

# Assessment of Alcohol Use and Consequences: Comparison of Data Collected Via Timeline Followback Interview and Daily Reports

JENNIFER E. MERRILL, PH.D.,<sup>a,\*</sup> PENGYANG FAN, M.P.H.,<sup>a</sup> TYLER B. WRAY, PH.D.,<sup>a</sup> & ROBERT MIRANDA JR., PH.D.<sup>a,b</sup>

<sup>a</sup>Center for Alcohol and Addiction Studies, Department of Behavioral and Social Sciences, Brown University School of Public Health, Providence, Rhode Island

<sup>b</sup>Department of Psychiatry and Human Behavior, Brown University, Providence, Rhode Island

**ABSTRACT. Objective:** The aim of this study was to compare data on both alcohol use and alcohol-related consequences between intensive longitudinal data collection and the retrospective Timeline Followback (TLFB) interview. **Method:** Heavy drinking college students ( $n = 96$ ; 52% women) completed daily reports across a 28-day period to assess alcohol use and positive and negative consequences of drinking. They returned to the lab at the end of this period to complete a TLFB assessing behavior over those same 28 days. First,  $t$  tests were used to compare variables aggregated across the full 28 days at the between-person level. Next, hierarchical linear modeling was used to examine within-person differences between methods for each variable in weekly and daily incre-

ments. **Results:** Many alcohol use and consequence variables were significantly different when derived from self-reports during TLFB versus daily reports. In contrast to prior work, we found that higher estimates of drinking were reported retrospectively on the TLFB than on the daily reports. In addition, discrepancies were greater on some variables for heavier drinkers and when more time had elapsed between the end of the daily reporting period and TLFB collection. **Conclusions:** Recall of drinking behavior during TLFB and daily reports may differ in systematic ways, with discrepancies varying based on participant and methodological characteristics. (*J. Stud. Alcohol Drugs*, 81, 212–219, 2020)

**H**EAVY DRINKING IS RELATED TO several longer and shorter term harms to self and others (Centers for Disease Control and Prevention, 2008; Holmes et al., 2016). Accurately assessing consumption and its consequences is essential for understanding and treating alcohol misuse. Research to measure alcohol use at the level of an individual, to understand its predictors and consequences for example, typically relies on either retrospective reports or intensive longitudinal assessments. Yet, there are important differences between these measures that could influence conclusions drawn. How these methods compare, particularly in assessment of alcohol-related consequences, deserves more attention.

The current gold standard for retrospective reporting of alcohol use is the Timeline Followback (TLFB; Sobell et al., 2003), a calendar-based interview about the number of standard drinks one consumed on each day over some time-frame (e.g., past month). The TLFB can be modified to assess outcomes other than alcohol use, such as consequences of drinking (Merrill et al., 2013). However, as a retrospective assessment, the TLFB may result in recall bias, such that

participants over- or underestimate their drinking behavior upon encountering difficulty remembering actual events (Hufford et al., 2002; Toll et al., 2006).

Intensive longitudinal measurements, collected in temporal proximity to actual drinking behavior and in the natural environment (e.g., daily diary and/or ecological momentary assessment), can minimize such biases (Dulin et al., 2017; Wray et al., 2014). A growing number of studies indicate that intensive assessments result in improved measurement validity when compared with traditional, retrospective methods (Dulin et al., 2017; Simons et al., 2015).

Although some studies show no significant difference between data derived from intensive longitudinal assessments (e.g., daily diaries) versus retrospective measures such as the TLFB (Chow et al., 2017), others show significantly lower alcohol consumption reports on retrospective assessments (Dulin et al., 2017; Monk et al., 2015; Patterson et al., 2019). Further, certain contextual and individual difference variables help explain differences across methods. A greater time interval between the end of intensive assessment and a retrospective assessment (e.g., TLFB) may increase the discrepancy (Dulin et al., 2017; Hoepfner et al., 2010), although others showed this had no effect (Rowe et al., 2016). Concordance between methods might also differ depending on aspects of the environmental context, such as drinking location or number of friends present (Monk et al., 2015).

Although prior studies provide valuable insights into how intensive longitudinal and retrospective methods compare, additional work is needed. Much of the prior research relied

---

Received: August 27, 2019. Revision: December 18, 2019.

This study was supported by National Institute on Alcohol Abuse and Alcoholism Grant K01AA022938 (to Jennifer E. Merrill).

\*Correspondence may be sent to Jennifer E. Merrill at the Department of Behavioral and Social Sciences, Center for Alcohol and Addiction Studies, Brown University, Box G-S121-5, Providence, RI 02912, or via email at: Jennifer\_Merrill@brown.edu.

on sample sizes of 50 or fewer (e.g., Dulin et al., 2017; Wray et al., 2019), or only examined aggregate-level variables (e.g., total drinking over 1 month) and not finer-grained, daily data (Chow et al., 2017; Poulton et al., 2018). It is important to note that no studies have compared data on alcohol consequences collected via both methods.

