



Published in final edited form as:

Alcohol (Hanover). 2023 February ; 47(2): 308–318. doi:10.1111/acer.14995.

Correspondence between alcohol use measured by a wrist-worn alcohol biosensor and self-report via ecological momentary assessment over a 2-week period

Veronica L. Richards^{1,2}, Nancy P. Barnett³, Robert L. Cook¹, Robert F. Leeman^{4,5}, Timothy Souza³, Stuart Case¹, Cindy Prins¹, Christa Cook⁶, Yan Wang¹

¹Department of Epidemiology, College of Public Health and Health Profession and College of Medicine, University of Florida, Gainesville, Florida, USA

²Edna Bennett Pierce Prevention Research Center, The Pennsylvania State University, University Park, Pennsylvania, USA

³Center for Alcohol and Addiction Studies, Brown University School of Public Health, Providence, Rhode Island, USA

⁴Department of Health Education and Behavior, College of Health and Human Performance, University of Florida, Gainesville, Florida, USA

⁵Department of Health Sciences, Bouvé College of Health Sciences, Northeastern University, Boston, Massachusetts, USA

⁶College of Nursing, University of Central Florida, Orlando, Florida, USA

Abstract

Background: Transdermal alcohol biosensors measure alcohol use continuously, passively, and non-invasively. There is little field research on the Skyn biosensor, a new-generation, wrist-worn transdermal alcohol biosensor, and little evaluation of its sensitivity and specificity and the day-level correspondence between transdermal alcohol concentration (TAC) and number of self-reported drinks.

Methods: Participants ($N = 36$; 61% male, $M_{\text{age}} = 34.3$) wore the Skyn biosensor and completed ecological momentary assessment (EMA) surveys about their alcohol use over 2 weeks. A total of 497 days of biosensor and EMA data were collected. Skyn-measured drinking episodes were defined by $\text{TAC} > 5 \mu\text{g/L}$. Skyn data were compared to self-reported drinking to calculate sensitivity and specificity (for drinking day vs. nondrinking day). Generalized estimating equations models were used to evaluate the correspondence between TAC features (peak TAC and TAC-area under the curve (AUC)) and number of drinks. Individual-level factors (sex, age, race/ethnicity, body mass index, human immunodeficiency virus status, and hazardous drinking) were examined to explore associations with TAC controlling for number of drinks.

Correspondence: Veronica L. Richards, Edna Bennett Pierce Prevention Research Center, The Pennsylvania State University, 320 Biobehavioral Health Building, University Park, PA 16802, USA. vlr5157@psu.edu.

CONFLICT OF INTEREST

The authors have no conflicts of interest to report.

Results: Using a minimum TAC threshold of 5 $\mu\text{g/L}$ plus coder review, the biosensor had sensitivity of 54.7% and specificity of 94.6% for distinguishing drinking from non-drinking days. Without coder review, the sensitivity was 78.1% and the specificity was 55.2%. Peak TAC ($\beta = 0.92$, $p < 0.0001$) and TAC-AUC ($\beta = 1.60$, $p < 0.0001$) were significantly associated with number of drinks. Females had significantly higher TAC levels than males for the same number of drinks.

Conclusions: Skyn-derived TAC can be used to measure alcohol use under naturalistic drinking conditions, additional research is needed to accurately identify drinking episodes based on Skyn TAC readings.

Keywords

alcohol; biosensor; ecological momentary assessment; real-time; transdermal alcohol concentration

INTRODUCTION

Transdermal alcohol biosensors allow for the remote, continuous, noninvasive, and objective measurement of alcohol use and can provide more fine-grained data on drinking patterns than self-reports, biomarkers, and breathalyzers (Barnett, 2015; Litten et al., 2010; Norman et al., 2021; Northcote & Livingston, 2011). Transdermal alcohol biosensors measure transdermal alcohol concentration (TAC) by capturing the approximately 1% of ethanol that is eliminated through the skin (Swift, 2003). The advantages offered by transdermal alcohol biosensors allow for innovative approaches to monitoring or intervening on alcohol use. For over a decade the SCRAM-Continuous Alcohol Monitor (SCRAM-CAM; Alcohol Monitoring Systems, Littleton, CO) ankle has been available and used extensively in intervention studies (i.e., contingency management trials; Alessi et al., 2017; Barnett et al., 2017; Dougherty et al., 2014; Dougherty, Karns, et al., 2015; Mathias et al., 2018). However, the SCRAM-CAM is challenging to use in “real-world” scenarios (e.g., with non-treatment seeking participants or in naturalistic alcohol use studies) due to its appearance/size and the stigma associated with it (as it is often used in the criminal justice system; Kilmer et al., 2013; Villalba et al., 2020).

