in substance abuse research. Statistics in Medicine, 36, 827–837. 10.1097/10.1002/sim.7177 [PubMed: 27873343]

Yang H, Cranford JA, Li R, & Buu A (2015). Two-stage model for time-varying effects of discrete longitudinal covariates with applications in analysis of daily process data. Statistics in Medicine, 34, 571–581. 10.1002/sim.6368 [PubMed: 25395042]

Yang S, Cranford JA, Li R, Zucker RA, & Buu A (2017). A time-varying effect model for studying gender differences in health behavior. Statistical Methods in Medical Research, 26, 2812–2820. 10.1177/0962280215610608 [PubMed: 26475829]

Highlights

- Time-varying effect modeling is a useful method to study measurement reactivity.
- Daily monitoring of drinking only had short-term reactivity in the first week.
- Daily monitoring of marijuana use did not have significant measurement reactivity.

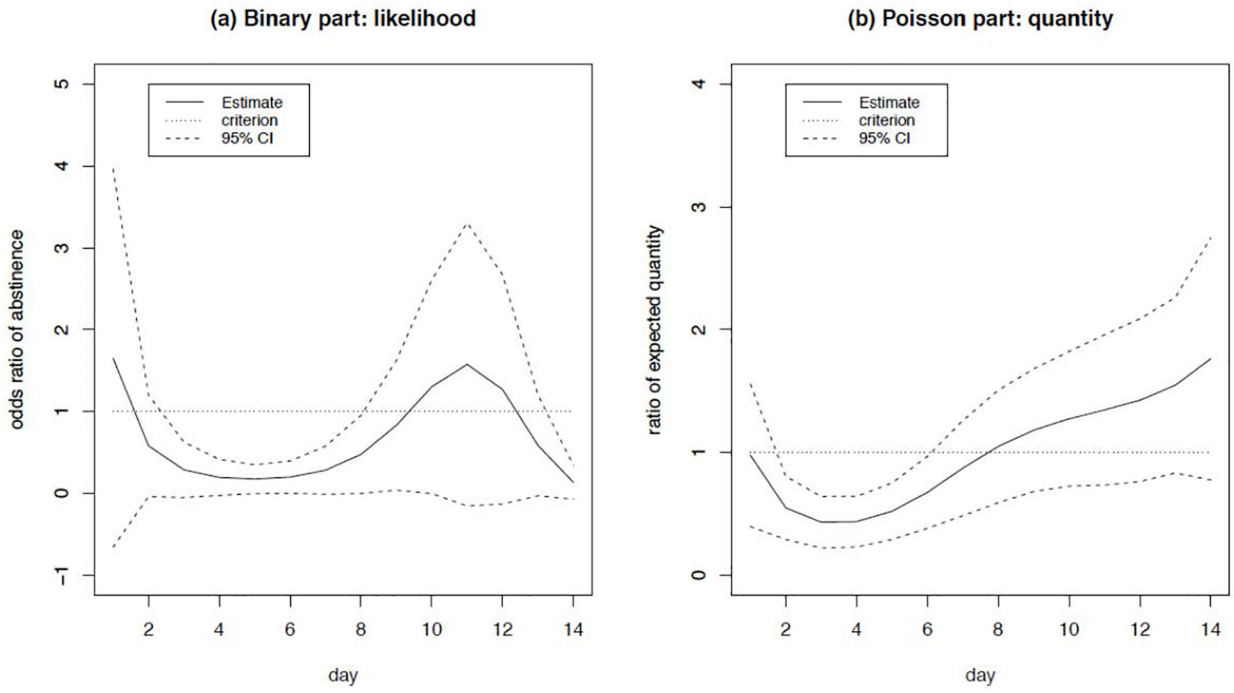

Figure 1.

The time-varying difference between daily and weekly groups in alcohol use likelihood and quantity (the weekly group is the reference group).