Drinking is associated with many consequences, both positive (e.g., making others laugh) and negative (e.g., nausea). In general, ability to accurately recall episodic memories declines with time. As such, we would expect reports of alcohol consequences to be less accurate when collected via the retrospective TLFB than intensive longitudinal assessment. It is possible, however, that the extent of discrepancies between methods differs for positive versus negative consequences. Emotional valence and arousal levels influence memory, and recall may be better for negative than for positive events (Earles et al., 2016; Holland & Kensinger, 2013; Kensinger, 2009; Mackay et al., 2004). Accurate measurement of drinking consequences is essential for answering many important research questions, such as whether consequences prompt naturalistic change in drinking behavior, or whether they are effectively reduced following intervention.

### *Present study*

Our aim was to compare data on alcohol use, positive consequences, and negative consequences collected via intensive longitudinal measurements (smartphone-based daily diaries) versus the TLFB. We tested differences between methods on variables aggregated across the full study time frame (28 days), as well as within-person differences (weekly, daily). In line with prior research, we hypothesized that self-reports of drinking behavior would be higher when measured via daily assessments versus TLFB, and greater discrepancies between the two methods would be observed as recall time increased and among heavier drinkers.

## **Method**

### *Participants*

A total of 101 heavy drinking college students enrolled; 100 completed the protocol. Eligibility criteria included ages 18–20, enrollment in a local 4-year college/university, and either (a) engaging in weekly heavy episodic drinking (HED; 4+ [women]/ 5+ [men] drinks in a single sitting) or (b) experiencing at least 1 (of 10) negative alcohol-related consequences in the past 2 weeks. Participants were excluded for past-2-week illicit drug use (other than marijuana), current participation in treatment for a substance use disorder, or no access to a smartphone data plan. Four participants who reported no drinking during daily reports were excluded, leaving an analytic sample of 96.

### *Procedure*

Procedures were approved by the university's institutional review board. Daily surveys were created using Metricwire Inc. software (<https://metricwire.com>). This allowed us to create custom surveys and a schedule of notifications and reminders (described below), which participants received upon downloading the Metricwire mobile application onto their phone and being enrolled into our study. On completion of any given survey, data are timestamped and automatically uploaded to a server.

Participants were recruited via social media advertisements and flyers. Interested participants completed an online screener. If eligible, they were routed to an online consent form and baseline questionnaire assessing demographics and drinking behavior. A 7-day grid similar to the Daily Drinking Questionnaire (Collins et al., 1985) was used to obtain number of drinks per typical week in the past 30 days. Single items assessed past-30-day drinks per typical/heaviest drinking day and number of HED days. The Brief Young Adult Alcohol Consequences Questionnaire (Kahler et al., 2005) assessed how many of 24 possible negative consequences resulted from drinking (past 30 days).

Next, participants attended an in-person group orientation to consent to and learn more about the remaining study procedures. Participants downloaded the Metricwire Inc. mobile data collection application onto their smartphones and completed practice reports. Participants were trained in the counting of standard drinks (12 oz. of beer, 5 oz. of wine, 1.5 oz. of distilled spirits). An image depicting these standard drink sizes accompanied questions about alcohol use in both the intensive assessment and TLFB interview. Participants were instructed to complete daily surveys via the mobile application each morning for 28 days. They then returned to the lab and completed the TLFB interview for the same 28-day period. Participants received \$25 for the baseline survey/orientation and \$30 for the interview. For longitudinal assessment, participants were paid based on percent compliance, earning from \$5 (<20% compliance) to \$45 (>90%) on Week 1. Potential payments increased slightly each week, to \$51 maximum.

*Intensive longitudinal assessment.* Participants were instructed to respond to daily diary reports via the mobile application as soon as possible after waking. Push notification reminders were sent via the application at 7 A.M. and 9 A.M., and the survey expired at midnight. Average time of completion was 10:39 A.M., and 78% were submitted before noon. In these surveys, participants indicated whether they had consumed alcohol “yesterday,” and if so, the number of standard drinks. During orientation, they received training that “yesterday” referred to a drinking event that could have extended across two calendar days (e.g., starting at 8 P.M., ending at 2 A.M.). They were also asked to report whether they had experienced any of eight negative consequences (nauseated/vomited, rude/

obnoxious, neglect school-related obligations, hurt or injured yourself by accident, behaved aggressively, embarrassed yourself, forgot what you did, hangover) and eight positive consequences (had something that normally would bother you fail to bother you, expressed feelings more easily, talked to someone probably wouldn't have otherwise, creative moment/experience, new friend/acquaintance, made others laugh, had something fun/exciting happen, and slept better). These consequences were selected based on prior research (Lee et al., 2017) and our formative work (Merrill et al., 2018)<sup>1</sup>.

*Timeline Followback.* The TLFB (Sobell & Sobell, 1992) assessment was collected an average of 9.84 days ( $SD = 6.71$ , range: 1–34) after the end of the 28 daily assessments. Participants' appointment books and/or smartphones (e.g., social media, photos, text messages) were used during the interview to aid recall. Importantly, information from the daily reports was not used to facilitate recall of drinking or alcohol-related consequences over the 28-day period. In addition to reporting drinks per day, participants were given a numbered list of the same positive and negative alcohol consequences assessed daily, and asked to report any experienced each day.

