








# Ecological Momentary Assessment of Alcohol Consumption and Its Concordance with Transdermal Alcohol Detection and Timeline Follow-Back Self-report Among Adults Experiencing Homelessness

Eun-Young Mun , Xiaoyin Li , Michael S. Businelle , Emily T. Hébert ,  
Zhengqi Tan , Nancy P. Barnett , and Scott T. Walters 

**Background:** Studies of alcohol use presume valid assessment measures. To evaluate this presumption, we examined the concordance of alcohol use as measured by ecological momentary assessment (EMA) self-reports, transdermal alcohol concentration readings via the Secure Continuous Remote Alcohol Monitor (SCRAM), and retrospective self-reports via the Timeline Follow-Back (TLFB) among adults experiencing homelessness.

**Methods:** Forty-nine adults who reported alcohol misuse (mean age = 47, SD = 9; 57% Black; 82% men) were recruited from a homeless shelter. For 4 weeks, alcohol use was assessed: (i) 5 times or more per day by EMA, (ii) every 30 minutes by a SCRAM device worn on the ankle, and (iii) by TLFB for the past month at the end of the study period. There were 1,389 days of observations of alcohol use and alcohol use intensity for 49 participants.

**Results:** EMA and SCRAM alcohol use data agreed on 73% of days, with an interrater agreement Kappa = 0.46. A multilevel analysis of concordance of 3 measures for alcohol use yielded statistically significant correlations of 0.40 (day level) and 0.63 (person level) between EMA and SCRAM. Alcohol use was detected on 49, 38, and 33% of days by EMA, SCRAM, and TLFB, respectively. For alcohol use intensity, EMA and SCRAM resulted in statistically significant correlations of 0.46 (day level) and 0.78 (person level). The concordance of TLFB with either EMA or SCRAM was weak, especially at the day level.

**Conclusions:** This is the first study to examine concordance of alcohol use estimates using EMA, SCRAM, and TLFB methods in adults experiencing homelessness. EMA is a valid approach to quantifying alcohol use, especially given its relatively low cost, low participant burden, and ease of use. Furthermore, any stigma associated with wearing the SCRAM or reporting alcohol use in person may be attenuated by using EMA, which may be appealing for use in studies of stigmatized and underserved populations.

**Key Words:** Ecological Momentary Assessment, Timeline Follow-Back, Transdermal Alcohol Sensor, Alcohol Assessment, Homelessness.

From the Department of Health Behavior and Health Systems (EYM, XL, STW), University of North Texas Health Science Center, Fort Worth, Texas; Health Promotion Research Center (MSB), Stephenson Cancer Center, Oklahoma City, Oklahoma; Family and Preventive Medicine (MSB), University of Oklahoma Health Sciences Center, Oklahoma City, Oklahoma; UTHealth School of Public Health in Austin (ETH), Austin, Texas; Department of Biostatistics and Epidemiology (ZT), University of North Texas Health Science Center, Fort Worth, Texas; Center for Alcohol and Addiction Studies (NPB), School of Public Health, Brown University, Providence, Rhode Island.

Received for publication October 2, 2020; accepted February 3, 2021.

Correspondence: Eun-Young Mun, PhD, Department of Health Behavior and Health Systems, School of Public Health, University of North Texas Health Science Center, 3500 Camp Bowie Blvd., EAD 709, Fort Worth, TX 76107-2699; E-mail: eun-young.mun@unthsc.edu

© 2021 The Authors. *Alcoholism: Clinical & Experimental Research* published by Wiley Periodicals LLC on behalf of Research Society on Alcoholism.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

DOI: 10.1111/acer.14571

STUDIES OF ALCOHOL use presume valid assessment measures. Though objective measures exist, alcohol intervention trials have primarily used self-report methods to estimate drinking outcomes (Carvalho et al., 2019; Witkiewitz et al., 2015b). This is mostly due to the convenience and low cost of self-reported drinking measures, compared with direct observational or biological measures. However, self-report is not without limitations. It is difficult to blind participants and treatment providers in behavioral intervention trials (Witkiewitz et al., 2015a). When participants are fully aware of the goal of a trial and their treatment allocation, biased reporting by participants (i.e., performance bias) is more likely due to participant demand characteristics and social desirability, a major source of outcome bias in randomized clinical trials (Higgins and Green, 2011). The lack of objective measures of alcohol use, therefore, can make the interpretation of effect sizes from clinical trials challenging, leading to a lower grade of evidence (Guyatt et al., 2011) for alcohol intervention trials (Grant et al., 2015).

Alcohol-related mortality among people experiencing homelessness is 6 to 10 times greater than the general population (Baggett et al., 2015), and about 38% of adults experiencing homelessness are alcohol dependent (Fazel et al., 2008), with 31% to 39% involved in high-risk drinking (Neisler et al., 2019; Reitzel et al., 2020; Taylor et al., 2016). As a result, individuals experiencing homelessness have a significantly greater need for alcohol use treatment than the general population. Despite this greater treatment need, existing alcohol treatments may not optimally engage this population in treatment (Orwin et al., 1999; Wenzel et al., 2001; see also Collins et al., 2019). For instance, a recent systematic review of treatment studies for adults with alcohol misuse who are experiencing homelessness found a limited evidence base, with only 17 studies meeting the inclusion criteria (Adams-Guppy and Guppy, 2016). The almost exclusive use of self-reported alcohol outcomes in these studies was raised as a potential measurement concern (Adams-Guppy and Guppy, 2016). Furthermore, alcohol misuse has long been stigmatized (Kulesza et al., 2014), and this stigma can hinder treatment seeking, treatment adherence and retention, and treatment outcomes (Carvalho et al., 2019), as well as self-report accuracy.

In treatment trials for alcohol use disorders, the most frequently used outcome measures have been frequency or quantity of alcohol consumption derived from the timeline follow-back method (TLFB; Sobell and Sobell, 1995). In this method, people are asked to recall the number of drinks they had on each day during a specific period (e.g., past 30 days) using a calendar and anchor events (e.g., weekends, holidays) as recall aids. The TLFB is easy to administer and has relatively good psychometric characteristics with various populations, including homeless individuals (McKenna et al., 2018; Sacks et al., 2003; Sobell and Sobell, 1995). However, when using retrospective self-report measures as the sole outcome measure, this may lead to recall bias due to having to recall drinking over relatively long periods (see Piasecki, 2019, for a review). For instance, as the time interval increases, people tend to smooth out the variation in alcohol consumed on each day and report less drinking overall (Hoepfner et al., 2010; Searles et al., 2002). Moreover, the TLFB's accuracy can be influenced by a person's impaired ability to recall alcohol consumption from memory, which may be more prevalent among people experiencing homelessness due to the high prevalence of comorbid mental and substance use disorders (Hurstak et al., 2017; Stasiewicz et al., 2008; Stone et al., 2019).

Recent studies suggest that exclusive reliance on the TLFB self-report outcomes in clinical trials may be problematic for detecting abstinence. For example, when the TLFB self-reports of alcohol abstinence were compared with eHealth biomarker and breath alcohol content (BrAC) data, 34% to 68% of the self-reports of alcohol abstinence were contradicted by positive readings by BrAC and biomarker data (Hämäläinen et al., 2020). Similarly, a large discrepancy between the TLFB reports of drinking and the Secure

Continuous Remote Alcohol Monitor (SCRAM) (47% vs. 92%, respectively) was reported in a study (Alessi et al., 2019) that analyzed data from 2 outpatient clinical trials.

Compared with a retrospective TLFB approach, there is evidence that more frequent assessments using interactive voice response, daily diary assessment, or ecological momentary assessment (EMA) are more accurate in estimating whether drinking occurred on a given day (Kaplan and Koffarnus, 2019; Tucker et al., 2007). For example, Kaplan and Koffarnus (2019) reported that the agreement between the past-day self-reports and breathalyzer data (92% and 71% in raw agreement for their treatment and control groups, respectively) was better than the agreement between the 30-day TLFB and breathalyzer data (71% in raw agreement for both groups). With respect to overall alcohol consumption level, evidence appears inconclusive and may be explained in part by participant and study characteristics (e.g., Kaplan and Koffarnus, 2019; Searles et al., 2000; Tucker et al., 2007).

Ecological momentary assessment protocols sample people's experiences within days using either a time-based assessment protocol and/or an event-based assessment protocol (Piasecki, 2019). Reports of alcohol consumption assessed by EMA tend to give higher overall consumption levels than reports assessed by TLFB (e.g., Dulin et al., 2017), although more data are needed to better characterize the relative under- or overreporting of the 2 methods. Notably, EMA can detect the dynamic nature of drinking over a relatively short period of time (Shiffman, 2009) as well as factors, such as urge, negative affect, and other situational variables that precede drinking at a particular moment. Furthermore, EMA has become increasingly feasible and attractive via applications installed on smartphones (Piasecki, 2019; Wray et al., 2014). However, to our best knowledge, the validity of EMA to measure alcohol use has not been studied among adults experiencing homelessness.