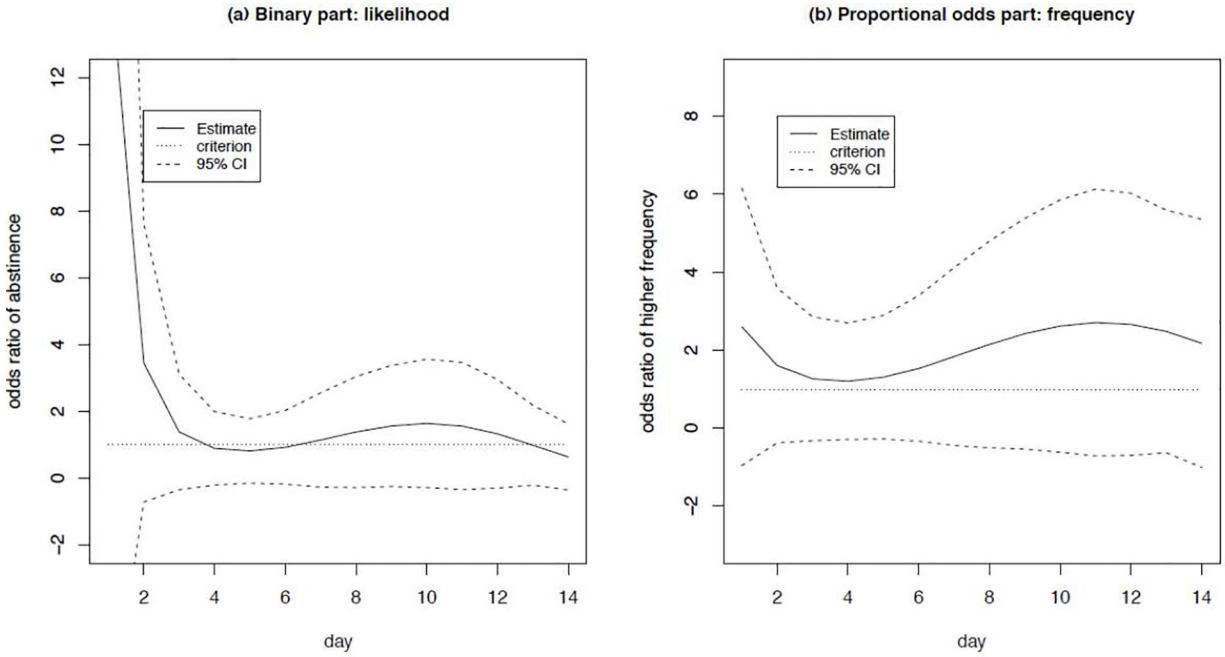

Figure 2.

The time-varying difference between daily and weekly groups in marijuana use likelihood and frequency (the weekly group is the reference group).

Table 1

The time-invariant effects in the time-varying effect model (TVEM) for alcohol consumption based on diary data.

	Coefficient	Standard error	p-value
Binary part: modeling probability of abstinence			
IVR (1) vs. SMS (0): γ_{11}	−0.665	0.410	0.108
Weekend (1) vs. weekday (0): γ_{12}	−1.028	0.233	<.001
Random effect: a_i	3.234	0.768	<.001
Poisson part: modeling quantity when engaging in drinking			
IVR (1) vs. SMS (0): γ_{21}	−0.273	0.201	0.177
Weekend (1) vs. weekday (0): γ_{22}	0.175	0.073	0.018
Random effect: b_i	0.642	0.133	<.001

Table 2

The time-invariant effects in the time-varying effect model (TVEM) for marijuana use based on diary data.

	Coefficient	Standard error	p-value
Binary part: modeling probability of abstinence			
IVR (1) vs. SMS (0): γ_{11}	1.468	0.560	0.010
Weekend (1) vs. weekday (0): γ_{12}	−0.143	0.202	0.481
Random effect: a_i	8.666	1.710	<.001
Proportional odds part: modeling frequency when engaging in marijuana use			
Intercept of weekly group:			
the baseline: θ_0 [a]	−2.964	0.429	<.001
the increment: d [a]	2.223	0.067	<.001
IVR (1) vs. SMS (0): γ_{21}	1.653	0.601	0.007
Weekend (1) vs. weekday (0): γ_{22}	−0.149	0.142	0.297
Random effect: b_i	11.767	1.792	<.001

[a] The intercept term for the lowest ordinal category: $-\theta_0$; the intercept terms for higher ordinal categories: $-\theta_l = -(\theta_0 + ld)$, $l = 1,...,5$.