### Statistical analysis

For both TLFB and intensive assessments, we created several matched variables reflecting alcohol use over the 28-day period: number of drinking days, number of HED days, total drinks, and drinks per drinking day. We also created several consequence variables: total positive and negative consequences and number of positive and negative consequences per drinking day. We created similar variables reflecting the total of each behavior/event each week, and day-level variables reflecting whether each day was a drinking day, whether it was an HED day, total drinks, total number of positive and negative consequences, and whether participants experienced any positive or negative consequences that day. We compared reports across the two methods in aggregate (i.e., summed across the 28 days) using  $t$  tests and Cohen's  $d$  effect sizes.

Next, we explored within-person differences across the two methods at weekly and daily levels using several hierarchical linear models (HLMs) in the HLM 7.02 program (Raudenbush et al., 2013). Models (one for each variable described above) were three levels (report type [TLFB vs. daily] nested within week/day nested within subjects). All weekly and most daily variables were normally distributed, so we specified Gaussian distributions. At the daily level,

<sup>1</sup>“Driving after having too much to drink” was also assessed, although never endorsed in this sample, and “Had an alcohol-facilitated romantic/sexual experience” was assessed but not included in the present study, given that it could have been perceived as either a negative or positive consequence. Sensitivity analyses with the addition of this item as a positive consequence were conducted, and model findings were similar.

binomial distributions were specified for dichotomous outcomes. Also at the daily level, number of negative consequences per day resembled a count distribution; therefore, a Poisson distribution was specified.

In all models, the predictor of interest was report type (0 = *daily*, 1 = *TLFB*). As such, model intercepts represented the average value of the outcome reported in the daily survey, controlling for covariates, and the slope effect of “report type” represented the difference in the outcome between the daily and TLFB report. Covariates included person-level (Level 3) drinks per typical week (baseline) and days between the end of the intensive longitudinal assessment and the TLFB. At Level 2, we controlled for the day/week of study. To evaluate moderation hypotheses, we specified cross-level interactions between report type and (a) total drinks per typical week reported at baseline (Level 3); (b) number of days between the end of daily assessment and in-person TLFB (Level 3); and (c) day/week of study, with 1 representing days/weeks more distal to the TLFB (Level 2). For parsimony, models without interaction terms are tabled, with any observed interactions described in the text. We assumed an unstructured covariance matrix. Intercepts were random; slope effects were tested and retained in the model only when significant. We used a criterion of  $p < .01$  for evaluating significance of both fixed and random effects.

## Results

### Descriptives

Participants ( $n = 96$ ) were 52% female, 80% first-year students, 72% White, 15% Hispanic, and on average age 18.7 ( $SD = 0.7$ ). At baseline, for the prior 30 days, they reported 10.48 ( $SD = 6.34$ ) drinks per typical week, 4.99 ( $SD = 2.20$ ) per typical drinking day, 7.35 ( $SD = 3.24$ ) on the heaviest day, 3.10 ( $SD = 2.27$ ) HED days, and 3.82 ( $SD = 3.20$ ) negative consequences. Almost all (98.7%, 2,653) daily reports were submitted.

### Aggregate variable (between-subjects) comparisons

Compared with TLFB, means derived from daily reports were significantly lower for total drinks, number of drinking days, and positive consequences (total across 28 days and average per drinking day) across the 4-week interval (Table 1). Effect sizes were moderate ( $d = 0.2$ – $0.5$ ). There were no statistically significant differences in mean number of HED days, drinks per drinking day, or negative consequences (total across 28 days or average per drinking day).

### Within-subjects comparisons: Weekly variables

For all outcomes except HED days and drinks per drinking day, reports were higher on the TLFB than on daily diaries,

TABLE 1. Aggregate (across all 28 days) means, standard deviations, and *t* tests between methods

Variable	TLFB		Daily assessment		<i>t</i> test <i>p</i>	Effect size (Cohen's <i>d</i> )
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		
Number of drinking days	5.90	2.89	5.07	2.65	.000**	0.30
Number of heavy drinking days	3.55	2.75	3.28	2.53	.153	0.10
Total number of drinks across the study	31.20	22.97	26.51	20.21	.001**	0.22
Average drinks per drinking day	5.01	2.07	4.97	2.15	.808	0.02
Total number of positive consequence experiences	17.06	10.47	12.60	9.27	.000**	0.45
Total number of negative consequence experiences	4.33	4.39	4.00	3.95	.345	0.08
Positive consequence per drinking day	2.98	1.30	2.45	1.39	.000**	0.39
Negative consequence per drinking day	0.72	0.66	0.76	0.72	.566	-0.05

Note: TLFB = Timeline Followback.

