Heaviness of Alcohol Use, Alcohol Problems, and Subjective Intoxication Predict Discrepant Drinking Reports in Daily Life

Angela K. Stevens, Alexander W. Sokolovsky, Hayley Treloar Padovano, Helene R. White, and Kristina M. Jackson

Background: Self-reported consumption is pervasive in alcohol research, though retrospective recall bias is a concern. Fine-grained methods are designed to limit retrospection; yet, discrepancies can arise when comparing responses on fine-grained surveys with responses to retrospective surveys across weeks or months. Many fine-grained studies use both repeated daily surveys (RDS) and end-of-day (EOD) summaries, but little research has examined whether these survey types are consistent. The purpose of this study was to quantify the magnitude and directionality of discrepancy between EOD summaries and RDS and identify alcohol-related predictors of discrepancy.

Methods: As a part of a larger study, college student alcohol and cannabis users ($N = 341$; 53% women; $M_{age} = 19.79$ years) were recruited to complete 56 days of data collection, including 5 daily assessments of their substance use and related constructs, one of which included an EOD summary of the previous day. Generalized linear mixed-effects models were used to examine between- and within-person predictors of a 5-category, discrepancy outcome: no discrepancy, low discrepancy where RDS < EOD, low discrepancy where EOD < RDS, high discrepancy where RDS < EOD, and high discrepancy where EOD < RDS.

Results: Discrepancies between EOD and RDS were observed in both directions. Alcohol problems predicted more alcohol consumption reported on the EOD survey than across RDS. Within-person alcohol quantity and hourly rate of consumption were most strongly related to less alcohol consumption reported on the EOD survey. Between- and within-person peak subjective intoxication and within-person liquor consumption were associated with discrepancies in both directions.

Conclusions: Surveys requiring more retrospection may overestimate alcohol consumption in problematic drinkers and underestimate consumption on days where more alcohol is consumed than typical. Evidence also suggests that greater day-to-day instability in alcohol behavior is linked to less consistent reporting overall. More research is needed to discern factors contributing to inconsistent reporting on fine-grained surveys to maximize the validity of reports.

Key Words: Daily Diary, Self-Report Discrepancy, Ecological Momentary Assessment, Alcohol, College Students.

ALCOHOL RESEARCH RELIES heavily on the validity of self-reported consumption, and the validity of this approach has been established (Glovannucci et al., 1991; Simons et al., 2015). Despite the advantages of self-report approaches (e.g., convenience, low cost, and low burden), retrospective recall bias is a concern. Indeed, fine-grained methods to assess alcohol consumption, such as experience sampling methods, like ecological momentary assessment (EMA), attempt to mitigate this concern by limiting the time period over which individuals are asked to recall their alcohol involvement (Shiffman, 2009; Shiffman et al., 2008; Wray et al., 2014). To maximize sensitivity of reports, studies often incorporate both "real-time" and delayed (e.g., next-day summary reports) assessments for the same day (e.g., Monk et al., 2015; see Stone and Shiffman, 2002).

This approach of retrospectively assessing 1 day of behavior is known as an end-of-day (EOD) summary (Shiffman, 2009). The EOD summary is attractive because it is comparatively less burdensome to participants, and it provides the opportunity to detect additional drinks not reported in real-time assessments. However, there is potential for these 2 self-reported assessments to be discrepant which is problematic in terms of establishing validity of the drinking outcome (Shiffman et al., 2008; Stone and Shiffman, 2002). Logically, number of drinks reported at the EOD summary should be equal to or greater than a count of real-time EMA drink reports. Alternately, the EOD summary may underestimate the number of prior-day drinks due to retrospection error.

From the Center for Alcohol and Addiction Studies,(AKS, AWS, HTP, KMJ), Brown University School of Public Health, Providence, Rhode Island; and Center of Alcohol and Substance Studies, Rutgers, (HRW), The State University of New Jersey, Piscataway, New Jersey.
Received for publication November 25, 2019; accepted May 5, 2020.
Reprint requests: Angela Stevens, PhD, MPH, Center for Alcohol and Addiction Studies, Brown University, Box G-S121-4, Providence, RI 02912; Tel.: 401-863-6489; Fax: 1-401-863-6697; E-mail: Angela_Stevens@brown.edu
© 2020 by the Research Society on Alcoholism.

DOI: 10.1111/acer.14362

Few studies have examined comparisons between EOD summary self-reports and real-time self-reports, despite many studies incorporating both approaches to assess alcohol and other substance use (e.g., Carney et al., 2000; Kiene et al., 2009; Patterson et al., 2019). Thus, the purpose of the present study was to quantify the directionality and magnitude of discrepancy between repeated daily surveys (RDS) and EOD summary reports, corresponding to the same day of behavior, and to examine both between- and within-person predictors of this discrepancy.

Technological advances in alcohol research have engendered myriad advances in the field, particularly with the opportunity to assess dynamic relations with minimal retrospective recall bias (i.e., EMA; Shiffman, 2009; Shiffman et al., 2008; Wray et al., 2014). EMA approaches are designed to reduce the possibility of retrospective recall bias that is often associated with self-report measures by: (i) collecting self-report data that are frequent and/or labile in real-world contexts; (ii) assessing individuals based on their recent and/or current behaviors (e.g., alcohol use) and experiences; (iii) completing repeated assessments over time; and (iv) utilizing random and/or event-contingent prompts (Shiffman et al., 2008; Stone and Shiffman, 1994; Stone and Shiffman, 2002; Trull and Ebner-Priemer, 2009; Wray et al., 2014). Extant research indicates significant discrepancies between self-reports collected retrospectively over weeks or months (e.g., timeline followback) and real-time assessments of mood, behavior, and mental health symptoms (e.g., Shiffman et al., 1997; Wray et al., 2016; see Solhan et al., 2009), with a general consensus that real-time assessments are more reliable and valid than retrospective reports (Hufford, 2007; Trull and Ebner-Priemer, 2009).

Discrepancies between reporting substance use behavior on self-reports requiring retrospection across weeks or months versus real-time self-reports are well-established (e.g., Carney et al., 1998; Dulin et al., 2017; Feunekes et al., 1999; Griffith et al., 2009; Leigh et al., 1998; Monk et al., 2015; Rowe et al., 2016; Shiffman et al., 1997; Simpura and Poikolainen, 1983), and there are many factors that can contribute to this discrepancy. Importantly, the process by which individuals retrospectively estimate the frequency or intensity of a behavior or experience is complex and is influenced by a variety of heuristics (see Gorin and Stone, 2001). For example, individuals' retrospective responses may be influenced by the context at the time of the assessment (Shiffman et al., 2008). This may be particularly important for substance use research, as situational contexts and cues aid in accurate reporting of use behavior (Godden and Baddeley, 1975; Shiffman, 2009). Further, retrospective recall is also influenced by individuals' variability in behavior, such that individuals with more inconsistent behavior (e.g., less patterned and more variable alcohol consumption) tend to be less consistent in their reports (Gmel and Daeppen, 2007; Perrine and Schroder, 2005; Solhan et al., 2009). Though evidence comparing aggregate retrospective self-report and real-time self-report via EMA points to significant discrepancy, many comparisons between these 2 approaches (e.g., Dulin et al., 2017) are limited by utilizing between-person analyses that neglect intraindividual variability in recall bias across an assessment window (Curran and Bauer, 2011; Shiffman et al., 2008).