The present study compares 3 methods of alcohol assessment among adults with alcohol misuse who are experiencing homelessness: self-reported alcohol use measured several times each day by EMA, self-reported alcohol use measured at a single time point by TLFB, and biologically detected alcohol use measured by a transdermal alcohol sensor worn during the study period (Leffingwell et al., 2013). The SCRAM (Alcohol Monitoring Systems, Inc.) has been used successfully in alcohol administration studies in laboratory settings (Dougherty et al., 2012; Fairbairn and Kang, 2019; Hill-Kapturczak et al., 2014; Hill-Kapturczak et al., 2015; Roache et al., 2015), in lab and field settings (Fairbairn et al., 2019; Roache et al., 2019), and in intervention trials (Alessi et al., 2019; Barnett et al., 2017; Dougherty et al., 2014; Dougherty et al., 2015; Rash et al., 2019). The SCRAM has also been utilized in alcohol treatment trials to verify and quickly reinforce alcohol abstinence or reductions using contingency management (e.g., Barnett et al., 2011).

Although studies note a number of drawbacks to using the SCRAM, including equipment failures, equipment loss or

tampering, discomfort, and challenges related to detecting lower-level drinking and interpreting transdermal alcohol concentration (TAC) data (Barnett et al., 2014; Roache et al., 2015; Roache et al., 2019), correspondence between self-reported alcohol use and SCRAM-assessed TAC appears to be adequate, and the SCRAM tends to be reasonably well tolerated by participants (see van Egmond et al., 2020; Piasecki, 2019 for reviews). For example, in a sample of soup kitchen attendees who reported heavy drinking, drinking days assessed by daily self-report (e.g., “. . .since I saw you yesterday?”) and TAC readings were correlated at 0.68 at the person-level (Rash et al., 2019). Similarly, a recent systematic review reported that TAC data from 3 biological detection methods (SCRAM, WristTAS, and Skyn) had a correlation of 0.62 with self-reported alcohol use (van Egmond et al., 2020).

The current study examines the extent to which TAC readings from adults with alcohol misuse who are experiencing homelessness correspond with self-reported alcohol use via EMA and TLFB. Given the challenges of conducting studies with this population and engaging them in clinical trials, the present study attempts to fill this knowledge gap by measuring the concordance between EMA- and TLFB-based self-reports of alcohol use, compared with continuously monitored TAC readings from the SCRAM.

## MATERIALS AND METHODS

### *Participants*

Data for the present study come from participants who participated in Phase I of an intervention development study (NCT03746808) conducted at a large homeless shelter in Dallas, TX. The North Texas Regional Institutional Review Board (IRB) approved the research protocol. The detailed study design, eligibility criteria, and recruitment procedure can be found elsewhere (Businelle et al., 2020). Briefly, participants were eligible if they scored an 8 or higher on the Alcohol Use Disorders Identification Test (AUDIT; Saunders et al., 1993); reported consuming at least 1 standard drink of alcohol in the past week; were concurrently receiving health services at the homeless shelter; were able to read English at the 7th-grade level; and did not have any physical or mental disabilities that could prevent them from participating. If they reported health conditions that might prohibit them from wearing the SCRAM (e.g., diabetes, swelling, nickel allergy), the study required that they be cleared by a medical professional at the homeless shelter's health clinic prior to enrollment.

Data come from the 49 participants who completed the baseline visit and 4-week follow-up visit, wore and returned the SCRAM device, and completed daily EMA surveys (see Fig. 1). Table 1 describes the sample's baseline characteristics. Participants were between the ages of 25 and 62 years ( $M = 47$ ,  $SD = 9$ ), 37% White, 57% Black, and 82% male. The median lifetime homelessness was 2.5 years (interquartile range [IQR] = 3.4 years) and 1.5 years for the current homeless episode (IQR = 3.2 years). Participants spent the night at the homeless shelter for an average of 23 days out of the past 30 days, followed by outside (3.2 days), at a friend's home (1 day), or at a relative's home (0.9 days). At baseline, participants reported using alcohol on 14.4 days in the past 30 days, 5.8 of which were heavy drinking days (4 drinks or more for women; 5 drinks or more for men).

The majority of participants reported that they were primarily diagnosed with Alcohol Use Disorder (59%), Cannabis Use Disorder (22%), or Cocaine/Opioid/Amphetamine Use Disorder (18%) in their lifetime. Their comorbid mental health conditions included Depression (71%), Anxiety Disorders (39%), Bipolar Disorder (51%), Posttraumatic Stress Disorder (37%), and Schizophrenia or Schizoaffective disorder (27%). All participants in the current study had a prescription for physical health conditions, such as heart disease, lung or pulmonary disease, cancer, high blood pressure, and diabetes. Participants were receiving a range of services at the homeless shelter, including crisis management, mental health counseling or medications, substance abuse counseling, smoking cessation counseling or medications, and legal aid.

With few exceptions, people who dropped out of the study ( $n = 29$ ; 37%) before the 4-week visit were not statistically different from those who completed the study. Black participants were more likely to drop out than non-Black participants ( $p = 0.047$ ). In addition, the prevalence of lifetime Schizophrenia or Schizoaffective Disorder among those who completed the study was lower, compared with those who did not complete the 4-week visit (see Table 1).

### *Procedures*

Data collection took place from 2/12/2019 to 2/27/2020. Participants completed structured interviews, including the TLFB, at baseline and 4 weeks. Approximately 5 days after the baseline assessment, participants completed an equipment setup visit where they were provided with a smartphone (Samsung Galaxy S3 or S7) preloaded with an app developed using the Insight<sup>TM</sup> mHealth platform (<https://otrc.stephensoncancercenter.org/Mobile-Health-Tech-Technology>) and fitted with the SCRAM “CAM” (Continuous Alcohol Monitoring) on the ankle. At 2 weeks, participants completed a brief equipment visit to verify the proper functioning of devices and receipt of SCRAM data. Participants received \$25 for completing the baseline assessment, \$25 for completing the 4-week assessment, up to \$25 each week for completing EMAs (based on percent completed), and \$25 for returning the phone and SCRAM in good condition at the end of the study.

### *Alcohol Assessment Methods*

During the 4-week study period, alcohol consumption was measured in 3 ways: (i) self-reported EMA via the smartphone app that prompted surveys once in the morning (30 minutes after the participant's self-reported wake time) and at random times during 4 epochs of the day (participants were also expected to self-initiate additional assessments if they started to drink); (ii) transdermal alcohol data passively collected every 30 minutes via the SCRAM worn on the ankle; and (iii) retrospective recall of daily alcohol use via the TLFB administered at the 4-week follow-up visit. Data sampling frequencies were up to 5 times or more per day for the EMA data, up to 48 times per day for the SCRAM data, and once per day for the TLFB data. Transdermal alcohol sensor data were transferred to a secure server administered by Alcohol Monitoring Systems (AMS) and downloaded through the SCRAMNet, a secure web interface. Given that the common sampling frequency of alcohol use across 3 assessment methods was at the day level, we aggregated intraday data from the EMA and SCRAM into day-level data.

### *Measures*

*Alcohol use.* For EMA, alcohol use on a given day was determined based on any positive answer to the following questions: (i) “Have you consumed alcohol today?”, which was scheduled up to 5 times per day (random EMAs); (ii) “Did you just drink or are you

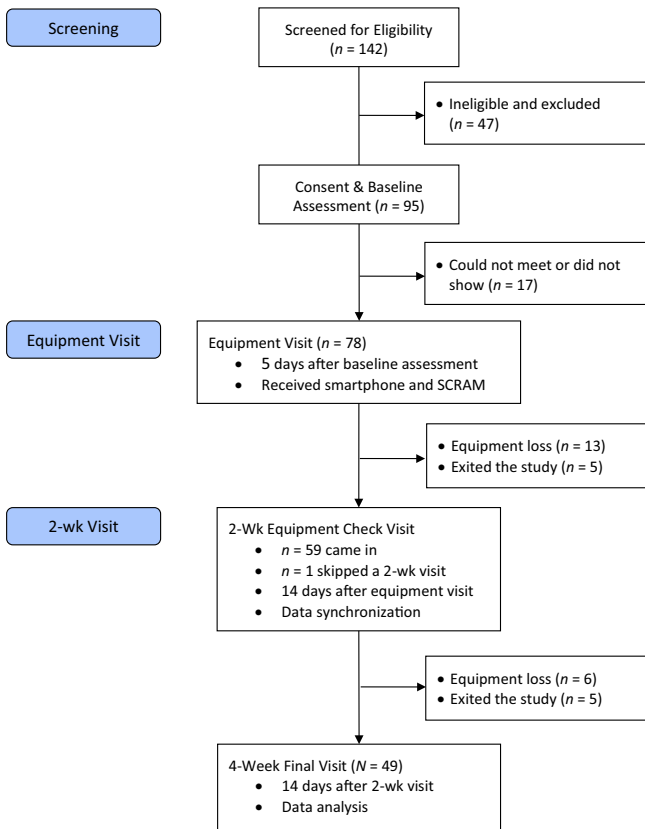


Fig. 1. Flowchart of the participants in the current study.

about to drink alcohol?”, which was asked when participants self-initiated drinking-related EMAs; and (iii) “Did you drink any alcohol yesterday?”, which was asked during the daily morning assessment about drinking the previous day (i.e., morning diary). If there was a positive answer to the morning diary question, we counted the preceding day as a drinking day. In the absence of a valid positive answer for the morning diary question, alcohol use was coded as 0 only when participants answered “No” to all EMA questions (random EMAs without any self-initiated drinking-related EMAs) the previous day. If not (i.e., some “No” responses and missing EMA responses within the day without the next morning diary answer), alcohol use was coded as missing.