A new-generation wrist-worn transdermal alcohol biosensor, the BACtrack Skyn (BACtrack, Inc.), resembles a fitness tracker and offers a promising alternative to the SCRAM-CAM. To date, small studies using the Skyn demonstrate high user acceptability (Merrill et al., 2022) and lab-based studies in which alcohol administration is used in a controlled setting show strong correlations between TAC and breath alcohol concentration (BrAC; Fairbairn & Bosch, 2021; Fairbairn & Kang, 2019; Merrill et al., 2022; Wang et al., 2021). Thus, although there is evidence of valid measurement of alcohol use in controlled settings with the Skyn, and acceptability of the Skyn in field studies, additional validation is needed in order for the Skyn to be incorporated into future large-scale studies and interventions.

Specifically, the ability of the Skyn to accurately detect alcohol use in the natural environment requires the identification of a criterion or set of criteria that can be applied to Skyn TAC data. Currently, there are no criteria that have been well tested and replicated to identify consumed alcohol and specifically to distinguish between a drinking day and

a non-drinking day. Using only a minimal TAC threshold (i.e., a positive reading that is significantly elevated from baseline) will likely produce false positive drinking episodes, as TAC rises quickly after an environmental alcohol exposure such as hand sanitizer. Thus, it is unlikely that solely relying on the minimal TAC threshold can accurately distinguish drinking from non-drinking days, without additional review by researchers to examine each potential episode.

Once an alcohol episode or drinking day is identified, TAC readings can be used to quantify the amount of alcohol consumed (using area under the curve, AUC, representing the individual's total exposure to alcohol) or level of intoxication (peak TAC, the highest level of TAC reached during the drinking episode). Such information can be used to characterize patterns of alcohol use and measure changes in drinking behavior over time. Amount of alcohol consumed (i.e., self-reported number of drinks) has been associated with both features from SCRAM-CAM data (Barnett et al., 2014; Dougherty et al., 2012; Richards et al., 2021; Russell et al., 2022) but has hardly been studied for the Skyn. Specifically, Ash and colleagues found that the Skyn had a high sensitivity for detecting heavy drinking and Rosenberg and colleagues found strong correlations between TAC and number of drinks in a sample of five (Ash et al., 2022; Rosenberg et al., 2021). By determining the level of correspondence between self-reported number of drinks and TAC, we can gain a better understanding of how Skyn TAC readings and derived features are associated with actual drinking behaviors. Additionally, the detection of alcohol using the SCRAM-CAM does not appear to be influenced by individual-level factors (e.g., sociodemographics, body mass index [BMI], chronic disease), other than sex and number of drinks (Barnett et al., 2014; Dougherty, Hill-Kapturczak, et al., 2015; Richards et al., 2021). It is unclear if these relationships hold for the Skyn biosensor.

In this study, we aimed to (1) identify the number of drinking days detected by a wrist-worn biosensor over 2 weeks defined by a minimum TAC alone (TAC-defined) or defined by a minimum TAC plus additional coder review (coder-defined); (2) compare the sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) of the two definitions of biosensor-detected drinking days compared to self-reported drinking days during a 2-week period; (3) using coder-defined drinking days, determine the level of correspondence between number of self-reported drinks and TAC features derived from biosensor readings (peak TAC and TAC-AUC); and (4) explore whether certain individual-level factors (sex, age, race/ethnicity, BMI, HIV status, and hazardous drinking) are associated with TAC features while controlling for self-reported number of drinks.

MATERIALS AND METHODS

This investigation used field data from the Alcohol Wearables And Research using ecological momentary assessment (AWARE; EMA) study, a pilot study that sought to evaluate the reliability and validity of the Skyn biosensor in people with and without HIV. All study procedures were approved by the University of Florida's Institutional Review Board (#IRB201801188), and all participants provided their informed consent prior to study participation.

Participants

We recruited participants using contact registries, flyers, and advertisements and in clinics that served persons living with HIV (PLWH) in North Central Florida. Persons were eligible if they were between 21 and 64 years old, living with or without HIV (confirmed through clinical/medical records or testing), reported drinking at least three drinks in a day at least once in the past 30 days, had at least 5 drinking days in the past 30 days, had a stable phone and residence, and could read and write English. Exclusion criteria included evidence of cognitive impairment, uncontrolled and clinically significant physical disease (e.g., uncontrolled diabetes mellitus), medical conditions other than HIV that would contraindicate alcohol use, current use of psychotropic drugs (excluding antidepressants), not using a reliable form of birth control (for women), currently seeking or having past 12-month alcohol treatment, past 30-day drug use (excluding cannabis), current substance use disorder (excluding cannabis), meeting criteria for current severe alcohol use disorder (AUD; 6 criteria endorsed on Structured Clinical Interview for the DSM (SCID)-5; First et al., 2016), or Clinical Institute Withdrawal Assessment (CIWA) score ≥ 8 (Sullivan et al., 1989).