Recent research in this area has started examining discrepancies in real-time self-reports at the daily level (i.e., within-person) rather than the aggregate level (i.e., between-person). For example, Rowe and colleagues (2016) used generalized estimating equations accounting for the clustering of assessments between persons and examined the concordance of alcohol and illicit substance use reported on a 14-day recall survey vs. EOD summary responses reflecting their behavior on the previous day. Findings showed that alcohol use was reported more frequently on EOD summary responses (vs. 14-day recall) and that binge drinking was associated with a lower concordance between the 2 approaches (Rowe et al., 2016). Despite the strengths of this approach, this study did not consider the direction or the magnitude of the discrepancy and instead used a binary (i.e., yes vs. no) assessment of the concordance between the 2 approaches. Further, EOD summaries, in addition to the 14-day recall survey, require some degree of retrospection; thus, little is known about the within-person concordance between EOD summaries and EMA surveys.

Only one study, to our knowledge, has examined the discrepancy between EMA surveys and EOD summary reports, which requires retrospection across 1 day. Specifically, Monk and colleagues (2015) found that alcohol type (e.g., beer, wine, and spirit) was the only significant predictor of a daily difference score between EMA surveys and EOD summary reports in a sample of 69 adults (M_{age} = 21.47). Interestingly, situational contexts (e.g., at a bar/club/pub) and alcohol-related problems were not significantly related to a drinking report discrepancy across 1 day. This landmark study used data from the first day of the EMA study where drinking was reported and subsequently assessed every hour for each participant. However, this study did not consider the magnitude of the discrepancy, restricting conclusions that can be drawn regarding how discrepancies fluctuate intraindividually across time.

Current Study

Building on prior research in this area, the purpose of this study was to examine the magnitude and directionality of discrepant drinking reports using an approach that disentangles between- and within-person effects in a large daily diary sample of college students spanning 56 days of assessment with 5 repeated assessments per day (RDS) and an EOD summary report each day. To our knowledge, we are the first to consider both the discrepancy magnitude (i.e., the absolute difference in the number of drinks reported on EOD summary vs. RDS) and directionality (i.e., RDS < EOD summary vs. EOD summary < RDS) when comparing EOD summaries and RDS reflecting substance use behavior on the same day. We first sought to evaluate the degree to which between-person alcohol use characteristics (i.e., Alcohol Use Disorders Identification Test [AUDIT]; average alcohol quantity) and sociodemographic (e.g., sex) and contextual (i.e., day of the week) predictors were related to discrepant drinking reports (Aim 1). We hypothesized that higher AUDIT scores and greater alcohol quantity aggregated across the study would be associated with drinking report discrepancies (vs. no discrepancy); relationships between predictors and the magnitude and directionality of these discrepancies were exploratory. We then sought to test whether characteristics of alcohol use on a given day (i.e., total number of drinks, peak hourly drinking rate, peak subjective intoxication, and consuming liquor) were related to discrepant drinking reports while disentangling between- and within-person effects and adjusting for covariates (Aim 2). Although we hypothesized that these between- and within-person characteristics of drinking would relate to drinking report discrepancy, findings from the analyses of these understudied predictors should be considered exploratory. Data were drawn from a parent study examining simultaneous alcohol and cannabis use; thus, cannabis use on a given day was included as a covariate in all models.

MATERIALS AND METHODS

Design and Sample

Screening Survey. Full-time students between ages 18 and 24 years were recruited from universities in 3 states with varying recreational cannabis policies to participate in a larger parent study on simultaneous alcohol and cannabis use. Eight thousand students were randomly chosen from each university's registrar database stratified by expected year of graduation (total N = 24,000) and were emailed an invitation to participate in an online screening survey. Screening completers (N = 7,000), relative to noncompleters, included more women, more White students, fewer Black students, more Asian students, more Hispanic/Latinx students, and more younger students (i.e., ages 18–21). Nonetheless, effects sizes were small (i.e., Cohen's h = 0.07–0.26). Of those screened as a part of the parent study, 2,874 (41.1%) were deemed eligible to participate based on age (i.e., 18–24), being enrolled full-time, and endorsing past-year alcohol and cannabis use. Students who completed the screening survey were eligible for a lottery to win \$100. See White and colleagues (2019) and Appendix S1 for further details regarding screening for the parent study.

Baseline Survey. Of those eligible for the larger parent study, a random sample of 2,501 students stratified by university was invited via email to participate in the baseline survey, and 1,524 (60.9%) of invitees completed the baseline survey. Of these, 1,390 (91.2%) were retained in analyses after excluding students who provided responses inconsistent with baseline survey eligibility criteria (see above) or whose surveys had technical problems. Students reporting past-month alcohol and cannabis use were oversampled. See White and colleagues (2019) and Appendix S1 for further details regarding the baseline survey of the parent study.

Daily Survey. Of those who completed the baseline survey, 693 used alcohol and cannabis at the same time "so that their effects overlapped" within the past month, making them eligible to participate in the daily phase of the study. In this phase, repeated daily surveys (RDS) were completed 5 times per day using a custom smartphone application. Daily survey recruitment was stratified based on frequency of past-month simultaneous use and sex to ensure roughly equal numbers of men and women and to oversample frequent simultaneous users. Daily data collection directly followed the longer surveys (baseline and 3-month follow-up) and comprised 28 days of RDS at each burst (56 total days) prompted at 9:00 am, 2:00 pm, 5:00 pm, 8:00 pm, and 11:00 pm. At the 9:00 am survey, participants were also asked additional questions retrospectively assessing yesterday's behavior through bedtime (i.e., EOD summary report), in addition to the RDS questions assessed at each subsequent survey. Participants were provided 4 hours to complete the 9:00 am survey and 2 hours to complete 2:00 pm, 5:00 pm, 8:00 pm, and 11:00 pm surveys. Reminders were provided to participants 15 minutes before the survey closed. See Appendix S1 for additional details regarding the daily phase of the parent study. See Fig. S1 for a flow chart of data collection for the parent study summarized above (adapted from Sokolovsky et al., 2020).

We retained data from 54 study days due to technological difficulties during the first 2 days of the study. The daily phase of the study included 343 participants; 2 participants only completed the first 2 days of data collection resulting in our final sample of 341 students. Among these students (53% women, M age = 19.79; 74% White; 10% Hispanic/Latinx), 32.3% of students were from School A (in a state where recreational cannabis use is illegal), 34.1% from School B (in a state where recreational cannabis use is decriminalized), and 33.5% from School C (in a state where recreational cannabis use is legal for adults 21 and older).[1] Participants were compensated \$25 for the baseline survey, \$35 for the follow-up survey, and \$1 for each completed daily survey, with weekly and overall bonuses. Aggregated across the 5 RDS, mean RDS compliance equaled 88.4%, and mean EOD summary compliance equaled 81.9%. Both compliance rates exceed the pooled compliance rate shown in a recent meta-analysis on EMA and substance use (Jones et al., 2019). As a part of the parent study, all participants were trained on standardized drink equivalences set forth by the National Institute of Alcoholism and Alcohol Abuse (NIAAA, 2007). All procedures were approved by the coordinating university's Institutional Review Board. A Certificate of Confidentiality was obtained from the National Institute on Drug Abuse.

Measures

Participants provided demographic information at the baseline survey, including age, sex, race, ethnicity, and fraternity/sorority affiliation status.

Baseline Survey. *Alcohol Use Disorders Identification Test (AUDIT)* was used to assess alcohol-related problems at baseline using a 10-item self-report measure (Babor et al., 2001). AUDIT scores ranged from 0 to 32, in this sample, with higher scores reflecting higher past-year alcohol-related problems (α=.74).