For the SCRAM data, alcohol consumption on a given day was detected via TAC readings. All TAC estimates and related values were processed using the Transdermal Alcohol Sensor Data Macro (TASMAC version 1.5; Barnett et al., 2015). A drinking episode was determined by a minimum TAC = 0.02 g/dl with the maximum absorption rate of 0.05% or the maximum elimination rate below 0.025% if peak TAC < 0.15 g/dl or below < 0.035% if peak TAC > 0.15 g/dl. Any positive drinking episode detected within 24 hours was recorded as 1 (1 = detected alcohol use or 0 = did not detect alcohol use).

All TAC data were analyzed based on a noon-to-noon 24-hour period. A positive drinking episode detected prior to noon on a given day was counted toward a previous day drinking episode because there is a longer time-to-peak TAC with a longer TAC descending curve than blood or breath alcohol concentrations (Karns-Wright et al., 2016). A noon-to-noon 24-hour period, compared to a midnight-to-midnight calendar day, is commonly used to assess daily-level drinking in natural settings (see Roache et al., 2019).

From the 30-day TLFB reports, if the participant recalled consuming any standard drinks on a given day, it was coded as 1 to indicate a drinking day; a nondrinking day was coded as 0.

Table 1. Means (Standard Deviations) and Percentages at Baseline between the Current Sample of Participants who Completed 4-week Follow-up, Compared with Those who did not

Variable	The Current Sample (N = 49)	Did not Complete 4 Weeks (N = 29)	Chi-square test or t-test	p-value
Age in years	47.0 (9.0)	45.9 (9.7)	0.50	0.62
Man (1; 0 = Woman)	81.6%	89.7%	0.90	0.34
Black (1; 0 = Other)	57.1%	79.3%	3.96	0.05
AUDIT <sup>a</sup> Score	20.7 (7.6)	20.8 (6.9)	-0.04	0.97
Lifetime homeless in months <sup>b</sup>	48.5 (60.2)	66.1 (54.1)	-1.29	0.20
Current homeless in months <sup>b</sup>	34.9 (43.6)	34.1 (44.1)	0.09	0.93
Number of days at homeless shelter in the past 30 days	23.0 (9.5)	25.7 (7.0)	-1.31	0.19
Drinking days in the past month <sup>c</sup>	14.4 (8.8)	17.9 (8.3)	-1.76	0.08
Drinks per day in the past month <sup>c</sup>	2.3 (1.9)	3.2 (2.0)	-1.98	0.05
Heavy drinking days in the past month <sup>c</sup>	5.8 (7.2)	8.1 (9.1)	-1.20	0.24
Lifetime Depression <sup>d</sup>	71%	90%	3.55	0.06
Lifetime Anxiety Disorders <sup>d</sup>	39%	52%	1.24	0.27
Lifetime Bipolar Disorder <sup>d</sup>	51%	59%	0.42	0.52
Lifetime Posttraumatic Stress Disorder <sup>d</sup>	37%	21%	2.20	0.14
Lifetime Schizophrenia or Schizoaffective Disorder <sup>d</sup>	27%	55%	6.40	0.01

Notes. The degrees of freedom for chi-square tests and t-tests were 1 and 76, respectively.

<sup>a</sup>AUDIT = The Alcohol Use Disorders Identification Test.

<sup>b</sup>The median lifetime and current homeless experience were 30 months and 18 months, respectively, with interquartile range (IQR) = 40.5 and 38.5 months, respectively, for the current sample. For those who did not complete all 4 weeks of the study, the median lifetime and current homeless experience were 54 months and 18 months, respectively, with respective IQR = 87.5 and 44 months.

<sup>c</sup>Timeline Follow-Back data at baseline, including the number of days using alcohol, average drinks per day, and the number of heavy drinking days (having 4 or more drinks per day for women and 5 or more drinks per day for men) in the past 30 days.

<sup>d</sup>All lifetime diagnoses were self-reported at baseline based on a single item (e.g., Have you ever been diagnosed with Depression?).

*Alcohol use intensity.* Alcohol use intensity was estimated by peak TAC from the TASMACH version 1.5 (Barnett et al., 2015). Peak TAC is a measure of drinking event intensity (Leffingwell et al., 2013). Non-zero peak TAC values were recoded as zero if no verified drinking episode was detected based on the prespecified criteria (see Alcohol use).

For EMA and TLFB, alcohol use intensity was estimated by the number of standard drinks reported each day. For EMA, people reported the number of drinks consumed during the morning assessment on each day (“How many standard drinks did you have yesterday?”) with a dropdown option ranging from 0 to 20. On the TLFB, people reported the number of drinks consumed on each day of the past month. Participants were asked to describe what they drank each day, and a research staff member recorded this answer in terms of standard drinks: a 12-ounce beer, a 5-ounce glass of wine, or a shot of liquor.

### Analysis

We first examined the extent of bivariate concordance among EMA, SCRAM, and TLFB measures regarding whether drinking occurred at the day-level in interrater agreement analyses. We subsequently examined the extent of correspondence among the 3 methods of assessing alcohol use and alcohol use intensity in multilevel analyses. Multilevel models take into account the clustered nature of data (i.e., days within persons) of the current study. The average cluster size was 28.3 days per person. With this multilevel modeling approach, day-level correlations are down-adjusted due to correcting design effects, compared with bivariate correlations in a single-level analysis, which ignores that day-level data are related to each other within persons. In addition, missing data are better accommodated in multilevel models. Finally, the associations between demographic covariates and 3 measures of alcohol intensity were examined in a “complex survey” structural equation model to account for data clustering. Analyses were performed with SAS version 9.4 (SAS Institute, Inc.), SPSS version 25 (IBM SPSS), and Mplus version 8.5 (Muthén and Muthén, 1998–2020).

## RESULTS

### Assessment Protocol Adherence

Overall, 91.8% of participants in this sample wore the SCRAM bracelet for 28 days, with a few participants wearing the bracelet up to 31 days for a total of 1,384 days of observation. During the study period, 23 study phones were lost or irreparably damaged, and 8 bracelets were lost by the 78 participants who completed the initial equipment setup visit (see Fig. 1). This number includes participants who were lost to follow-up, which most times resulted in equipment loss as well. Of those who completed the 4-week follow-up (the current sample), 2 phones were lost. The costs of administering SCRAM included replacement costs for lost bracelets (\$1,200 to 1,500 each), which were leased for the current study, and monitoring fees for each active bracelet (\$6/day per device) in the AMS system. With phones, the costs included the price of refurbished Android smartphones (about \$100 to 120 each) and data plans (\$15/month per device). All SCRAM bracelets and smartphones returned in good working condition were reset to be used by new

participants throughout the study. The cost of replacing 2 bracelets was equivalent to replacing all phones lost in the current study.

Random EMA and daily diary prompts were scheduled 4,776 and 1,347 times, respectively, for the participants in this sample, of which “actual” notifications were sent 3,943 (82.6%) and 1,078 (80%) times, respectively, for random and daily diary EMA assessments. In response to the “actual” notifications received, 3,303 (83.8%) random and 842 (78.1%) daily diary responses were provided. Based on the number of “scheduled” EMA prompts, the overall EMA response rate of this study would be 67.7%. Based on the “actual” prompts participants received, the EMA response rate was 82.6%. In addition, participants self-initiated 395 EMAs before or after drinking. In total, there were 1,001 days of EMA alcohol data before being matched by day to 2 other alcohol measures.

The gap between “scheduled” vs. “actual” EMA prompts could be attributed to app malfunctions, phone or battery replacements, and phones being turned off for more than a couple of days for various reasons (e.g., phones lost or sold). When participants complained about not receiving EMA prompts, we reinstalled the app or exchanged phones or batteries or both. Most of the issues were addressed by a project staff member who was available 3 times a week at the site of participant recruitment.