\*\**p* < .01.

when aggregated to the week level (Table 2). In a subsequent moderation model predicting HED days, a significant interaction indicated that as number of days between the two assessments increased, the degree of discrepancy between HED reported on the TLFB versus daily reports also increased ( $B = 0.02$ ,  $SE = 0.001$ ,  $p = .004$ ). A second significant interaction suggested that the discrepancy in HED between reports increased as participants' baseline drinking increased ( $B = 0.02$ ,  $SE = 0.01$ ,  $p = .001$ ). When predicting the number of positive consequences, a significant interaction indicated that the discrepancy between the two reporting methods increased in later study weeks ( $B = 0.34$ ,  $SE = 0.13$ ,  $p = .009$ ).

#### *Within-subjects comparisons: Daily variables*

Reports from TLFB were higher than those from daily diaries at the day level for number of drinks, number of positive consequences, and drinking day but not HED day, any negative consequence, any positive consequence, or number of negative consequences (Table 3). In a subsequent model of HED, a significant cross-level interaction suggested that the odds of discrepant reports across the two methods increased as days between the two assessments increased (odds ratio [OR] = 1.03, 95% CI [1.01, 1.05],  $p = .003$ ). A second significant cross-level interaction suggested that as level of baseline drinking increased, the extent to which the likelihood of HED was higher when reported on TLFB relative to daily report also increased (OR = 1.03, 95% CI [1.01, 1.04],  $p = .004$ ). For several outcomes, day in the study (Level 2) also moderated the effect of report type. As the daily assessment day was closer in time to the in-person TLFB interview, the extent to which outcomes were higher on TLFB than daily reports also increased for number of drinks ( $B = 0.03$ ,  $SE = 0.01$ ;  $t = 3.10$ ,  $p = .002$ ), number of positive consequences ( $B = 0.04$ ,  $SE = 0.01$ ;  $t = 3.53$ ,  $p < .001$ ), and likelihood of any positive consequence (OR = 1.07, 95% CI [1.03, 1.12],  $p < .001$ ).

## Discussion

Using two different methods—intensive longitudinal assessment and TLFB—over 28 days, we advanced prior

work by examining discrepancies not only in self-reports of alcohol use, but also negative and positive alcohol-related consequences. Overall, results suggest that recall of drinking behavior during TLFB and daily reports may differ, depending on the specific variable and level of analysis. However, contrary to our hypotheses, higher estimates of drinking were reported retrospectively on the TLFB than on daily reports.

Comparisons of person-level means on alcohol use indicated higher aggregate TLFB estimates of number of drinking days and total drinks, but not HED days or drinks per drinking day. For some variables, the magnitude and clinical significance of these differences was substantial—on average over a month, participants reported about five more drinks on the TLFB than daily diaries. There were also significantly higher estimates on the TLFB for positive consequences. On the other hand, negative consequence reports did not differ by assessment type. As such, researchers can feel confident that similar levels of negative (but perhaps not positive) consequences will be reported via either method, at least when aggregated across 28 days.

There are a few potential explanations for report type discrepancies in positive but not negative consequences. First, it is possible that, because of social desirability (Davis et al., 2010; Joinson, 1999), participants endorsed more positive effects of drinking on the TLFB to appear in a favorable light to the interviewer. Further, positive consequences of drinking (e.g., having fun) tend to be more common than negative consequences (Barnett et al., 2015). Accordingly, one possibility is that during the TLFB participants had a response bias toward endorsing the positive consequences more frequently than they were reported to occur in daily assessments. Another possibility is that the more rare negative consequences (e.g., injury, embarrassment) may be more distinct in one's memory and therefore more accurately reported in retrospect.

According to the accessibility model of emotional self-report (Robinson & Clore, 2002), some of the sources of information that an individual draws on when reporting retrospectively on emotion may result in bias. Although this model is specific to recall of emotions, such sources of information may also color one's recall for drinking conse-



TABLE 2. Correspondence between TLFB and daily reports at the weekly level

Fixed effects	Models predicting alcohol use outcomes															
	Drinking days				Heavy drinking days				Total drinks				Drinks per drinking day			
	<i>B</i>	<i>SE</i>	<i>t</i>	<i>p</i>	<i>B</i>	<i>SE</i>	<i>t</i>	<i>p</i>	<i>B</i>	<i>SE</i>	<i>t</i>	<i>p</i>	<i>B</i>	<i>SE</i>	<i>t</i>	<i>p</i>
Intercept	1.57	0.11	13.92	<.001	0.98	0.10	9.99	<.001	7.96	0.72	11.03	<.001	3.98	0.31	12.72	<.001
Report type (L1)	<b>0.22</b>	<b>0.05</b>	<b>4.16</b>	<b>&lt;.001</b>	0.07	0.05	1.45	.151	<b>1.09</b>	<b>0.31</b>	<b>3.55</b>	<b>&lt;.001</b>	0.30	0.15	2.04	.043
Week no. (L2)	-0.12	0.03	-3.50	<.001	-0.06	0.03	-2.15	.033	-0.53	0.22	-2.37	.018	-0.05	0.11	-0.49	.628
Time between methods (L3)	-0.01	0.01	-0.89	.378	0.00	0.01	0.67	.503	-0.00	0.06	-0.06	.949	0.01	0.03	0.43	.666
Baseline typical weekly drinks (L3)	0.03	0.01	3.07	.003	0.05	0.01	4.60	<.001	0.41	0.09	4.69	<.001	0.15	0.03	4.54	<.001
	Models predicting alcohol consequences outcomes															
	No. of positive consequences				No. of negative consequences											
	<i>B</i>	<i>SE</i>	<i>t</i>	<i>p</i>	<i>B</i>	<i>SE</i>	<i>t</i>	<i>p</i>								
Intercept	4.42	0.38	11.70	<.001	4.62	0.40	11.65	<.001								
Report type (L1)	<b>1.11</b>	<b>0.20</b>	<b>5.51</b>	<b>&lt;.001</b>	<b>1.13</b>	<b>0.21</b>	<b>5.32</b>	<b>&lt;.001</b>								
Week no. (L2)	-0.51	0.11	-4.68	<.001	-0.49	0.11	-4.30	<.001								
Time between methods (L3)	-0.06	0.03	-2.25	.027	-0.06	0.03	-2.24	.027								
Baseline typical weekly drinks (L3)	0.08	0.05	1.72	.090	0.09	0.05	1.71	.091								