EOD Summary Report. *Daily Reports of Alcohol Quantity*. Each morning, participants were presented a sliding scale ranging from <1 to 15+ and were asked to recall their drinking levels from the prior day. Specifically, they were asked, "How many drinks did you consume in total between the time you woke until the time you went to bed yesterday?" This EOD summary report occurred at the 9:00 am survey reflecting yesterday's behavior.

[1]Descriptive information is provided for the full daily sample. To achieve study aims, we pared down the full daily sample, as described in the Analytic Strategy, to the "restricted daily sample." Descriptive statistics for the restricted daily sample (used in analyses) are provided in Table 1 alongside those for the full daily sample for comparison.

Repeated Daily Survey (RDS). *Peak Subjective Intoxication.* At each of the 5 RDS, participants reported their subjective effects from alcohol on a graphical interface, with anchors for not at all drunk, a little drunk, moderately drunk, and very drunk; the maximum subjective intoxication for a given person on a given day was their highest rating reported across RDS, ranging from 0 (*not at all drunk*) to 3 (*very drunk*). *Quantity of Alcohol Consumption.* Participants indicated the number of drinks using the same graphical interface, tapping the screen for each drink consumed since their last RDS. The sum of drinks reported at each completed RDS determined the total drinks reported across RDS that day. (See Appendix S1 for screenshots of this graphical interface.) *Peak Hourly Drinking Rate.* Peak hourly drinking rate was defined by computing the maximum number of drinks consumed within 1 hour on a given day for a given person. *Liquor Consumption.* At each RDS following endorsement of alcohol use, participants were asked, "What type of alcohol had you been drinking between X and Y?" Options included "beer" (no/yes), "wine" (no/yes), "liquor" (no/yes), and "beer alternative" (no/yes). *Cannabis Product (covariate).* At each RDS survey following endorsement of cannabis use, participants were asked, "In what form was the marijuana you used between X and Y?" Options included "dry leaf" (no/yes), "concentrate" (no/yes), and "edible" (no/yes). On days when cannabis use was not endorsed, values for leaf, concentrate, and edible cannabis were recoded as "no" (0).

Analytic Strategy

Data management and coding were conducted in R version 3.6.0 (R Core Team, 2013) and SAS 9.4™ software (SAS Institute Inc., 2012). For the present investigation, all RDS were aggregated to the daily level to match the level of analysis (daily) of the EOD summaries. Observations were lagged such that EOD summaries and RDS reflected the same assessment day ($n_{\text{observations}}$ = 15,798). The data were restricted to drinking days reported on RDS and EOD summaries ($n_{\text{drinking days}}$ = 3,818; 24% of original observations). Daily observations were excluded if 2 adjacent RDS were missed which would result in an incomplete coverage of the survey day (*n* excluded = 841 observations). A subset of daily data was also excluded from analyses due to technological errors during data collection (*n* excluded = 130 observations). Due to concerns about reliability of reports of excessive consumption that might unduly influence our study findings, daily observations exceeding 15 drinks were capped at 15 drinks (*n* truncated = 100 observations), which is consistent with prior work (O'Grady et al., 2011). (See Fig. S2 for a flow chart of data selection from the full daily sample to the restricted sample.)

Of the 2,847[2] observations retained in analyses (henceforth referred to as the restricted sample), 46.05% (*n* = 1,311) were not discrepant such that the number of drinks reported across RDS on a given day for a given person was identical to the number of drinks reported on the EOD summary. A quarter (*n* = 707; 24.83%) of observations were discrepant by 1 drink in either direction (i.e., alcohol consumption reported on the EOD summary exceeded alcohol consumption reported across RDS vs. alcohol consumption reported on the EOD summary was lower compared to alcohol consumption reported across RDS). For observations where alcohol use reported on the EOD summary exceeded alcohol use reported across RDS, the number of discrepant drinks in a day ranged from 1 to 9. For observations where total number of drinks reported across RDS exceeded alcohol use reported on the EOD summary, the number of discrepant drinks ranged from 1 to 14. The median absolute difference between discrepant RDS and EOD summaries equaled approximately 1 drink. Thus, to quantify both the magnitude and the direction of discrepant observations, we created a dependent variable with 5 categories using a median split. Categories included (1) no discrepancy (46%); (2) RDS < EOD by 1 drink ("low"; 11%); (3) EOD < RDS by 1 drink ("low"; 14%); (4) RDS < EOD by 2 or more drinks ("high"; 8%); and (3) EOD < RDS by 2 or more drinks ("high"; 22%) (see Fig. S3).[3]

To determine the effects of between- and within-person predictors on repeated discrepant observations (level-1) nested within participants (level-2), 5 generalized linear mixed-effects models (GLMMs) were conducted using PROC GLIMMIX in SAS using Laplace approximation given the multinomial nature of the outcome (Hedeker, 2005). GLMMs, an extension of multilevel modeling, are necessary to account for clustering that is inherent for nested data (repeated days at level-1 nested within participants at level-2), which would violate the assumption of independent errors in ordinary least squares regression (Curran and Bauer, 2011; Hox et al., 2017; Raudenbush and Bryk, 2002; Singer, 1998).[4] For Aim 1, we examined a multivariate model containing between-person sociodemographic characteristics (i.e., age, sex, school, and fraternity/sorority status), between-person alcohol consumption (baseline AUDIT score and total number of daily drinks reported across RDS averaged across the 54 study days), and day of the week (i.e., weekday [Sunday–Thursday = 0] vs. weekend [Friday–Saturday = 1]) to predict discrepant grouping (0 = reference group).[5] We adjusted for cannabis use products in the Aim 1 model. All continuous predictors were grand-mean-centered for Aim 1.

For Aim 2, 4 additional models evaluated between- and within-person alcohol-related predictors reported on RDS: total number of drinks, peak hourly drinking rate, peak subjective intoxication, and consuming liquor (1 = yes, 0 = no). Between-person and within-person effects of daily-level predictors were disaggregated in each of these 4 models. For continuous daily-level predictors, within-person variations were assessed by creating a person-centered variable (e.g., each individual's daily alcohol quantity was centered around their own average level of alcohol quantity across days), such that a

[2]Despite paring down observations from the parent study for our specific analyses in the present study, with 2,847 observations across 310 participants, we had adequate power to detect significant effects. Simulation studies indicate sufficient power to provide unbiased parameter estimates when level-2 units are approximately 100 and when level-1 units are large (≥ 18; (Maas and Hox, 2005; Mathieu et al., 2012); our restricted sample far exceeds these recommendations (i.e., 310 participants with an average of 38 level-1 units [SD = 14] for each participant).

[3]Other similar works examining discrepancies between retrospective and real-time reports (e.g., Dulin et al., 2017; Monk et al., 2015) have used a difference score as the outcome. The authors considered this approach, which yielded a zero-inflated overdispersed outcome, to be best suited for a zero-inflated negative binomial distribution. Considering that the directionality of the discrepancy could be positive or negative (i.e., EOD < RDS vs. RDS < EOD), this difference score also contained negative values, which is prohibited by negative binomial regression (see Cameron and Trivedi, 1986; Agresti, 2002). However, an absolute value of this difference score would obscure the directionality of the discrepancy. Thus, we created a 5-level multinomial outcome to examine both the directionality and the magnitude of the discrepancy.