### Alcohol Use

Table 2 describes the interrater agreement of 3 methods regarding whether alcohol use occurred on a given day. EMA and SCRAM data agreed on 73% of days. Cohen’s interrater agreement coefficient, Kappa of 0.46, indicated that there was 46% greater agreement than what was expected after accounting for main effects or marginal differences. A Cohen’s Kappa between 0.4 and 0.6 is considered a moderate level of agreement (Landish and Koch, 1977). An odds ratio (*OR*) of 8.58 suggested that it was 8.58 times more likely that EMA and SCRAM would agree that there was a drinking day rather than disagree. Between EMA and TLFB and between SCRAM and TLFB, there was relatively weaker agreement, although all Kappa estimates and *ORs* were statistically significant at  $p < 0.05$ .

The nature of disagreement was that 38.9% of the EMA-reported drinking days were contradicted as nondrinking days by the SCRAM data, while 19.8% of the SCRAM-detected drinking days were self-reported as nondrinking days via EMA. These discrepancies were higher for EMA and TLFB, with 53.7% of the EMA-reported drinking days being nondrinking days according to TLFB self-reports and 33.0% of the TLFB-reported drinking days as nondrinking days as per EMA self-reports. Similarly, discrepancies between SCRAM and TLFB were fairly high. Fifty-eight percent of the SCRAM-detected drinking days were nondrinking days based on TLFB, whereas 52.0% of the

**Table 2.** Agreement of Day-level Drinking by the EMA, SCRAM, and TLFB Methods

	SCRAM		Total
	Yes	No	
EMA			
Yes	300	191	491
No	74	404	478
Raw agreement	0.73		
Kappa (SE)	0.46	0.03	$t = 14.58$
OR (95% CI)	8.58	6.31, 11.66	

	TLFB		Total
	Yes	No	
EMA			
Yes	223	259	482
No	110	367	477
Raw agreement	0.62		
Kappa (SE)	0.23	0.03	$t = 7.55$
OR (95% CI)	2.87	2.18, 3.79	

	TLFB		Total
	Yes	No	
SCRAM			
Yes	216	298	514
No	234	611	845
Raw agreement	0.61		
Kappa (SE)	0.15	0.03	$t = 5.44$
OR (95% CI)	1.89	1.50, 2.39	

All coefficients are statistically significant at  $p < 0.05$ . OR = odds ratio. CI = confidence interval.

drinking days based on TLFB self-reports were nondrinking days based on the SCRAM data.

Table 3 shows day-level (Level 1) and person-level (Level 2) estimates for the EMA, SCRAM, and TLFB measures of alcohol use and their correlations from a multilevel analysis ( $N = 1,389$  days for 49 participants). Whether drinking occurred could be determined for 69.8%, 99.6%, and 98.2% of the days, respectively, with EMA, SCRAM, and TLFB measures. Alcohol use was detected on 49%, 38%, and 33% of days by EMA, SCRAM, and TLFB, respectively (see level-2 means in the bottom row block). The intraclass correlations (ICCs) for EMA, SCRAM, and TLFB were 0.30, 0.24, and 0.24, respectively, which means that 30% of the estimated variability in EMA data was due to between-person variability, with the remaining variability at the day level. For SCRAM and TLFB data, 24% of the variability was due to between-person differences, and the remaining variability was due to day-to-day variability within persons.

At the day level (Level 1), TLFB's associations with EMA and SCRAM alcohol use measures were low,  $r = 0.11$  ( $SE = 0.04$ ,  $p < 0.05$ ) or trivial and statistically not significant,  $r = 0.05$  ( $SE = 0.03$ ,  $p = 0.10$ ). However, at the person level (Level 2), the associations among all 3 measures improved, ranging from 0.48 to 0.63. All person-level (Level 2) correlations were statistically significant ( $p < 0.05$ ).

Overall, EMA and SCRAM were correlated 0.40 and 0.63, respectively, at Level 1 and Level 2.

*Alcohol Use Intensity*

Table 4 shows the results from a multilevel analysis of alcohol use intensity measures ( $N = 1,389$  days for 49 participants). Alcohol use intensity could be determined for 59.1%, 98.2%, and 99.6% of the days, respectively, with EMA, SCRAM, and TLFB measures. ICCs for EMA, SCRAM, and TLFB were 0.31, 0.31, and 0.32, respectively, suggesting that 31% to 32% of the variability in alcohol intensity measures was due to between-person differences, and the remaining variability was due to day-to-day variability. Means and variances of alcohol use intensity measures are shown (see Level 2 data in the bottom row block). On average, participants self-reported having 1.83 drinks and 1.61 drinks, respectively, per day, via TLFB and EMA during the observed period.

Of all 3 measures, EMA and SCRAM were correlated modestly ( $r = 0.46$ ,  $SE = 0.06$ ,  $p < 0.05$ ) at Level 1. TLFB, in contrast, had statistically insignificant, weak correlations with EMA ( $r = 0.10$ ,  $SE = 0.07$ ,  $p = 0.13$ ) and SCRAM ( $r = 0.02$ ,  $SE = 0.04$ ,  $p = 0.54$ ). At Level 2, the associations among all 3 measures improved, ranging from 0.30 to 0.78 ( $p < 0.05$ ). Overall, EMA and SCRAM were correlated 0.46 and 0.78, respectively, at Level 1 and Level 2 (see also Fig. 2).

*Demographic Covariates of Alcohol Consumption*

We regressed all 3 alcohol intensity measures on age, sex (female = 1, male = 0), and race (Black = 1, other = 0) simultaneously in a complex survey structural equation model (see Fig. 3). Women tended to report less drinking on the TLFB (coeff. =  $-0.15$ ,  $SE = 0.04$ ,  $p < 0.05$ ); but not on EMA (coeff. =  $-0.07$ ,  $SE = 0.10$ ,  $p = 0.47$ ), holding all other covariates and associations constant. This sex difference was not evident in the SCRAM data (coeff. =  $-0.01$ ,  $SE = 0.08$ ,  $p = 0.95$ ). Age or being Black was not significantly associated with the intensity of alcohol consumption across all 3 methods.

*Sensitivity Analysis: Different 24-hour SCRAM Periods*

We examined 2 other 24-hour periods for TAC data: a midnight-to-midnight period and a 6-am-to-6-am period. Based on the midnight-to-midnight SCRAM data, drinking was observed in 42% of the days, compared to 38% of the days (for both the 6-am-to-6-am and noon-to-noon periods). This difference translates to approximately a day out of 28 days, suggesting that positive TAC readings near midnight may be counted twice if a calendar day is used. Overall, the results reported in Tables 3 and 4 and Figs 2 and 3 were, by and large, unchanged when we used different 24-hour

**Table 3.** Alcohol Use (1 = Drinking, 0 = No Drinking): Robust Maximum-likelihood Estimates for Level 1 (Day) and Level 2 (Person) Data across 3 Methods

Level 1: Day-to-Day, Within Persons ( $N = 1,389$ Days)									
Corr	EMA	SCRAM	TLFB	$N$	Mean	Var			
EMA	1.00			969	0	0.17*			
	–					(0.01)			
SCRAM	0.40*	1.00		1,384	0	0.18*			
	(0.04)	–				(0.01)			
TLFB	0.11*	0.05	1.00	1,364	0	0.17*			
	(0.04)	(0.03)	–			(0.01)			
Level 2: Between Persons ( $N = 49$ )									
Corr	EMA	SCRAM	TLFB	$N$	Mean	Var	Min	Max	ICC
EMA	1.00			49	0.49*	0.08*	0.00	1.00	0.30
	–				(0.04)	(0.01)			
SCRAM	0.63*	1.00		49	0.38*	0.06*	0.00	1.00	0.24
	(0.11)	–			(0.04)	(0.01)			
TLFB	0.61*	0.48*	1.00	49	0.33*	0.05*	0.00	1.00	0.24
	(0.12)	(0.12)	–		(0.04)	(0.01)			

\* $p < 0.05$ . Corr = correlation, Var = variance, ICC = intraclass correlation. Means at Level 2 can be understood as proportions of drinking days. Numbers in parenthesis indicate standard errors. SCRAM data were based on a noon-to-noon 24-hour period.