Notes: TLFB = Timeline Followback; no. = number; L1 = Modeled at Level 1 portion of the model; L2 = Level 2; L3 = Level 3; report type coded 0 = daily report, 1 = TLFB. As such, intercept = outcome reported daily and effect of report type = difference in the outcome between daily and TLFB report; **bold** effects represent significant differences between methods ( $p < .01$ ); Weekly models included a random slope effect of report type on all outcomes with the exception of drinks per drinking day, for which there was instead a significant random slope effect of study week.

quences. Specifically, one's retrospective recall for the drinking consequences of a specific event might be influenced by expectancies for the types of consequences that are likely to result from a drinking event (i.e., situation-specific beliefs, such as "when people drink alcohol, they have a good time"), or one's knowledge about the types of consequences he/she typically experiences (i.e., identity-related beliefs, such as "when I drink alcohol, I usually make others laugh"). Further, as noted in the introduction, recall for negative events is typically better than for positive events (Earles et al., 2016; Holland & Kensinger, 2013; Kensinger, 2009; Mackay et al., 2004). This may explain, at least in part, why we found less evidence for reporting discrepancies in negative than in positive consequences of drinking.

Regardless of the mechanism of differences between report types of positive (but not negative) consequences, it is notable that when asked by a researcher to recall one's drinking experiences, participants reported more positive consequences than when completing daily assessments using a mobile device. This discrepancy might suggest that, with time, students may recall drinking events as characterized by more positive consequences than they actually were. This shift may reinforce drinking behavior over time and provide prime opportunities for novel interventions. For example, by providing daily-level feedback to participants about the true extent of their positive and negative alcohol-related experiences, interventions can provide corrective feedback and potentially break the instrumental link between false expectancies and alcohol use.

In contrast to our results, the majority of other studies

have shown that participants report less alcohol use on the TLFB when compared with daily diary (Dulin et al., 2017; Monk et al., 2015; Rowe et al., 2016). Methodological differences among studies may explain, in part, the divergent findings. For example, Monk et al. compared number of drinks measured in real-time (rather than next day, as in the present study) to those reported retrospectively at the end of 1 week (rather than 1 month). Although Dulin et al. used methods more similar to ours (daily reports compared to TLFB), they used a 6-week (rather than 1-month) timeframe. There were also differences in the sample characteristics across studies; Dulin et al. studied individuals with alcohol use disorders engaging in mobile treatment alongside the assessments, and Rowe et al. (2016) studied men who have sex with men and endorse both alcohol and methamphetamine use.

There are two additional potential explanations for the higher estimates on TLFB we observed, given the specifics of our methods. One is that the TLFB interviewer directly assisted participants in ensuring that larger drinks (e.g., 16 oz. beer) were counted in standard drink sizes during the TLFB. In addition, interviewers encouraged participants to rely on their smartphones (e.g., photos, text messages, social media) for extra "clues" about when they drank, how much they consumed, and what else occurred during the drinking event, which may have resulted in reports of drinking behavior that were closer to their actual experience. A similar level of active guidance could not be provided remotely each day. Although more research is needed (particularly because it is unclear whether our findings suggest that TLFB reports are more accurate versus simply higher), findings suggest

TABLE 3. Differences between TLFB and daily reports at the daily level

Fixed effects	Models predicting alcohol use outcomes									
	No. of drinks				Drinking day			Heavy drinking day		
	<i>B</i>	<i>SE</i>	<i>t</i>	<i>p</i>	OR	[95% CI]	<i>p</i>	OR	[95% CI]	<i>p</i>
Intercept	1.36	0.10	13.05	<.001	0.31	[0.26, 0.38]	<.001	0.16	[0.13, 0.21]	<.001
Report (L1)	<b>3.16</b>	<b>0.18</b>	<b>17.50</b>	<b>&lt;.001</b>	<b>1.25</b>	<b>[1.12, 1.39]</b>	<b>&lt;.001</b>	1.11	[0.96, 1.29]	.145
Day no. (L2)	-0.03	0.00	-5.99	<.001	0.96	[0.95, 0.97]	<.001	0.96	[0.95, 0.98]	<.001
Time between methods (L3)	-0.01	0.01	-0.56	.580	1.00	[0.98, 1.01]	.696	1.01	[0.99, 1.03]	.421
Baseline typical weekly drinks (L3)	0.06	0.01	4.36	<.001	1.03	[1.01, 1.06]	<.001	1.06	[1.04, 1.09]	<.001