[4]There are limited documented procedures for calculating an intraclass correlation coefficient (ICC) when using generalized linear mixed models with a multinomial outcome (Nakagawa et al., 2017; Sommet and Morselli, 2017). When treating our outcome variable as dichotomous (i.e., discrepancy vs. no discrepancy), the ICC is equal to 0.114, which indicates 11% of the variance occurs at the between-person level. This indicates significant clustering that would violate assumptions of OLS regression. Thus, multilevel modeling is needed in this case to avoid drawing erroneous conclusions (Curran and Bauer, 2011; Hox et al., 2017; Raudenbush and Bryk, 2002; Singer, 1998).

[5]Note that previous analyses of these data indicate that drinking quantity on Thursdays is closer to weekday than weekend amounts

positive (or negative) value for alcohol quantity, for a given person on a particular day, reflected above (or below) average alcohol quantity (Curran and Bauer, 2011; Raudenbush and Bryk, 2002). Between-person variations were constructed by centering a given individual's overall mean around the sample grand mean (Curran and Bauer, 2011). For categorical predictors (i.e., any daily liquor endorsement), the between-person effect was constructed by calculating a proportion of study days where liquor was endorsed for each person. A random intercept was included in all models. Aim 2 models adjusted for day of the week, cannabis use mode, age, sex, and school. In all models, odds ratios were determined to be significantly larger than others if their 95% confidence intervals did not overlap (Cumming and Finch, 2005).

RESULTS

See Table 1 for descriptive statistics of the full study sample and the restricted sample (used in all analyses in the present study). Using logistic regression to predict participants being excluded from analyses, higher AUDIT scores, more frequent liquor consumption, lower daily alcohol quantity, lower peak subjective intoxication, less frequent leaf/concentrate/edible cannabis use, and no fraternity/sorority affiliation were associated with increased odds of being excluded from the present analyses. See Table 2 for drinking report discrepancy descriptive statistics for the restricted sample overall and by age, sex, race, and fraternity/sorority affiliation status.

Aim 1: Average Alcohol Use and Alcohol Problems

As shown in Table 3, baseline AUDIT scores were associated with significantly greater odds of a low RDS < EOD (vs. no discrepancy) and a high RDS < EOD (vs. no discrepancy) but not low or high EOD < RDS. Alcohol quantity averaged across study days was also linked to significantly greater odds of a low RDS < EOD, high RDS < EOD, and a high EOD < RDS as compared to no discrepancy. This effect did not disentangle the between-person and within-person effects of alcohol quantity (see below). Weekend (vs. weekday) and leaf cannabis use (vs. not) were associated with all discrepancy categories. Being male was linked to lower odds of a high EOD < RDS (see Table 3).

Table 1. Demographic and Alcohol-Related Characteristics of Full Study Sample and Restricted Sample

	M (SD) or %	
Variable	Full sample N = 341 n = 15,798	Restricted sample[a] N = 310 n = 2,847
Age	19.79 (1.31)	20.11 (1.32)
Sex (female)	52.60%	49.77%
White	73.80%	78.61%
Hispanic/Latinx	9.80%	7.80%
Fraternity/sorority	29.40%	36.32%
AUDIT	9.50 (5.11)	10.35 (5.04)
Daily alcohol quantity[b]	1.28 (2.90)	5.03 (2.23)
Daily peak hourly drinking rate[b]	0.66 (0.01)	2.56 (0.98)
Daily peak subjective effects[b]	0.94 (0.01)	1.72 (0.61)
Daily liquor consumption[b]	53.35%	52.27%
Daily leaf cannabis use[b]	31.19%	42.08%
Daily cannabis concentrate use[b]	9.22%	13.87%
Daily edible cannabis use[b]	1.77%	2.49%

AUDIT, Alcohol Use Disorders Identification Test; N, number of persons; n, number of daily observations.

[a]Restricted sample is limited to observations on end-of-day (EOD) summary reports and repeated daily surveys (RDS) endorsing any alcohol consumption and excludes daily observations where technological errors occurred and where 2 or more surveys in a row were missed. See Fig. S2 for a full description of data selection from the full daily sample to the restricted sample used in analyses.

[b]Reported on RDS.

Aim 2: Between- and Within-Person Alcohol-Related Predictors

Model 1: Alcohol Quantity. After adjusting for age, sex, school, day of the week, and cannabis use products, between-person alcohol quantity was associated with each drinking report discrepancy category when compared to no discrepancy (see Table 4). That is, heavier drinkers were less "consistent" in both directions. Likewise, a within-person positive variation in alcohol quantity (relative to each person's mean) was linked to greater odds of a low EOD < RDS, high RDS < EOD, and high EOD < RDS all compared to no discrepancy, with the largest effect for high EOD < RDS (OR = 1.82). Leaf cannabis use was associated with greater odds of all discrepancy categories, whereas weekend (vs. weekday) and being male were linked to lower odds of a high EOD < RDS (see Table S1).

Model 2: Peak Hourly Drinking Rate. At the between-person level, peak hourly drinking rate was associated with greater odds of all discrepancy categories, with the largest effect for high EOD < RDS (OR = 2.25) after adjusting for covariates. Within-person positive variations in peak hourly drinking rate were linked to greater odds of all drinking report discrepancy categories. The largest effect was found for high EOD < RDS (OR = 2.25) after adjusting for covariates (see Table 4). Leaf cannabis use was linked to greater odds of all discrepancy categories. Use of cannabis concentrates was linked to greater odds of high EOD < RDS. Weekend (vs. weekday) was associated with greater odds of high RDS < EOD and high EOD < RDS (see Table S2).

Model 3: Subjective Intoxication. Between-person subjective intoxication was significantly related to greater odds of all drinking report discrepancy categories after adjusting for covariates, with comparable effects. Likewise, within-person positive variations in subjective intoxication were linked to significantly greater odds of all drinking report discrepancy categories, with comparable effects (see Table 4). Weekend (vs. weekday) was associated with greater odds of high RDS < EOD and high EOD < RDS, and leaf cannabis use was related to low RDS < EOD. Being younger was linked to lower odds of high EOD < RDS (see Table S3).

Table 2. Drinking Report Discrepancy Descriptive Statistics

	No discrepancy *n* (%)	Low RDS < EOD *n* (%)	Low EOD < RDS *n* (%)	High RDS < EOD *n* (%)	High EOD < RDS *n* (%)
Restricted sample	1311 (46.05%)	308 (10.82%)	399 (14.01%)	217 (7.62%)	612 (21.50%)
Age 21+	598 (49.50%)	139 (11.51%)	179 (14.82%)	84 (6.95%)	208 (17.22%)
Age < 21	713 (43.50%)	169 (10.31%)	220 (13.42%)	133 (8.11%)	404 (24.65%)
Female	654 (46.15%)	139 (9.81%)	210 (14.82%)	105 (7.41%)	309 (21.81%)
Male	657 (45.94%)	169 (11.82%)	189 (13.22%)	112 (7.83%)	303 (21.19%)
White	999 (44.64%)	250 (11.17%)	306 (13.67%)	186 (8.31%)	497 (22.21%)
Non-White	312 (51.23%)	58 (9.52%)	93 (15.27%)	31 (5.09%)	115 (18.88%)
Frat/Sor Affiliated	446 (43.13%)	111 (10.74%)	132 (12.77%)	103 (9.96%)	242 (23.40%)
Nonaffiliated	865 (47.71%)	197 (10.87%)	267 (14.73%)	114 (6.29%)	370 (20.41%)

n = 2,847. "Low" refers to a discrepancy by 1 drink. "High" refers to a discrepancy by 2 or more drinks. RDS < EOD reflects that alcohol quantity reported on the end-of-day (EOD) summary survey exceeded the alcohol quantity reported across repeated daily surveys (RDS). EOD < RDS reflects that alcohol quantity reported across RDS exceeded the alcohol quantity reported on the EOD summary survey. Frat/Sor = fraternity/sorority. Percentages were computed separately for each row.