**Table 4.** Alcohol Use Intensity: Robust Maximum-likelihood Estimates for Level 1 (Day) and Level 2 (Person) Data across 3 Methods

Level 1: Day-to-Day, Within Persons ( $N = 1,389$ days)									
Corr	EMA	SCRAM	TLFB	$N$	Mean	Var			
EMA	1.00			821	0	4.34*			
	–					(0.82)			
SCRAM	0.46*	1.00		1,384	0	0.01*			
	(0.06)	–				(0.00)			
TLFB	0.10	0.02	1.00	1,364	0	9.69*			
	(0.07)	(0.04)	–			(3.04)			
Level 2: Between Persons ( $N = 49$ )									
Corr	EMA	SCRAM	TLFB	$N$	Mean	Var	Min	Max	ICC
EMA	1.00			49	1.61*	1.90*	0.00	11.00	0.31
	–				(0.21)	(0.70)			
SCRAM	0.78*	1.00		49	0.04*	0.00*	0.00	0.44	0.31
	(0.10)	–			(0.01)	(0.00)			
TLFB	0.54*	0.30*	1.00	49	1.83*	4.64*	0.00	24.00	0.32
	(0.11)	(0.09)	–		(0.32)	(2.28)			

\* $p < 0.05$ . Corr = correlation, Var = variance, ICC = intraclass correlation. Numbers in parenthesis indicate standard errors. SCRAM data were based on a noon-to-noon 24-hour period.

periods for SCRAM in this study (i.e., small differences in the second decimal points).

## DISCUSSION

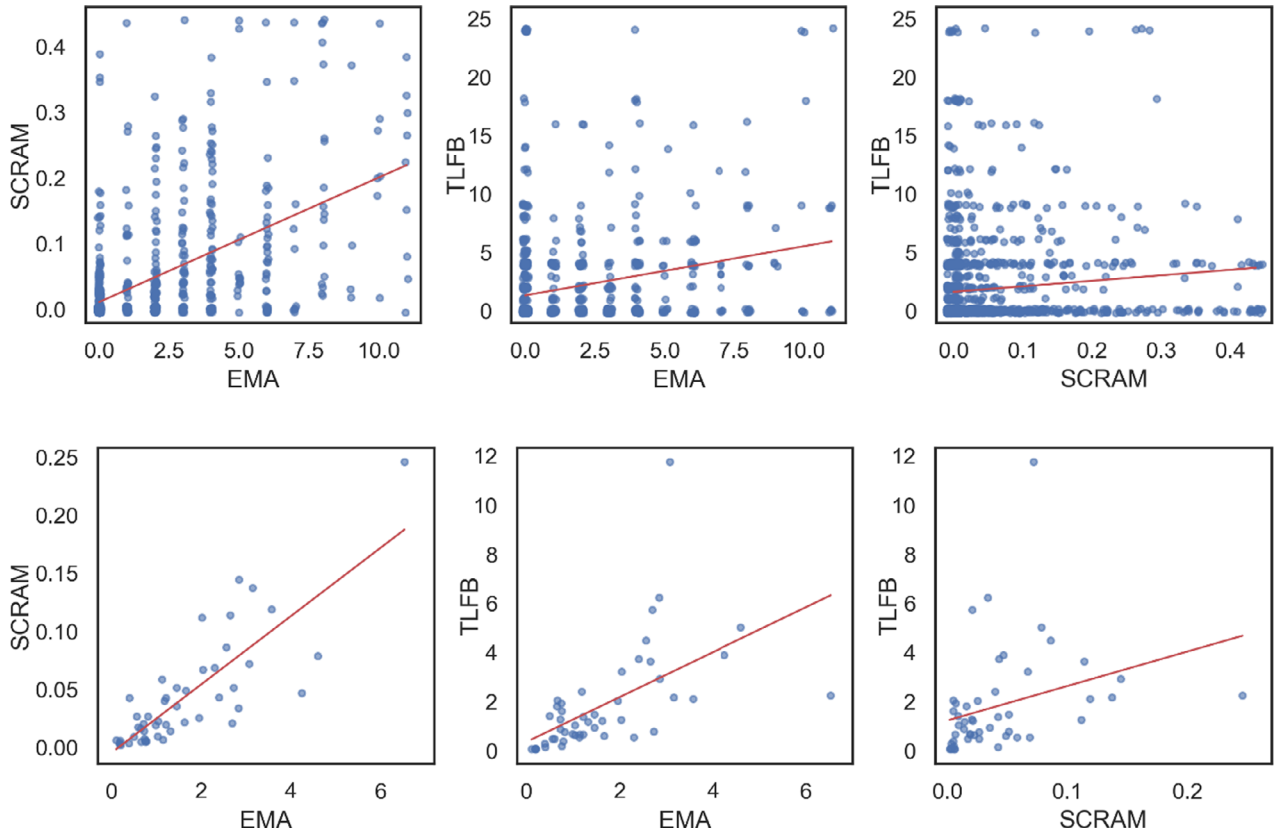
### *Day-level and Person-level Concordance*

We found moderate concordance between EMA and SCRAM over 4 weeks among adults with alcohol misuse who were experiencing homelessness. More specifically, the concordance between EMA and SCRAM was modest at the day level but strong at the person level. TLFB self-reports, in

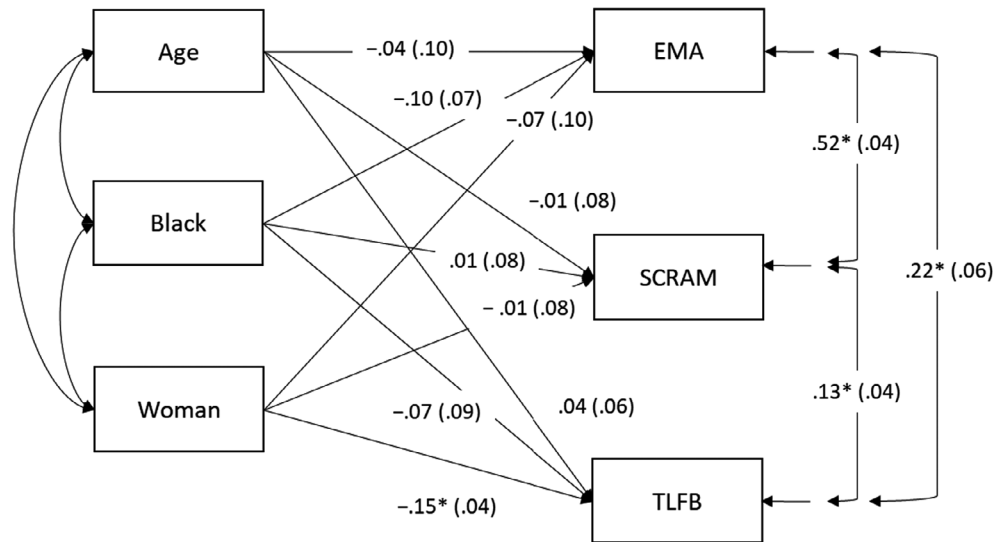
contrast, showed weak and sometimes statistically insignificant concordance with EMA and SCRAM, especially at the day level. The concordance of the 3 measures at the person level was better than at the day level.

Similarly, for alcohol use intensity, day-level correspondence between EMA and SCRAM ( $r = 0.46$ ) was modest, compared with person-level correspondence ( $r = 0.78$ ). The person-level correlation of 0.78 in alcohol use intensity between EMA and SCRAM slightly exceeds the reported correlation of 0.62 between TAC readings and self-reports in a recent systematic review (van Egmond et al., 2020). In an alcohol administration study in a highly controlled





**Fig. 2.** A multilevel analysis of correspondence among 3 alcohol intensity measures: Day-to-day within-person data (Level 1; Top) and between-person data (Level 2; Bottom). The magnitude of estimated correlations at 2 levels are shown in Table 4.



**Fig. 3.** Demographic covariates of alcohol intensity as measured by the EMA, SCRAM, and TLFB methods ( $N = 49$ ). A complex survey option with robust maximum-likelihood estimation was used. Reported path coefficients are standardized estimates. Numbers in parenthesis indicate standard errors. \* $p < 0.05$ .

laboratory setting, the correlation between transdermal and breath assessments of alcohol across 24 subjects was 0.84 (Sakai et al., 2006). Therefore, the findings from the current study conducted in a field setting are promising, particularly

given that the measures of intensity were not analogous between SCRAM and self-reports via EMA or TLFB. However, it is possible that EMA self-reports corresponded relatively well with the TAC data because participants were



aware of being continuously monitored by the SCRAM. In addition, due to the study inclusion criteria, which included a willingness to wear the SCRAM for 4 weeks, participants in the current study might have been somewhat more conscientious than other high-risk populations.

Note that the average number of drinks per day based on TLFB self-reports was 1.83, whereas the average number of drinks per day based on EMA reports was 1.61 in this current study, which represents a 6.2 drink difference spread over 28 days and a 1.4 drink difference per drinking day between 2 self-report methods. Participants reported, on average, 3.84 drinks and 5.23 drinks per drinking day, respectively, via EMA and TLFB. Overall, TLFB had low and statistically insignificant correlations with the EMA and SCRAM measures of alcohol use intensity at the day level.