  

Fixed effects	Models predicting alcohol consequences outcomes												
	No. positive consequences				No. negative consequences			Any positive consequence			Any negative consequence		
	<i>B</i>	<i>SE</i>	<i>t</i>	<i>p</i>	OR	[95% CI]	<i>p</i>	OR	[95% CI]	<i>p</i>	OR	[95% CI]	<i>p</i>
Intercept	2.81	0.15	18.20	<.001	0.60	[0.49, 0.75]	<.001	10.72	[6.73, 17.05]	<.001	0.79	[0.56, 1.11]	.172
Report type (L1)	<b>0.51</b>	<b>0.11</b>	<b>4.53</b>	<b>&lt;.001</b>	0.97	[0.84, 1.13]	.736	1.64	[1.05, 2.56]	.029	0.80	[0.61, 1.05]	.114
Day number (L2)	-0.03	0.01	-4.67	<.001	1.00	[0.99, 1.01]	.886	0.96	[0.94, 0.98]	<.001	1.00	[0.98, 1.02]	.948
Time between methods (L3)	-0.03	0.01	-2.35	.021	0.99	[0.97, 1.01]	.488	0.97	[0.93, 1.01]	.174	1.00	[0.97, 1.03]	.773
Baseline typical weekly drinks (L3)	-0.01	0.01	-0.50	.616	1.01	[0.99, 1.03]	.477	0.99	[0.95, 1.03]	.699	1.02	[0.99, 1.05]	.207

Notes: TLFB = Timeline Followback; no. = number; OR = odds ratio; CI = confidence interval; LL=lower limit; UL=upper limit; L1 = Modeled at Level 1 portion of the model; L2 = Level 2; L3 = Level 3; Report type coded 0 = daily, 1 = TLFB. As such, intercept = outcome reported daily and effect of report = difference in the outcome between daily and TLFB report; **bold** effects represent significant differences between methods ( $p < .01$ ). Daily models included a random slope effect of report type for number of drinks and number of positive consequences.

that perhaps more detailed TLFB procedures that allow participants to rely on personal visual cues (e.g., photos, social media) could help to reduce some of the recall issues demonstrated in prior research.

Second, in our daily assessment protocol, if participants endorsed alcohol use they received several follow-up questions about consequences. Likewise, if they endorsed consequences, they received follow-up questions about those consequences. Similar detailed follow-up on each drinking day was not requested during the TLFB. Participants may have intentionally denied alcohol use and/or consequences during daily assessment to avoid the burden of these follow-up questions on repeated assessments (Wray et al., 2014). Such behavior could have resulted in the systematically lower estimates observed on the daily reports relative to TLFB.

Our moderation findings highlight important distinctions in when reporting discrepancies emerge and for whom. The extent to which reports on the TLFB indicated more HED days relative to daily assessments increased among those who had a longer break between the end of daily assessment and the in-person TLFB. This was not surprising given the increase in bias that may occur with time. This discrepancy also increased among those with higher levels of baseline drinks per week, inconsistent with prior work, in which no impact of drinking involvement on correspondence between reports was observed (Monk et al., 2015). Of note, drinking involvement was assessed differently in these two studies, as Monk et al. used the Alcohol Use Disorders Identification Test, which assesses both alcohol use and consequences.

Our findings suggest that heavier drinkers may have greater error in reporting on one or both methods. It is possible that heavier drinkers are more likely to underreport on daily assessments because of an inability to recall how many drinks were consumed the prior night. Alternatively, they may be more likely to overreport on the TLFB because they are more likely to remember a heavier “typical” pattern of drinking than they actually consumed. Additional work to better understand the impact of drinking levels on reporting is warranted.

Moderation findings also revealed that as daily reporting day approached the time of the TLFB interview, the discrepancy between reports (number of drinks, positive consequences) increased. In prior work, discrepancies were greater at days or weeks more distal from the time of TLFB (Dulin et al., 2017; Hoepfner et al., 2010); however, the discrepancy itself was in the opposite direction (higher reports on intensive longitudinal assessment than TLFB) than observed here. Again, in this study, it is possible that as daily assessments went on, participants became more burdened and denied drinking and consequences to avoid follow-up questions, which would result in even lower estimates from daily reports at these more recent assessments.

It is important to note that although we examined differences in reporting between two methods, we are unable to determine which method provides more accurate information. Others have assumed that intensive longitudinal measurements collect more accurate and valid data because of the minimization of social desirability and recall biases (Rowe et al., 2016). Future work that objectively measures

and/or biologically verifies alcohol use, to compare with these data collection methods, would be useful. For example, transdermal alcohol sensors such as the Secure Continuous Remote Alcohol Monitor bracelet could be used to passively assess consumption in order to compare it with both daily and retrospective reports. Other, more creative, methods are needed to understand how to obtain the most accurate reports of alcohol-related consequences.