Procedures

Participants were first screened over the phone, and those who were initially eligible were invited for an in-depth in-person or remote screening which included a 30-Day Timeline Followback, medical history, the SCID, and the CIWA. If eligible following this additional screening, participants completed an alcohol administration lab session in which they consumed three standard drinks in a row (30 min for each drink), wore the Skyn biosensor, and provided breathalyzer readings (findings not presented here). Participants then completed the 2-week field test in which they wore the Skyn biosensor and completed EMA surveys. After the field test, participants returned for a second alcohol administration lab session. Due to social distancing necessary during the COVID-19 pandemic, 12 participants only participated in the field test, with all research procedures (field use only without lab sessions) conducted remotely (i.e., the Skyn was mailed to the participant and instructions were provided via Zoom).

BACtrack Skyn—The Skyn biosensor uses fuel cell-based sensors and measures TAC every 20 s. The Skyn also captures the wearer's body temperature. In this study, an early production version obtained in mid-2019 was used (see Figure 1). This version had a battery life of approximately 2 to 3 days. Participants were instructed to wear the Skyn in their daily life for a 2-week period. They were instructed to wear the device continuously, except when charging the device or when exposed to water (e.g., when showering or in the rain) because the bracelet was not waterproof. Participants were instructed to charge the device for at least 1 h each day. The device had to be turned off to charge, so participants were reminded to turn their device off when charging and turn it back on before replacing it on their wrist. It was emphasized that participants keep the device on while drinking and sleeping, thus it was suggested to charge the device when they would need to take it off anyway, for example, while they were showering. Participants were also instructed to avoid contact with alcohol-based products such as hand sanitizer or perfume while wearing the device. Participants installed the Skyn data collection application (app) on their smartphones. The

Skyn app is only compatible with iOS so participants who did not have an iPhone were provided a study phone for the field test period ($n = 20$). The version of the mobile app (version 0.1) used in this study required participants to sync the data stored on the Skyn device by opening the app and waiting for the loading bar to reach 100%. Participants were instructed to sync their data three times per day. When missing data were identified (e.g., 1 day without data), research staff would reach out to the participant to identify and resolve the problem (e.g., technical issues related to uploading) and provide reminders. Upon completion of the field test, a csv file containing time stamped TAC readings with temperature data was downloaded from the BACtrack database for each participant.

TAC processing—Data from the Skyn were processed using the Transdermal Alcohol Sensor Data Macro (TASMAC Software) version 1.8.4 (Barnett et al., 2022). The TASMAC was originally developed for use with SCRAM-CAM data (Barnett et al., 2011) and has been widely used (Alessi et al., 2019; Barnett et al., 2014; Mun et al., 2021; Richards et al., 2021). Participant TAC csv files are uploaded into the TASMAC. Modifications were made to the TASMAC to allow for processing of Skyn data. Since there was no established criteria for detecting alcohol use (when ground truth is not known, that is, in a field setting where number of drinks consumed is unknown to the researcher) based on Skyn TAC readings, we used a low threshold such that the start of a possible alcohol episode was defined as when TAC value reached $>5 \mu\text{g/L}$. Due to the high intensity and noisy nature of the Skyn data, the TASMAC for Skyn can apply a centered moving average to smooth the data¹; a 30-min moving average was chosen for the current study (in 20-s resolution Skyn data, this is equivalent to a moving average of 90 data points). Using these settings defined by the user, the TASMAC identifies drinking episodes and calculates summary data for each, including absorption rate (peak TAC divided by time from baseline to peak TAC), elimination rate (peak TAC divided by time from peak TAC to baseline), peak TAC, and TAC-AUC, and produces graphs for each detected episode (Figure 2).

Drinking episodes were next reviewed by two independent coders (91% agreement) to eliminate detected episodes that were unlikely to be alcohol consumption. Episodes with steep absorption rates, missing or steep elimination rates, very short durations, and very low peak TACs were eliminated based on the team's prior experience with laboratory obtained and pilot field Skyn TAC data (Wang et al., 2021). All possible episodes were coded independently of EMA self-reports (i.e., the coders did not compare possible episodes to EMA self-reports). A third coder was consulted to resolve primary coder discrepancies. Episodes coded as likely to be real drinking episodes were further categorized by data quality (high vs. low). Episodes were considered to be low quality if more than 60 min of data were missing, body temperature dropped below 86°F (a cutoff provided by BACtrack and incorporated into the TASMAC code that produces a warning associated with the episode), or if there was obvious distortion in peak TAC (e.g., an extreme value not likely to be a function of consumed alcohol occurred during a drinking episode)². Drinking episodes

¹This smoothing reduces the effect of extreme values in TAC, something that is commonly seen in Skyn data (Wang et al., 2021).