Findings from this study suggest that for studies focusing on daily alcohol abstinence as an outcome measure, TLFB may not be appropriate because it has a very weak convergence with other measurement approaches, perhaps because TLFB relies on fading memory, which is not surprising. However, for studies focusing on the proportion of alcohol abstinence over 30 days (which means aggregated data across days within participants), TLFB self-reports may still provide an adequate outcome measure for clinical alcohol treatment trials. The Food and Drug Administration (FDA) tentatively recommended that both alcohol abstinence and moderation in alcohol use (i.e., no heavy drinking days) are acceptable efficacy endpoints in pharmaceutical trials for alcohol use disorder in 2015 (Food and Drug Administration, 2015). Based on this recommendation, if one's goal is to reduce the number of drinking days in the past month, TLFB self-reports may provide reasonable outcome data at a low cost.

In contrast, for trials to provide real-time monitoring and immediate reinforcement or other interactions with participants, EMA self-reports or the SCRAM may be more appropriate than TLFB self-reports. However, there is a need to improve EMA assessment and analysis in future studies. We could determine the drinking days and number of standard drinks for 69.8% and 59.1% of the days, using EMA self-reports in the current study, due to occasional nonresponse. Using SCRAM and TLFB methods, we could obtain comparable information for 99.6% and 98.2% of the days, respectively. From a clinical trial standpoint, the findings suggest that each alcohol assessment method, in its current form, has comparative strengths and weaknesses. When used with patient-reported outcomes such as EMA and TLFB, available alcohol biosensors may improve clinical trials' internal and statistical conclusion validity.

### *Special Considerations for Using SCRAM*

We reported TAC data based on the noon-to-noon 24-hour period, as this selection is typical for studies of day-level drinking in natural settings (Roache et al., 2019). When we examined 2 other 24-hour periods in this current sample, 3

different 24-hour periods did not yield meaningful differences. This may be partly because this was an adult homeless sample who spent most of their nights at the shelter where drinking was prohibited. However, in studies of populations under different situations, it may be necessary to carefully examine when to start a 24-hour period because the transdermal alcohol curve is delayed relative to blood or breath alcohol levels (Barnett et al., 2017). It has been discussed that for drinking episodes that span across 2 calendar days (midnight to midnight), TAC levels in the morning can reflect drinking that occurred the day before, which presumably would be self-reported as a drinking episode the day before on EMA or TLFB (Barnett et al., 2017).

Furthermore, laboratory-based research has shown that the initial, times-to-peak TAC tend to lag behind the times-to-peak curve of blood or breath concentration considerably, with its time lag increasing as a function of the amount of alcohol consumed, from 83 minutes for 1 drink to 162 minutes for 5 drinks, and it takes longer to be eliminated (Karns-Wright et al., 2016). Therefore, hour-by-hour correspondence among measures would be challenging to report without a more careful, case-by-case examination of TAC data. The fact that TAC levels are difficult to adjust in real-time or in large batches remains a barrier to overcome in future studies.

TAC data from the SCRAM are obtained through AMS that uses a propriety software program. The AMS detection of a drinking event tends to be conservative, with most low-level drinking (1-3 standard drinks) going undetected (see Roache et al., 2015; Roache et al., 2019). Although we used less stringent criteria (Barnett et al., 2014) in the current study, these relaxed criteria may still be too conservative for detecting a drinking episode for use in clinical trials. The poorer ability of the SCRAM to detect low-level drinking may help explain the difference in the rate of alcohol use between EMA (49%) and SCRAM (38%). Given the importance of objective measures of drinking in clinical trials, further refinement in detecting and calibrating alcohol use and alcohol use intensity and easier processing of TAC data would be a high-value target in future research.

### *Compliance and Feasibility*

We found the SCRAM was generally well tolerated over 4 weeks. Most of the sample (91.8%) wore the SCRAM for 28 days or more. For this sample that completed all 4 weeks, we did not experience any SCRAM failures. This may be because we incorporated a mid-point visit to ensure that the phone and the SCRAM were working properly, check the SCRAM for proper fit, and upload data from the SCRAM. We also screened participants with health issues that could contraindicate wearing the SCRAM and provided a sheet with suggestions about how to decrease any potential physical and mental discomfort from wearing the SCRAM. Finally, in order to reduce potential stigma around wearing the SCRAM, we placed a prominent sticker on the SCRAM

and provided people with a laminated ID card certifying their participation in a research study.

We also found relatively good compliance with EMA survey prompts. Participants were compensated during the mid-point visit for their EMA responses completed thus far, which might have helped with adherence to the EMA protocol. In the current study, participants responded to EMA prompts 82.6% of the time, which exceeds the pooled compliance rate of 75% reported in the meta-analysis of substance use studies (Jones et al., 2019). Our rate is on par with the estimates ranging from 84% to 93% in samples of young adults and outpatients or 63% to 90% in college student samples (see Piasecki, 2019 for a review). Therefore, the current study's EMA protocol, a "medium resolution" EMA protocol (Piasecki, 2019), may be a reasonable balance between participant burden and data quality.

Despite this promising data, it is important to note that there was considerable equipment loss during the study (Figure 1). A total of 23 study phones and 8 SCRAMs were lost or irreparably damaged during the study period. Our initial protocol was to replace the phone 1 time if a participant reported it lost or stolen; we later revised the protocol so that we would replace the phone 1 time only if a participant had completed at least 50% EMA prompts for at least 1 week prior to the phone loss. Because of the high cost, we did not replace any SCRAMs. As depicted in the flowchart (Fig. 1), 19 participants did not complete the 4-week follow-up following equipment loss, and 10 additional participants did not attend the follow-up visit. Because of the residential and financial circumstances faced by people experiencing homelessness, this level of attrition was expected to a certain extent.

Furthermore, some participants complained about the size and physical discomfort of wearing the SCRAM, which has been discussed in previous studies (Alessi et al., 2017; Caluzzi et al., 2019). When the 29 participants who did not complete the study were compared with those who did complete the study, 2 statistically significant differences were found (Table 1). These differences may be chance findings due to multiple statistical null hypothesis tests. However, data in Table 1 cautiously suggest that people who completed the 4-week follow-up tended to function better than those who dropped out. Therefore, although the reported findings do not appear to be biased by those who experienced equipment loss or had early exits, it may be desirable to accommodate potential dropout mechanisms when developing a data model for inference on outcomes.

#### *Covariates of Alcohol Use Measures*

We examined whether alcohol use intensity differed across demographic groups and assessment measures. There have been reports that the SCRAM may be less able to detect drinking in women (e.g., Marques and McKnight, 2009, approximately at the same 0.08 g/dl blood alcohol concentration), although this finding has not been replicated with

newer SCRAM device versions. Indeed, Barnett and colleagues (2014) reported that at the level of 4 drinks or less, women's drinking episodes were more likely to be detected and peak TAC was higher (with older versions of the SCRAM: SCRAMII and SCRAMx), compared with men. Though we did not make this comparison using the same drinking categories as Barnett et al., we did not find any evidence of higher levels of peak TAC among women. Because we did not control alcohol dose and blood alcohol concentration levels, our observation should be cautiously interpreted. The relationship between peak TAC and breath alcohol concentration levels may differ for men and women (Hill-Kapturczak et al., 2015; Karns-Wright et al., 2016), which needs to be further studied.

We discovered that women reported relatively lower levels of drinking on the TLFB, whereas there was no sex difference in the SCRAM or EMA alcohol use intensity. This finding may be related to "social desirability," where participants tend to describe their behavior in more positive terms. Given that the TLFB was the only method that required face-to-face interactions with a male research staff member, participant bias might have played a role. In addition, consistent with prior studies (e.g., Kaplan and Koffarnus, 2019), participants overall reported the lowest levels of drinking based on the TLFB method (33% vs. 49% and 38% for EMA and TAC, respectively). Finally, all participants in the current study were receiving health services at the homeless shelter, including alcohol and substance use treatment services, which might have contributed to reductions in the percentage of drinking days and number of drinks during the study, compared to their baseline data.

#### *Limitations and Future Directions*

The current study has a number of limitations. First, because of our focus on comparing 3 alcohol assessment approaches with a common sampling interval, we aggregated within-day EMA self-reports and the SCRAM TAC readings. A more fine-grained analysis focusing on truly "event-level" drinking data may shed light on the relative advantages of EMA self-report vs. TAC data. On a related point, for interrater agreement analysis, we ignored data clustering within persons in part because sparse cells would result if participants are used as strata and also because some individuals did not have any variability in data (e.g., no drinking). With more data at each data level, interrater agreement may be better investigated in future studies.

Second, in the EMA protocol, the number of drinks (i.e., intensity in the current study) was asked via a dropdown response option, which visibly displayed 0 through 10. Participants had to touch the phone screen and scroll to the final value (i.e., 11 or more). Therefore, we coded 11 drinks or more as 11, which created truncated data. Without this data truncation, the full range of drinks would have been slightly larger. There were 9 observations with a score of 11 (1.1% of valid observations) in the current study.