### Limitations

We acknowledge several study limitations. First, although daily surveys may minimize recall bias more so than retrospective interviews, they typically do not assess behavior in real-time. Monk et al. (2015) showed discrepancies between daily recording and real-time alcohol measures, indicating that daily reports could also involve recall bias. In addition, our study was conducted on a relatively homogenous sample of all heavy drinking college students, with limited racial/ethnic diversity, and who were especially compliant with the daily survey protocol. As such, generalizability is limited and future studies should replicate this examination in larger and more diverse samples. Further, we did not measure potentially important contextual variables that may affect the correspondence between reporting types (Monk et al., 2015). Future studies are needed that continue to seek to understand why reports from various assessment methods differ.

### Conclusion

We observed significant mean and within-person (daily, weekly) differences between daily reports and TLFB on several measures of alcohol use and consequences. Where differences existed, data collected on the TLFB suggested higher drinking involvement than data collected via daily reports, inconsistent with most previous research. This study was unique in comparing reports of alcohol-related consequences, with results suggesting that young adults reported more positive consequences on the TLFB than daily diaries. Together, these findings suggest that there may be value in assessing alcohol-related behaviors and constructs closer to when they occur, but that additional research is needed to better understand which methods of reporting are most accurate for understanding drinking behavior.

### References

Barnett, N. P., Merrill, J. E., Kahler, C. W., & Colby, S. M. (2015). Negative evaluations of negative alcohol consequences lead to subsequent reductions in alcohol use. *Psychology of Addictive Behaviors, 29*, 992–1002. doi:10.1037/adb0000095

Centers for Disease Control and Prevention. (2008). *Alcohol-related disease impact (ARDI) application*. Atlanta, GA: Author.

Chow, P. I., Lord, H. R., MacDonnell, K., Ritterband, L. M., & Ingersoll, K. S. (2017). Convergence of online daily diaries and timeline follow-

back among women at risk for alcohol exposed pregnancy. *Journal of Substance Abuse Treatment, 82*, 7–11. doi:10.1016/j.jsat.2017.08.004

Collins, R. L., Parks, G. A., & Marlatt, G. A. (1985). Social determinants of alcohol consumption: The effects of social interaction and model status on the self-administration of alcohol. *Journal of Consulting and Clinical Psychology, 53*, 189–200. doi:10.1037/0022-006X.53.2.189

Davis, C. G., Thake, J., & Vilhena, N. (2010). Social desirability biases in self-reported alcohol consumption and harms. *Addictive Behaviors, 35*, 302–311. doi:10.1016/j.addbeh.2009.11.001

Dulin, P. L., Alvarado, C. E., Fitterling, J. M., & Gonzalez, V. M. (2017). Comparisons of alcohol consumption by time-line follow back vs. smartphone-based daily interviews. *Addiction Research & Theory, 25*, 195–200. doi:10.1080/16066359.2016.1239081

Earles, J. L., Kersten, A. W., Vernon, L. L., & Starkings, R. (2016). Memory for positive, negative and neutral events in younger and older adults: Does emotion influence binding in event memory? *Cognition & Emotion, 30*, 378–388. doi:10.1080/02699931.2014.996530

Hoepfner, B. B., Stout, R. L., Jackson, K. M., & Barnett, N. P. (2010). 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, 35*, 1138–1143. doi:10.1016/j.addbeh.2010.08.013

Holland, A., & Kensinger, E. (2013). Emotion in episodic memory: The effects of emotional content, emotional state, and motivational goals. In J. Armony & P. Vuilleumier (Eds.), *The Cambridge handbook of human affective neuroscience* (pp. 465–488). New York, NY: Cambridge University Press.

Holmes, J., Angus, C., Buykx, P., Ally, A., Stone, T., Meier, P., & Brennan, A. (2016). *Mortality and morbidity risks from alcohol consumption in the UK: Analyses using the Sheffield Alcohol Policy Model (v. 2.7) to inform the UK Chief Medical Officers' review of the UK lower risk drinking guidelines*. Sheffield, England: ScHARR, University of Sheffield.