Table 3. Average Alcohol Use and Problems as Predictors of Drinking Report discrepancy

	Drinking report discrepancy groups			
Predictors	Low RDS < EOD OR (95% CI)	Low EOD < RDS OR (95% CI)	High RDS < EOD OR (95% CI)	High EOD < RDS OR (95% CI)
AUDIT	**1.04** (1.02, 1.07)	1.01 (0.98, 1.04)	**1.06** (1.02, 1.11)	1.02 (1.00, 1.05)
Alcohol quantity	**1.11** (1.04, 1.19)	1.06 (0.99, 1.14)	**1.23** (1.12, 1.35)	**1.42** (1.33, 1.51)
Leaf	**1.83** (1.42, 2.37)	**1.54** (1.21, 1.95)	**1.94** (1.39, 2.72)	**1.95** (1.55, 2.45)
Concentrate	1.20 (0.83, 1.73)	1.02 (0.71, 1.45)	1.17 (0.72, 1.87)	1.30 (0.95, 1.79)
Edible	1.54 (0.73, 3.23)	1.19 (0.56, 2.49)	1.03 (0.34, 3.13)	1.13 (0.56, 2.27)
Age	1.00 (0.90, 1.10)	0.97 (0.88, 1.07)	0.90 (0.78, 1.04)	0.91 (0.83, 1.00)
Male	1.04 (0.80, 1.36)	0.85 (0.66, 1.09)	0.93 (0.64, 1.36)	**0.74** (0.58, 0.94)
Non-White	0.86 (0.62, 1.21)	1.07 (0.79, 1.44)	0.76 (0.46, 1.26)	0.99 (0.74, 1.34)
School (A)	1.02 (0.74, 1.42)	1.20 (0.89, 1.63)	0.82 (0.50, 1.34)	0.97 (0.72, 1.31)
School (B)	1.18 (0.86, 1.62)	1.22 (0.90, 1.66)	1.29 (0.83, 2.02)	0.98 (0.73, 1.31)
Frat/Sor	0.93 (0.71, 1.23)	0.92 (0.70, 1.20)	1.29 (0.88, 1.91)	0.97 (0.75, 1.25)
Weekend	**1.38** (1.07, 1.78)	**1.34** (1.06, 1.69)	**1.96** (1.43, 2.67)	**1.83** (1.47, 2.26)

n = 2,897. AIC = 7556.10; BIC = 7761.43; −2 log-likelihood = 7179.26. No drinking report discrepancy is the reference group for the 5-level multinomial outcome. "Low" refers to a discrepancy by 1 drink. "High" refers to a discrepancy by 2 or more drinks. RDS < EOD reflects that alcohol quantity reported on the end-of-day (EOD) summary survey exceeded the alcohol quantity reported across repeated daily surveys (RDS). EOD < RDS reflects that alcohol quantity reported across RDS exceeded the alcohol quantity reported on the EOD summary survey. OR = odds ratio. AUDIT = Alcohol Use Disorders Identification Test; Leaf = leaf cannabis use (yes/no); Concentrate = concentrated cannabis use (yes/no); Edible = edible cannabis use (yes/no); Frat/Sor = fraternity/sorority affiliation. AUDIT, alcohol use, and age are grand-mean-centered. Statistically significant effects determined by 95% CIs around the OR that do not overlap the value 1.00 are in bold typeface.

Model 4: Liquor Consumption. Within-person liquor consumption was significantly associated with greater odds of low RDS < EOD, high RDS < EOD, and high EOD < RDS after adjusting for covariates, with comparable effects (see Table 4). Between-person liquor consumption was not significantly linked to drinking report discrepancy categories. Leaf cannabis use was linked to greater odds of all discrepancy categories, and use of cannabis concentrates was associated with greater odds of high EOD < RDS. Weekend (vs. weekday) was related to greater odds of low EOD < RDS, high RDS < EOD, and high EOD < RDS, whereas being younger was linked to lower odds of high EOD < RDS (see Table S4).

DISCUSSION

The present study was the first, to our knowledge, to: (i) quantify both the direction and the magnitude of the discrepancy between alcohol use reported on EOD summary reports (requiring 1 day of retrospection) compared with that reported on RDS (with less retrospection) that reflect the same day of behavior, and (ii) examine alcohol-related predictors of this discrepancy while disentangling between- and within-person effects and adjusting for relevant covariates. Importantly, observations were largely consistent when comparing RDS and EOD summary responses, with 71% of observations varying by 1 drink or less. Of the discrepant observations, findings suggest a discrepancy in both directions (i.e., RDS < EOD vs. EOD < RDS) of varying magnitude (i.e., 1 drink, or a "low" discrepancy vs. 2 or more drinks, or a "high" discrepancy). Consistent with hypotheses, between- and within-person alcohol-related predictors were associated with drinking report discrepancies in both directions. Though considered exploratory, we also found evidence that some predictors were differentially related to the direction of the discrepancy and/or the magnitude of the discrepancy.

Table 4. Between- and Within-Person Alcohol-Related Predictors of Drinking Report Discrepancy

	Drinking report discrepancy groups			
Predictors	Low RDS < EOD OR (95% CI)	Low EOD < RDS OR (95% CI)	High RDS < EOD OR (95% CI)	High EOD < RDS OR (95% CI)
Model 1				
Alcohol use (BP)	**1.17** (1.10, 1.25)	**1.12** (1.05, 1.20)	**1.35** (1.24, 1.47)	**1.56** (1.43, 1.69)
Alcohol use (WP)	1.05 (1.00, 1.11)	**1.27** (1.21, 1.33)	**1.08** (1.02, 1.14)	**1.82** (1.72, 1.92)
Model 2				
Peak rate (BP)	**1.47** (1.27, 1.70)	1.13 (0.97, 1.32)	**1.80** (1.48, 2.19)	**2.25** (1.92, 2.64)
Peak rate (WP)	**1.18** (1.04, 1.32)	**1.64** (1.47, 1.83)	**1.21** (1.06, 1.38)	**2.25** (2.03, 2.49)
Model 3				
PSI (BP)	**1.43** (1.11, 1.83)	**1.25** (1.01, 1.57)	**2.63** (1.80, 3.85)	**2.51** (1.93, 3.27)
PSI (WP)	**1.27** (1.06, 1.53)	**1.53** (1.30, 1.81)	**2.13** (1.70, 2.67)	**2.82** (2.40, 3.31)
Model 4				
Liquor (BP)	0.94 (0.50, 1.76)	1.06 (0.58, 1.93)	1.35 (0.53, 3.48)	0.57 (0.29, 1.11)
Liquor (WP)	**1.96** (1.47, 2.61)	1.20 (0.92, 1.56)	**2.47** (1.73, 3.54)	**2.62** (2.06, 3.34)

n = 2,847. Model 1: AIC = 6726.41; BIC = 6922.03; −2 log-likelihood = 6187.48. Model 2: AIC = 7319.79; BIC = 7495.41; −2 log-likelihood = 6789.61. Model 3: AIC = 7547.17; BIC = 7726.53; −2 log-likelihood = 6955.49. Model 4: AIC = 7701.11; BIC = 7876.72; −2 log-likelihood = 7088.93. No drinking report discrepancy is the reference group for the 5-level multinomial outcome. "Low" refers to a discrepancy by 1 drink. "High" refers to a discrepancy by 2 or more drinks. RDS < EOD reflects that alcohol quantity reported on the end-of-day (EOD) summary survey exceeded the alcohol quantity reported across repeated daily surveys (RDS). EOD < RDS reflects that alcohol quantity reported across RDS exceeded the alcohol quantity reported on the EOD summary survey. OR = odds ratio. WP = within-person (person-centered); BP = between-person (grand-mean-centered). Alcohol use = number of drinks consumed; Peak rate = peak hourly drinking rate on a given day for a given person; PSI = peak subjective intoxication; liquor (binary) = liquor consumption vs. not. Statistically significant effects determined by 95% CIs around the OR that do not overlap the value 1.00 are in bold typeface. Each model (Models 1–4) was examined separately while adjusting for age, sex, school, cannabis use product, and day of the week (weekend vs. weekday). See Tables S1–S4 for covariate effects of each model presented in this table.