Third, we limited the sample to people who completed the TLFB at the 4-week follow-up visit; attrition of those who started the study was 37%. This might have resulted in selection bias, yielding a sample that was overrepresented by conscientious individuals. Moreover, those who completed the 4-week follow-up tended to function better than those who dropped out.

Fourth, this was a single-site study with a predominantly male sample (82%). A larger or more balanced sample may help to explore the role of sex in SCRAM-detected drinking and possible biased reporting on the TLFB. On a related point, we did not assess participants' tendency to provide a socially acceptable response. Given that we had an objective measure of drinking via SCRAM, any response bias might have been curtailed. However, it is challenging to estimate the extent to which the participant's awareness of being monitored influenced their response. A social desirability measure might have explained the sex difference in TLFB reports.

Finally, participants' EMA response rates could be improved by having research staff onsite for more days and by allocating more time during equipment visit. Unfamiliarity with the smartphone app, especially initially, may have contributed to some missing EMA responses. Upon completing the study, most (78%) of the current study participants indicated that a smartphone app could help change their behavior, a positive change compared with their baseline attitude (59%).

Accurately assessing alcohol use via objective biosensors remains an important goal, especially for randomized clinical trials for alcohol misuse. Although self-reported outcomes will likely continue to be an important part of patient-oriented clinical outcomes research, it can be problematic when there are no objective outcome measures to interpret these self-reports (see Mun et al., 2015 for an example). Existing biosensors, although highly promising, have room to improve in terms of cost, functionality, and durability for use in clinical trials. Therefore, the current finding that EMA self-reports of alcohol use among adults experiencing homelessness showed good concordance with TAC levels from the SCRAM, especially at the person level, is encouraging not only because EMA data are accurate and valid but also because EMA can be used to gather ancillary data, such as a person's mood, cognition, or location that might help to detect factors that place them at risk for drinking. With the identification of "in the moment" risk factors, tailored intervention messages can be provided based on the person's unique risk factors and drinking goal for each day (Businelle et al., 2016; Walters et al., 2021).

## CONCLUSIONS

EMA self-report of alcohol use among adults experiencing homelessness is feasible, accurate, and promising in connection with "just-in-time" intervention development and delivery. EMA is straightforward to administer, provides participant privacy, which could help address stigma, and

can potentially be enriched with other passive data for better detection and treatment of alcohol use, an important breakthrough for underserved, highly marginalized populations.

## ACKNOWLEDGMENTS

We would like to thank Charles Williams and the staff at the Bridge for their help with participant recruitment and Recovery Monitoring Solutions for their help with the SCRAM setup and execution. Xiaoyin Li is now affiliated with the Tobacco Research and Evaluation, The University of Texas at Austin.

## CONFLICT OF INTEREST

MSB is an inventor of the Insight mHealth Platform and receives royalties related to use of this platform. However, MSB did not receive royalties for the use of the platform for this study.

## FUNDING INFORMATION

This study was supported by grants R34 AA024584 and 3R34 AA024584-02S1 awarded to MSB and STW and, in part, by R01 AA019511 awarded to EYM. In addition, this work was partially supported by the Oklahoma Tobacco Settlement Endowment Trust grant 092-016-0002 and NCI grant P30 CA225520 awarded to the Stephenson Cancer Center and used the mHealth Shared Resource.

## REFERENCES

- Adams-Guppy JR, Guppy A (2016) A systematic review of interventions for homeless alcohol-abusing adults. *Drugs* 23:15–30.
- Alessi SM, Barnett NP, Petry NM (2017) Experiences with SCRAMx alcohol monitoring technology in 100 alcohol treatment outpatients. *Drug Alcohol Depend* 178:417–424.
- Alessi SM, Barnett NP, Petry NM (2019) Objective continuous monitoring of alcohol consumption for three months among alcohol use disorder treatment outpatients. *Alcohol* 81:131–138.
- Baggett TP, Chang Y, Singer DE, Porneala BC, Gaeta JM, O'Connell JJ, Rigotti NA (2015) Tobacco-, alcohol-, and drug-attributable deaths and their contribution to mortality disparities in a cohort of homeless adults in Boston. *Am J Public Health* 105(6):1189–1197.
- Barnett NP, Celio MA, Tidey JW, Murphy JG, Colby SM, Swift RM (2017) A preliminary randomized controlled trial of contingency management for alcohol use reduction using a transdermal alcohol sensor. *Addiction* 112:1025–1035.
- Barnett NP, Meade EB, Glynn TR (2014) Predictors of detection of alcohol use episodes using a transdermal alcohol sensor. *Exp Clin Psychopharmacol* 22:86–96.
- Barnett NP, Souza T, Rosen G, Luczak S, Glynn T, Swift RM (2015) Transdermal Alcohol Sensor Data Macro (Version 1.5) [Computer Program]. Brown University, Providence, RI.
- Barnett NP, Tidey J, Murphy JG, Swift R, Colby SM (2011) Contingency management for alcohol use reduction: a pilot study using a transdermal alcohol sensor. *Drug Alcohol Depend* 118(2–3):391–399.
- Businelle MS, Ma P, Kendzor DE, Frank SG, Vidrine DJ, Wetter DW (2016) An ecological momentary intervention for smoking cessation: Evaluation of feasibility and effectiveness. *J Med Internet Res* 18:e321.