Hufford, M. R., Shields, A. L., Shiffman, S., Paty, J., & Balabanis, M. (2002). Reactivity to ecological momentary assessment: An example using undergraduate problem drinkers. *Psychology of Addictive Behaviors, 16*, 205–211. doi:10.1037/0893-164X.16.3.205

Joinson, A. (1999). Social desirability, anonymity, and Internet-based questionnaires. *Behavior Research Methods, Instruments, & Computers, 31*, 433–438. doi:10.3758/BF03200723

Kahler, C. W., Strong, D. R., & Read, J. P. (2005). Toward efficient and comprehensive measurement of the alcohol problems continuum in college students: The Brief Young Adult Alcohol Consequences Questionnaire. *Alcoholism: Clinical and Experimental Research, 29*, 1180–1189. doi:10.1097/01.ALC.0000171940.95813.A5

Kensinger, E. A. (2009). Remembering the details: Effects of emotion. *Emotion Review, 1*, 99–113. doi:10.1177/1754073908100432

Lee, C. M., Crounce, J. M., Baldwin, S. A., Fairlie, A. M., Atkins, D. C., Patrick, M. E., . . . Leigh, B. C. (2017). Psychometric analysis and validity of the daily alcohol-related consequences and evaluations measure for young adults. *Psychological Assessment, 29*, 253–263. doi:10.1037/pas0000320

Mackay, D. G., Shafto, M., Taylor, J. K., Marian, D. E., Abrams, L., & Dyer, J. R. (2004). Relations between emotion, memory, and attention: Evidence from taboo Stroop, lexical decision, and immediate memory tasks. *Memory & Cognition, 32*, 474–488. doi:10.3758/BF03195840

Merrill, J. E., Rosen, R. K., Walker, S. B., & Carey, K. B. (2018). A qualitative examination of contextual influences on negative alcohol consequence evaluations among young adult drinkers. *Psychology of Addictive Behaviors, 32*, 29–39. doi:10.1037/adb0000339

Merrill, J. E., Vermont, L. N., Bachrach, R. L., & Read, J.P. (2013). Is the pregame to blame? Event-level associations between pregameing and alcohol-related consequences. *Journal of Studies on Alcohol and Drugs, 74*, 757–764. doi:10.15288/jsad.2013.74.757

- Monk, R. L., Heim, D., Qureshi, A., & Price, A. (2015). "I have no clue what I drank last night" using Smartphone technology to compare in-vivo and retrospective self-reports of alcohol consumption. *PLoS One*, *10*, e0126209. doi:10.1371/journal.pone.0126209
- Patterson, C., Hogan, L., & Cox, M. (2019). A comparison between two retrospective alcohol consumption measures and the daily drinking diary method with university students. *American Journal of Drug and Alcohol Abuse*, *45*, 248–253. doi:10.1080/00952990.2018.1514617
- Poulton, A., Pan, J., Bruns, L. R., Jr., Sinnott, R. O., & Hester, R. (2018). Assessment of alcohol intake: Retrospective measures versus a smartphone application. *Addictive Behaviors*, *83*, 35–41. doi:10.1016/j.addbeh.2017.11.003
- Raudenbush, S. W., Bryk, A. S., & Congdon, R. (2013). *HLM 7.01 for Windows* [Computer software]. Skokie, IL: Scientific Software International, Inc.
- Robinson, M. D., & Clore, G. L. (2002). Belief and feeling: Evidence for an accessibility model of emotional self-report. *Psychological Bulletin*, *128*, 934–960. doi:10.1037/0033-2909.128.6.934
- Rowe, C., Hern, J., DeMartini, A., Jennings, D., Sommers, M., Walker, J., & Santos, G. M. (2016). Concordance of text message ecological momentary assessment and retrospective survey data among substance-using men who have sex with men: A secondary analysis of a randomized controlled trial. *JMIR mHealth and uHealth*, *4*(2), e44. doi:10.2196/mhealth.5368
- Simons, J. S., Wills, T. A., Emery, N. N., & Marks, R. M. (2015). Quantifying alcohol consumption: Self-report, transdermal assessment, and prediction of dependence symptoms. *Addictive Behaviors*, *50*, 205–212. doi:10.1016/j.addbeh.2015.06.042
- Sobell, L. C., Agrawal, S., Sobell, M. B., Leo, G. I., Young, L. J., Cunningham, J. A., & Simco, E. R. (2003). Comparison of a quick drinking screen with the timeline followback for individuals with alcohol problems. *Journal of Studies on Alcohol*, *64*, 858–861. doi:10.15288/jsa.2003.64.858
- Sobell, L. C., & Sobell, M. B. (1992). Timeline follow-back: A technique for assessing self-reported alcohol consumption. In R. Z. Litten & J. P. Allen (Eds.), *Measuring alcohol consumption: Psychosocial and biochemical methods* (pp. 41–72). Totowa, NJ: Humana Press.
- Toll, B. A., Cooney, N. L., McKee, S. A., & O'Malley, S. S. (2006). Correspondence between Interactive Voice Response (IVR) and Timeline Followback (TLFB) reports of drinking behavior. *Addictive Behaviors*, *31*, 726–731. doi:10.1016/j.addbeh.2005.05.044
- Wray, T. B., Adia, A. C., Pérez, A. E., Simpanen, E. M., Woods, L. A., Celio, M. A., & Monti, P. M. (2019). Timeline: A web application for assessing the timing and details of health behaviors. *American Journal of Drug and Alcohol Abuse*, *45*, 141–150. doi:10.1080/00952990.2018.1469138
- Wray, T. B., Merrill, J. E., & Monti, P. M. (2014). Using ecological momentary assessment (EMA) to assess situation-level predictors of alcohol use and alcohol-related consequences. *Alcohol Research: Current Reviews*, *36*, 19–27.