Aim 1: Average Alcohol Use and Alcohol Problems

As expected, baseline AUDIT scores and alcohol quantity aggregated across all study days were both associated with drinking report discrepancy. Specifically, AUDIT scores were linked to greater odds of reporting more alcohol consumption on the EOD summary compared with that reported across RDS of both low and high magnitudes. This same finding was not statistically significant in a similar study by Monk and colleagues (2015), though important methodological differences may contribute to these differences in findings, including that Monk and colleagues (2015) examined 1 day of EMA surveys (vs. 54 daily survey days used in the present study) and considered the directionality of the discrepancy, but not the magnitude. Nevertheless, our findings suggest that individuals with more alcohol-related problems as indexed by the AUDIT tend to report more alcohol consumption the next day, which is in line with our understanding of heuristics (Gorin and Stone, 2001; Shiffman et al., 2008). For example, individuals endorsing more alcohol-related problems on the AUDIT may be differentially weighting behaviors and/or experiences when responding to both the AUDIT and the EOD summary given both self-report assessments require a degree of retrospection (at least compared to RDS). Though speculative, individuals who endorse more alcohol-related problems may identify as more problematic drinkers—by extension, heavier alcohol users—and may rely on this heuristic when responding retrospectively about their alcohol involvement.

Average alcohol consumption across the study was linked to all discrepancy categories except for a low EOD < RDS. These findings are consistent with Monk and colleagues (2015) who found that more alcohol consumption was linked to a greater difference score between real-time EMAs of alcohol consumption and EOD summaries. Practically, as the number of drinks consumed increases, the likelihood of a discrepancy between survey types increases. Building off work by Monk and colleagues (2015), our findings provide preliminary evidence that these discrepancies occur in both directions. Though speculative, unexplored moderators (e.g., context of alcohol consumption and context where survey was completed) may further differentiate the relation between heavy drinking and the directionality of the discrepancy.

Aim 2: Between- and Within-Person Alcohol-Related Predictors

After disentangling between- and within-person effects, between-person alcohol quantity exhibited increased odds for each discrepancy category with comparable effect sizes. Within-person alcohol quantity was linked to greater odds of all categories except for low RDS < EOD, with the largest effect for high EOD < RDS. Similar to Monk and colleagues (2015), this suggests that within-person increases in alcohol quantity are more strongly associated with reporting fewer drinks on the EOD summary than across RDS. This same pattern was evident for between- and within-person peak hourly drinking rate. Again, this finding suggests that experiencing greater variability in alcohol consumption relative to an individual's own average is linked to underestimated alcohol consumption reported on a survey requiring more retrospection. Differential relations to the magnitude of discrepancy (i.e., low vs. high) are also consistent with heuristics, suggesting that individuals who experience greater instability in their behavior (i.e., intraindividual variability) tend to be less consistent reporters (i.e., a greater magnitude of discrepancy; Gmel and Daeppen, 2007; Perrine and Schroder, 2005; Solhan et al., 2009).

Between- and within-person peak subjective intoxication exhibited greater odds for all discrepancy categories. At the between-person level, the largest effects were found for a high magnitude of discrepancy (2 or more drinks) in both directions. At the within-person level, the largest effect was found for high EOD < RDS, suggesting fewer drinks were reported on the EOD summary as individuals experience within-person increases in subjective intoxication. This may be due to the subjective nature of rating one's level of intoxication and also may relate to heuristics used in retrospective recall (Gorin and Stone, 2001; Shiffman et al., 2008). For example, within-person variations in peak subjective intoxication inherently suggest instability in behavior, which can result in more inconsistent reporting by recalling some experiences more than others and/or by responding based on one's typical, rather than actual, behavior, which might yield underestimating alcohol consumption on the EOD summary (Gmel and Daeppen, 2007; Perrine and Schroder, 2005; Solhan et al., 2009). By contrast, individuals may differentially weight their experience of peak subjective intoxication and overestimate their alcohol consumption based on 1 or 2 particularly salient experiences (see Shiffman et al., 2008).

Interestingly, within-person, but not between-person, liquor consumption was also associated with all discrepancy categories except for low EOD < RDS. This suggests that inconsistent reporting is more likely to occur in both directions when liquor is consumed that day over and above their average frequency of liquor consumption across the study. In this context, liquor consumption may serve as a proxy for a rapid rise in blood alcohol concentration, which converges with our findings on peak subjective intoxication.

Across all models, weekend days and leaf cannabis use days were most consistently associated with drinking report discrepancies in both directions over and above alcohol-related predictors. This suggests that individuals are less consistent on weekend days and on days when using leaf cannabis, irrespective of their alcohol consumption and/or level of intoxication. This is particularly interesting because many fine-grained, daily studies prioritize capturing data on weekends because alcohol behavior occurs more often on weekends (e.g., Lipperman-Kreda et al., 2017; Stevens et al., 2017; Suffoletto et al., 2018; Wright et al., 2018). Further, much recent work has examined alcohol and cannabis co-use using fine-grained designs (Jackson et al., 2020; Linden-Carmichael et al., 2019; Patrick et al., 2018), and our preliminary findings indicate leaf cannabis use may be impacting the accuracy of alcohol use reporting. More research in this area is needed to determine factors contributing to inconsistent reporting on weekend days and on leaf cannabis use days to increase the validity of substance use reports.

Indeed, the implications of discrepant reports depend on the level of accuracy needed by researchers and clinicians. For example, a participant reporting 8 drinks yet consuming 10 drinks (i.e., discrepancy by 2 drinks) may not unduly affect findings if the outcome of interest is alcohol quantity (i.e., continuous range of drinks). On the other hand, this discrepancy might be more meaningful if the use of the variable is to compute an estimated blood alcohol concentration. Nevertheless, it is difficult to determine the potential impact of discrepant reports and to identify when this level of accuracy matters. (see Jackson, 2008, for a discussion of the predictive utility of drinking thresholds).