- Businelle MS, Walters ST, Mun E-Y, Kirchner T, Hébert ET, Li X (2020) Reducing drinking among people experiencing homelessness: protocol for the development and testing of a just-in-time adaptive intervention. *JMIR Res Protoc* 9:e15610.
- Caluzzi G, Pennay A, Cook M, Wright C, Norman T, Kuntsche E (2019) Transdermal monitors to assess alcohol consumption in real-time and real-life—a qualitative study on user-experience. *Addict Res Theory* 27:354–361.
- Carvalho AF, Heilig M, Perez A, Probst C, Rehm J (2019) Alcohol use disorders. *Lancet* 394:781–792.
- Collins SE, Clifasefi SL, Nelson LA, Stanton J, Goldstein SC, Taylor EM, Hoffmann G, King VL, Hatsukami AS, Cunningham ZL, Taylor E, Mayberry N, Malone DK, Jackson TR (2019) Randomized controlled trial of harm reduction treatment for alcohol (HaRT-A) for people experiencing homelessness and alcohol use disorder. *Int J Drug Policy* 67:24–33.
- Dougherty DM, Charles NE, Acheson A, John S, Furr RM, Hill-Kapturczak N (2012) Comparing the detection of transdermal and breath alcohol concentrations during periods of alcohol consumption ranging from moderate drinking to binge drinking. *Exp Clin Psychopharmacol* 20:373–381.
- Dougherty DM, Hill-Kapturczak N, Liang Y, Karns TE, Cates SE, Lake SL, Mullen J, Roache JD (2014) Use of continuous transdermal alcohol monitoring during a contingency management procedure to reduce excessive alcohol use. *Drug Alcohol Depend* 142:301–306.
- Dougherty DM, Karns TE, Mullen J, Liang Y, Lake SL, Roache JD, Hill-Kapturczak N (2015) Transdermal alcohol concentration data collected during a contingency management program to reduce at-risk drinking. *Drug Alcohol Depend* 148:77–84.
- 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(3):195–200.
- Fairbairn CE, Kang D (2019) Temporal dynamics of transdermal alcohol concentration measured via new-generation wrist-worn biosensor. *Alcohol Clin Exp Res* 43:2060–2069.
- Fairbairn CE, Rosen IG, Luczak SE, Venerable WJ (2019) Estimating the quantity and time course of alcohol consumption from transdermal alcohol sensor data: A combined laboratory-ambulatory study. *Alcohol* 81:111–116.
- Fazel S, Khosla V, Doll H, Geddes J (2008) The prevalence of mental disorders among the homeless in western countries: systematic review and meta-regression analysis. *PLoS Medicine* 5:e225.
- Food and Drug Administration (2015) Alcoholism: Developing Drugs for Treatment Guidance for Industry. US Department of Health and Human Services, Rockville, MD.
- Grant S, Pedersen ER, Osilla KC, Kulesza M, D'Amico EJ (2015) Reviewing and interpreting the effects of brief alcohol interventions: comment on a Cochrane review about motivational interviewing for young adults. *Addiction* 111:1521–1527.
- Guyatt GH, Oxman AD, Vist G, Kunz R, Brozek J, Alonso-Coello P, Montori V, Akl EA, Djulbegovic B, Falck-Ytter Y, Norris SL, Williams JW Jr, Atkins D, Meerpohl J, Schünemann HJ (2011) GRADE guidelines: 4. Rating the quality of evidence—study limitations (risk of bias). *J Clin Epidemiol* 64:407–415.
- Hämäläinen MD, Zetterström A, Winkvist M, Söderquist M, Öhagen P, Andersson K, Nyberg F (2020) Breathalyser-based eHealth data suggest that self-reporting of abstinence is a poor outcome measure for alcohol use disorder clinical trials. *Alcohol Alcohol* 55:237–245.
- Higgins JPT, Green S (2011) *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0. The Cochrane Collaboration, Available from [www.cochrane-handbook.org](http://www.cochrane-handbook.org)
- Hill-Kapturczak N, Lake SL, Roache JD, Cates SE, Liang Y, Dougherty DM (2014) Do variable rates of alcohol drinking alter the ability to use transdermal alcohol monitors to estimate peak breath alcohol and total number of drinks? *Alcohol Clin Exp Res* 38:2517–2522.
- Hill-Kapturczak N, Roache JD, Liang Y, Karns TE, Cates SE, Dougherty DM (2015) Accounting for sex-related differences in the estimation of breath alcohol concentrations using transdermal alcohol monitoring. *Psychopharmacology* 232:115–123.
- Hoepfner BB, Stout RL, Jackson KM, Barnett NP (2010) How good is fine-grained Timeline Follow-back data? Comparing 30-day TLFBI and repeated 7-day TLFBI alcohol consumption reports on the person and daily level. *Addict Behav* 35:1138–1143.
- Hurstak E, Johnson JK, Tieu L, Guzman D, Ponath C, Lee CT, Jamora CW, Kushel M (2017) Factors associated with cognitive impairment in a cohort of older homeless adults: Results from the HOPE HOME study. *Drug Alcohol Depend* 178:562–570.
- Jones A, Remmerswaal D, Verveer I, Robinson E, Franken IHA, Wen CKF et al (2019) Compliance with ecological momentary assessment protocols in substance users: a meta-analysis. *Addiction* 114(4):609–619.
- Kaplan BA, Koffarnus MN (2019) Timeline followback self-reports underestimate alcohol use prior to successful contingency management treatment. *Alcohol Alcohol* 54(3):258–263.
- Karns-Wright TE, Roache JD, Hill-Kapturczak N, Liang Y, Mullen J, Dougherty DM (2016) Time delays in transdermal alcohol concentrations relative to breath alcohol concentrations. *Alcohol Alcohol* 52:35–41.
- Kulesza M, Ramsey S, Brown R, Larimer M (2014) Stigma among individuals with substance use disorders: Does it predict substance use, and does it diminish with treatment? *J Addict Behav Ther Rehabil* 3:1000115.
- Landish JR, Koch GG (1977) The measurement of observer agreement for categorical data. *Biometrics* 33:159–174.
- Leffingwell TR, Cooney NJ, Murphy JG, Luczak S, Rosen G, Dougherty DM, Barnett NP (2013) Continuous objective monitoring of alcohol use: twenty-first century measurement using transdermal sensors. *Alcohol Clin Exp Res* 37:16–22.
- Marques PR, McKnight AS (2009) Field and laboratory alcohol detection with 2 types of transdermal devices. *Alcohol Clin Exp Res* 33(4):703–711.
- McKenna H, Treanor C, O'Reilly D, Donnelly M (2018) Evaluation of the psychometric properties of self-reported measures of alcohol consumption: a COSMIN systematic review. *Subst Abuse Treat Prev Policy* 13:6.
- Mun E-Y, Atkins DC, Walters ST (2015) Is motivational interviewing effective at reducing alcohol misuse in young adults? A critical review of Foxcroft et al. (2014). *Psychol Addict Behav* 29:836–846.
- Muthén LK, Muthén BO (1998–2020) *Mplus User's Guide*, 8th edn. Muthén & Muthén, Los Angeles, CA.
- Neisler J, Shree S, Reitzel LR, Chen TA, Kendzor DE, Obasi EM, Wrighting Q, Businelle MS (2019) Characterizing alcohol use behaviors among homeless men and women. *Am J Health Behav* 43:37–49.
- Orwin RG, Garrison-Mogren R, Jacobs ML, Sonnefeld LJ (1999) Retention of homeless clients in substance abuse treatment. Findings from the National Institute on Alcohol Abuse and Alcoholism Cooperative Agreement Program. *J Subst Abuse Treat* 17:45–66.
- Piasecki TM (2019) Assessment of alcohol use in the natural environment. *Alcohol Clin Exp Res* 43:564–577.
- Rash CJ, Petry NM, Alessi SM, Barnett NP (2019) Monitoring alcohol use in heavy drinking soup kitchen attendees. *Alcohol* 81:139–147.
- Reitzel LR, Chinamuthevi S, Daundasekara SS, Hernandez DC, Chen TA, Harkara Y, Obasi EM, Kendzor DE, Businelle MS (2020) Association of problematic alcohol use and food insecurity among homeless men and women. *Int J Environ Res Public Health* 17:3631.
- Roache JD, Karns TE, Hill-Kapturczak N, Mullen J, Liang Y, Lamb RJ, Dougherty DM (2015) Using transdermal alcohol monitoring to detect low-level drinking. *Alcohol Clin Exp Res* 39:1120–1127.
- Roache JD, Karns-Wright TE, Goros M, Hill-Kapturczak N, Mathias CW, Dougherty DM (2019) Processing transdermal alcohol concentration (TAC) data to detect low-level drinking. *Alcohol* 81:101–110.
- Sacks JAY, Drake RE, Williams VF, Banks SM, Herrell JM (2003) Utility of the timeline follow-back to assess substance use among homeless adults. *J Nerv Ment Dis* 191:145–153.
- Sakai JT, Mikulich-Gilbertson SK, Long RJ, Crowley TJ (2006) Validity of transdermal alcohol monitoring: fixed and self-regulated dosing. *Alcohol Clin Exp Res* 30:26–33.
- Saunders JB, Aasland OG, Babor TF, de la Fuente JR, Grant M (1993) Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO collaborative project on early detection of persons with harmful alcohol consumption—II. *Addiction* 88:791–804.

- Searles JS, Helzer JE, Rose GL, Badger GJ (2002) Concurrent and retrospective reports of alcohol consumption across 30, 90 and 366 days: interactive voice response compared with the timeline follow back. *J Stud Alcohol* 63:352–362.
- Searles JS, Helzer JE, Walter DE (2000) Comparison of drinking patterns measured by daily reports and timeline follow back. *Psychol Addict Behav* 14(3):277–286.
- Shiffman S (2009) Ecological momentary assessment (EMA) in studies of substance use. *Psychol Assess* 21:486–497.
- Sobell LC, Sobell MB (1995) Alcohol consumption measures, in *Assessing Alcohol Problems: A Guide for Clinicians and Researchers*. Vol. 2 (Allen JP, Columbus M eds), pp. 75–99. National Institute on Alcohol Abuse and Alcoholism, Bethesda, MD.
- Stasiewicz PR, Vincent PC, Bradizza CM, Connors GJ, Maisto SA, Mercer ND (2008) Factors affecting agreement between severely mentally ill alcohol abusers' and collaterals' reports of alcohol and other substance abuse. *Psychol Addict Behav* 22:78–87.
- Stone B, Dowling S, Cameron A (2019) Cognitive impairment and homelessness: A scoping review. *Health Soc Care Community* 27:e125–e142.
- Taylor EM, Kendzor DE, Reitzel LR, Businelle MS (2016) Health risk factors and desire to change among homeless adults. *Am J Health Behav* 40:455–460.
- Tucker JA, Foushee HR, Black BC, Roth DL (2007) Agreement between prospective interactive voice response self-monitoring and structured retrospective reports of drinking and contextual variables during natural resolution attempts. *J Stud Alcohol Drugs* 68:538–542.
- van Egmond K, Cassandra JCW, Livingston M, Kuntsche E (2020) Wearable transdermal alcohol monitors: A systematic review of detection validity, relationship between transdermal and breath alcohol concentration and influencing factors. *Alcohol Clin Exp Res* 44:1918–1932.
- Walters ST, Businelle MS, Suchting R, Li X, Hebert ET, Mun E-Y (2021) Using machine learning to identify predictors of imminent drinking and create tailored messages for at-risk drinkers experiencing homelessness. *J Subst Abuse Treat* (in press).
- Wenzel SL, Audrey Burnam M, Koegel P, Morton SC, Miu A, Jinnett KJ, Sullivan JG (2001) Access to inpatient or residential substance abuse treatment among homeless adults with alcohol or other drug use disorders. *Med Care* 39:1158–1169.
- Witkiewitz K, Finney JW, Harris AHS, Kivlahan DR, Kranzler HR (2015a) Guidelines for the reporting of treatment trials for alcohol use disorders. *Alcohol Clin Exp Res* 39:1571–1581.
- Witkiewitz K, Finney JW, Harris AHS, Kivlahan DR, Kranzler HR (2015b) Recommendations for the design and analysis of treatment trials for alcohol use disorders. *Alcohol Clin Exp Res* 39:1557–1570.
- 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.