Strengths and Limitations

The present study has several strengths, including being a novel investigation of the discrepancy between RDS (with less retrospection) compared with an EOD summary (with more retrospection) spanning 54 days of assessments in a large sample of college students. We are the first to examine both the magnitude and the direction of this discrepancy, which provides a more nuanced understanding of relations. We also disentangled between- and within-person effects for alcohol-related predictors of discrepancy categories (Curran and Bauer, 2011). However, findings should be interpreted considering limitations. First, data were drawn from a larger parent study on college students that required at least 1 occasion of past-month use of alcohol and cannabis at the same time so that their effects overlapped to be eligible for the daily survey phase. Thus, findings from the present study may not generalize beyond college students who use both alcohol and cannabis. Second, the RDS did not capture behavior in real time, as is typical in EMA studies. RDS and EOD summary reports required a degree of retrospective recall, as individuals were asked to recall their behavior since their last completed assessment. Future research in this area should examine discrepancy between retrospective reports and real-time surveys collected using time-stratified random sampling (e.g., Moberly and Watkins, 2008; Monk and Heim, 2014). Third, it is also possible that individuals had difficulty monitoring drinks previously recorded that may have led to double counting of drinks on RDS (Johnson and Schultz, 2005). Replication is needed to determine the reliability of discrepant reports across fine-grained methodologies and across samples before drawing firm conclusions about processes that contribute to this discrepancy. Fourth, days where 2 or more adjacent surveys were missed were excluded from this study, and it is possible these observations also reflected heavier drinking days.

Fifth, following a drinking day, the EOD summary administered at 9:00 am did not re-assess if alcohol was consumed the prior day if a participant had indicated on any RDS that they had consumed alcohol; thus, participants could not recant their alcohol use behavior from the prior day. Future research is needed to determine factors that influence recanting on an EOD summary following endorsement of substance use behavior on real-time surveys. Sixth, of the observations included in the analyses, only 2.85% of observations included endorsing alcohol-related blackouts. Future studies with a higher daily prevalence of blackout endorsement should examine whether this contributes to response inconsistency intraindividually, as experiencing a blackout could affect memory and contribute to discrepancies between reports. Seventh, cannabis use products, rather than cannabis use quantity, were included as covariates, given the likelihood of discrepant reporting for cannabis use, as we show for alcohol use reporting. This may be compounded for cannabis use because there are no established standard units for cannabis use as there are for alcohol use. However, cannabis use quantity may also impact discrepant alcohol reporting, given the established link between cannabis use and acute memory impairment (Volkow et al., 2016). Future research should further examine how cannabis co-use affects reporting consistency of substance use. Eighth, it is possible that individuals misperceive their own level of drunkenness, such that their subjective intoxication does not reflect their actual intoxication.

We did not consider context of the drinking event as a predictor of discrepant reports given RDS were aggregated to the daily level, which hindered our ability to discern which context of the drinking event was linked to inconsistent reporting. Future studies should consider context as a predictor of inconsistent reporting, as Monk and colleagues (2015) found that, when comparing alcohol reported on weekly retrospective surveys to real-time surveys, the situational context (e.g., being at a bar) and the social context (e.g., being with 2 or more friends) significantly related to discrepant reports.

CONCLUSIONS

In sum, more than half of all observations were discrepant, such that the number of drinks reported across RDS was not in line with the number of drinks reported on the EOD summary report the next morning. On the other hand, almost three-fourths of all comparisons of RDS to EOD summaries were consistent within 1 drink. Findings suggest that surveys requiring more retrospection may overestimate alcohol consumption for individuals with more problematic alcohol behavior. Findings also support that greater within-person fluctuations in alcohol behavior is linked to less consistent reporting overall and suggest that a variety of memory heuristics may be at play. Future research is needed to discern the optimal use and timing of survey types administered in fine-grained survey designs to maximize the validity of reports.

ACKNOWLEDGMENTS

This work was supported by the National Institute on Drug Abuse (R01 DA040880, MPIs: Jackson and White; T32 DA016184, PI: Rohsenow) and the National Institute on Alcohol Abuse and Alcoholism (K23 AA024808, PI: Treloar).

CONFLICTS OF INTEREST

The authors have no conflicts of interest to declare.

REFERENCES

Babor TF, Higgins-Biddle J, Saunders J, Monteiro M. (2001) The alcohol use disorders identification test: Guidelines for use in primary health care. World Health Organization – Department of Mental Health and Substance Dependence.

Carney MA, Armeli S, Tennen H, Affleck G, O'Neil TP (2000) Positive and negative daily events, perceived stress, and alcohol use: a diary study. J Consult Clin Psychol 68:788–798.

Carney MA, Tennen H, Affleck G, Del Boca FK, Kranzler HR (1998) Levels and patterns of alcohol consumption using timeline follow-back, daily diaries and real-time "electronic interviews". J Stud Alcohol 59:447–454.

Cumming G, Finch S (2005) Inference by eye confidence intervals and how to read pictures of data. Am Psychol 60:170–180.

Curran PJ, Bauer DJ (2011) The disaggregation of within-person and between-person effects in longitudinal models of change. Annu Rev Psychol 62:583–619.

Dulin PL, Alvarado CE, Fitterling JM, Gonzalez VM (2017) Comparisons of alcohol consumption by timeline follow back vs. smartphone-based daily interviews. Addict Res Theory 25:195–200.

Feunekes GI, van'tVeer P, vanStaveren WA, Kok FJ (1999) Alcohol intake assessment: The sober facts. Am J Epidemiol 150(1):105–112.

Glovannucci E, Colditz G, Stampfer MJ, Rimm EB, Litin L, Sampson L, Willett WC (1991) The assessment of alcohol consumption by a simple self-administered questionnaire. Am J Epidemiol 133:810–817.

Gmel G, Daeppen JB (2007) Recall bias for seven-day recall measurement of alcohol consumption among emergency department patients: implications for case-crossover designs. J Stud Alcohol Drugs 68:303–310.

Godden DR, Baddeley AD (1975) Context-dependent memory in two natural environments: on land and underwater. Br J Psychol 66:325–331.

Gorin AA, Stone AA. (2001) Recall biases and cognitive errors in retrospective self-reports: A call for momentary assessments, in *Handbook of Health Psychology* (Baum A, Revenson TA, Singer JE eds.), pp. 405–413. Erlbaum, Mahwah, NJ.

Griffith SD, Shiffman S, Heitjan DF (2009) A method comparison study of timeline followback and ecological momentary assessment of daily cigarette consumption. Nicotine Tob Res 11:1368–1373.

Hedeker D. (2005) Generalized linear mixed models, in Encyclopedia of Statistics in Behavioral Science. John Wiley & Sons Ltd, Chichester, UK.

Hox J, Moerbeek M, Van de Schoot R (2017) Multilevel Analysis: Techniques and Applications, 3rd edn. Taylor & Francis, New York, NY.

Hufford MR. (2007) Special methodological challenges and opportunities in ecological momentary assessment. in The Science of Real-Time Data Capture: Self-Reports in Health Research (Stone AA, Shiffman S, Atienza AA, Nebeling L eds.), pp. 54–76. Oxford University Press, Oxford, UK.

Jackson KM (2008) Heavy episodic drinking: determining the predictive utility of five or more drinks. Psychol Addict Behav 22:68–77.

Jackson KM, Sokolovsky AW, Gunn RL, White HR (2020) Consequences of alcohol and marijuana use among college students: prevalence rates and attributions to substance-specific versus simultaneous use. Psychol Addict Behav 34:370–381.

Johnson EO, Schultz L (2005) Forward telescoping bias in reported age of onset: an example from cigarette smoking. Int J Methods Psychiatr Res 14:119–129.

Jones A, Remmerswaal D, Verveer I, Robinson E, Franken IHAA, Wen CKFF, Field M (2019) Compliance with ecological momentary assessment protocols in substance users: a meta-analysis. Addiction 114:609–619.

Kiene SM, Barta WD, Tennen H, Armeli S (2009) Alcohol, helping young adults to have unprotected sex with casual partners: findings from a daily diary study of alcohol use and sexual behavior. J Adolesc Heal 44:73–80.

Leigh BC, Gillmore MR, Morrison DM (1998) Comparison of diary and retrospective measures for recording alcohol consumption and sexual activity. J Clin Epidemiol 51(2):119–127.

Linden-Carmichael AN, Stamates AL, Lau-Barraco C (2019) Simultaneous use of alcohol and marijuana: patterns and individual differences. Subst Use Misuse 54:2156–2166.

Lipperman-Kreda S, Gruenewald PJ, Grube JW, Bersamin M (2017) Adolescents, alcohol, and marijuana: Context characteristics and problems associated with simultaneous use. Drug Alcohol Depend 179:55–60.

Maas CJM, Hox JJ (2005) Sufficient sample sizes for multilevel modeling. Methodology 1:86–92.

Mathieu JE, Aguinis H, Culpepper SA, Chen G (2012) Understanding and estimating the power to detect cross-level interaction effects in multilevel modeling. J Appl Psychol 97:951–966.

Moberly NJ, Watkins ER (2008) Ruminative self-focus and negative affect: an experience sampling study. J Abnorm Psychol 117:314–323.

Monk RL, Heim D (2014) A real-time examination of context effects on alcohol cognitions. Alcohol Clin Exp Res 38:2454–2459.

Monk RL, Heim D, Qureshi A, Price A (2015) "I have no clue what I drunk last night" using smartphone technology to compare in-vivo and retrospective self-reports of alcohol consumption. PLoS One 10: e0126209.

Nakagawa S, Johnson PC, Schielzeth H (2017) The coefficient of determination R 2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. J R Soc Interface 14(134): 1–11.

NIAAA (2007) Helping patients who drink too much: A clinician's guide: updated 2005 edition (no. 7). US Department of Health and Human Services, National Institutes of Health, National Institute on Alcohol Abuse and Alcoholism.

O'Grady MA, Cullum J, Tennen H, Armeli S (2011) Daily relationship between event-specific drinking norms and alcohol use: a four-year longitudinal study. J Stud Alcohol Drugs 72:633–641.

Patrick ME, Kloska DD, Terry-McElrath YM, Lee CM, O'Malley PM, Johnston LD (2018) Patterns of simultaneous and concurrent alcohol and marijuana use among adolescents. Am J Drug Alcohol Abuse 44:441–451.

Patterson C, Hogan L, Cox M (2019) A comparison between two retrospective alcohol consumption measures and the daily drinking diary method with university students. Am J Drug Alcohol Abuse 45:248–253.

Perrine MWB, Schroder KEE (2005) How many drinks did you have on September 11, 2001? J Stud Alcohol 66:536–544.

R Core Team (2013) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna. http://www.R-project.org/

Raudenbush SW, Bryk AS (2002) Hierarchical Linear Models. 2nd ed. SAGE, Thousand Oaks, CA.

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 uHealth 4:e44.

Shiffman S (2009) Ecological momentary assessment (EMA) in studies of substance use. Psychol Assess 21:486–497.

Shiffman S, Hufford M, Hickcox M, Paty JA, Gnys M, Kassel JD (1997) Remember that? A comparison of real-time versus retrospective recall of smoking lapses. J Consult Clin Psychol 65:292–300.

Shiffman S, Stone AA, Hufford MR (2008) Ecological momentary assessment. Annu Rev Clin Psychol 4:1–32.

Simons JS, Wills TA, Emery NN, Marks RM (2015) Quantifying alcohol consumption: self-report, transdermal assessment, and prediction of dependence symptoms. Addict Behav 50:205–212.

Simpura J, Poikolainen K (1983) Accuracy of retrospective measurement of individual alcohol consumption in men; a reinterview after 18 years. J Stud Alcohol 44:911–917.

Singer JD (1998) Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. J Educ Behav Stat 24:323–355.

Sokolovsky AW, Gunn RL, Micalizzi L, White HR, Jackson KM (2020) Alcohol and marijuana co-use: Consequences, subjective intoxication, and the operationalization of simultaneous use. Drug Alcohol Depend. https://doi.org/10.1016/j.drugalcdep.2020.107986

Solhan MB, Trull TJ, Jahng S, Wood PK (2009) Clinical assessment of affective instability: comparing EMA indices, questionnaire reports, and retrospective recall. Psychol Assess 21:425–436.

Sommet N, Morselli D (2017) Keep calm and learn multilevel logistic modeling: a simplified three-step procedure using Stata, R, Mplus, and SPSS. Int Rev Soc Psychol 30:203–218.

Stevens AK, Littlefield AK, Talley AE, Brown JL (2017) Do individuals higher in impulsivity drink more impulsively? A pilot study within a high risk sample of young adults. Addict Behav 65:147–153.

Stone AA, Shiffman S (1994) Ecological momentary assessment (EMA) in behavioral medicine. Ann Behav Med 16:199–202.

Stone AA, Shiffman S (2002) Capturing momentary, self-report data: a proposal for reporting guidelines. Ann Behav Med 24:236–243.

Suffoletto B, Chung T, Muench F, Monti P, Clark DB (2018) A text message intervention with adaptive goal support to reduce alcohol consumption among non-treatment-seeking young adults: non-randomized clinical trial with voluntary length of enrollment. JMIR mHealth uHealth 6:e35.

Trull T, Ebner-Priemer UW (2009) Using experience sampling methods/ecological momentary assessment (ESM/EMA) in clinical assessment and clinical research: introduction to the special section. Psychol Assess 21:457–462.

Volkow ND, Swanson JM, Evins AE, DeLisi LE, Meier MH, Gonzalez R, Bloomfield MAP, Curran HV, Baler R (2016) Effects of cannabis use on human behavior, including cognition, motivation, and psychosis: a review. JAMA Psychiatry. 73:292–297.

White HR, Kilmer JR, Fossos-Wong N, Hayes K, Sokolovsky AW, Jackson KM (2019) Simultaneous alcohol and marijuana use among college students: patterns, correlates, norms, and consequences. Alcohol Clin Exp Res 43:1545–1555.

Wray TB, Kahler CW, Monti PM (2016) Using ecological momentary assessment (EMA) to study sex events among very high-risk men who have sex with men (MSM). AIDS Behav 20(10):2231–2242.

Wray TB, Merrill JE, Monti PM (2014) Using ecological momentary assessment (EMA) to assess situation-level predictors of alcohol use and alcohol-related consequences. Alcohol Res 36:19–27.

Wright C, Dietze PM, Agius PA, Kuntsche E, Livingston M, Black OC, Room R, Hellard M, Lim MS (2018) Mobile phone-based ecological momentary intervention to reduce young adults' alcohol use in the event: a three-armed randomized controlled trial. JMIR mHealth uHealth 6:e149.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Fig. S1. Flowchart from screening to the daily survey phase of the parent study.

Fig. S2. Flowchart of data selection from the full daily sample to the restricted daily sample used in analyses.

Fig. S3. Drinking report discrepancy frequencies by category.

Table S1. Between- and within-person alcohol quantity as predictors of drinking report discrepancy.

Table S2. Between- and within-person peak hourly drinking rate as predictors of drinking report discrepancy.

Table S3. Between- and within-person peak subjective intoxication as predictors of drinking report discrepancy.

Table S4. Between- and within-person liquor consumption as predictors of drinking report discrepancy.

Appendix S1. Eligibility and recruitment for the parent study.