²As hand sanitizer use, in particular, is common and can be detected by alcohol sensors, it is critical that work with wrist-worn sensors identify optimal methods for excluding data from hand sanitizer and other interferents to ensure accurate identification of alcohol consumption episodes.

that spanned multiple days were manually assigned to the day when peak TAC occurred to capture the day with heaviest alcohol use.

EMA—All EMA assessments were collected via the mobile app mEMA, developed by Ilumivu, Inc. Every day, participants received prompts to complete four surveys: one morning survey and three random surveys assessing items including craving, mood, and pain (not included in current analysis). The morning survey was delivered each day at 10 a.m. and assessed past 24-h alcohol use, including how many standard drinks participants consumed and over what time period the drinking occurred; participants had 30 min to answer this survey. Participants were first asked, “Have you used any alcohol in the past 24 hours?”, if they reported yes, they were asked “When did you start drinking?”, “When did you stop drinking?” and “About how many drinks did you have?” Participants were also instructed to initiate event-contingent surveys during or after drinking episodes to record the number of drinks. Specifically, the event-contingent surveys included the questions, “Do you have a drinking event to report?” and “How many drinks did you have?” Number of drinks reported on these event-contingent surveys were averaged with number of drinks reported on the morning survey when both were available.

Baseline measures

Demographics—Sex (male or female), age (continuous), race (White, Black/African American, Native American, Asian, Multi-racial, Other), and ethnicity (Hispanic/Latino or non-Hispanic/Latino) were collected at the in-person screening.

Clinical characteristics—Research staff measured height and weight (to calculate BMI) for most participants, although data are missing for the 12 who were enrolled during the COVID pandemic.³ HIV status was confirmed through clinical/medical records for PLWH. Persons with unknown or negative HIV status were tested using the Oraquick rapid HIV oral swab test to confirm status.

Hazardous drinking—Participants completed the Alcohol Use Disorder Identification Test (AUDIT), a validated 10-item self-report assessment of past 12-month drinking at the screening visit (Saunders et al., 1993). A score ≥ 8 is consistent with hazardous drinking (Saunders et al., 1993).

Data analysis

Analyses were conducted in SAS 9.4 (SAS Institute). We examined data collected between June 2019 and February 2022. Descriptive statistics on individual-level factors were calculated to characterize the study sample.

A day (defined by social day, 10 a.m. to 9:59 a.m.) was considered a drinking day if there was any drinking episode on that day. All days with both Skyn and EMA data (drinking and non-drinking days), regardless of drinking episode quality, were included in the sensitivity/

³These persons only participated in the field test. All screening and field procedures were conducted remotely. The field test procedures for these participants were identical to those for non-COVID affected participants, but devices were mailed to participants and instructions were provided over Zoom instead of in-person.

specificity analyses ($n = 497$ days from 36 participants). Two by two tables were created to calculate the sensitivity, specificity, PPV, and NPV (using conventional definitions) of the Skyn biosensor in the field by comparing with self-report data.

For evaluating the correspondence between TAC features and self-reported number of drinks, only self-reported drinking days that included a single, high-quality drinking episode were included to ensure that the episode could be paired with self-reported drinking. This was done for two reasons: (1) low-quality drinking episodes were deemed to have inaccurate TAC features (e.g., obviously misrepresented peak TACs) and (2) days with multiple drinking episodes would require aggregating data (e.g., total number of self-reported drinks) but still have just one peak TAC (i.e., highest of the day), thus would result in inaccurate associations. However, ad hoc correlation analyses were conducted using all single-episode drinking days ($n = 150$). Distributions of self-reported number of drinks, peak TAC, and TAC-AUC were examined and descriptive statistics were calculated. For analyses, self-reported number of drinks, peak TAC, and TAC-AUC were log-transformed due to right skewness. To analyze the level of correspondence between the number of self-reported drinks and TAC features from the Skyn, generalized estimating equations models (GEE; normal distributions, autoregressive correlation structure, identity link; Liang & Zeger, 1986) were used.

To explore the association between individual-level factors and TAC features, each factor was separately analyzed in its own GEE model controlling for number of drinks (e.g., GEE model with TAC feature = self-reported number of drinks + one individual-level factor).

RESULTS

A total of 39 participants were included in the parent study, and 36 participants were included in the present analyses due to data processing issues. Participants contributed 497 days worth of data. The sample consisted of 61.1% males with a mean age of 34.3 (SD = 14.8). Approximately half (51.4%) of the participants identified as Non-Hispanic White, 20.0% as Non-Hispanic Black, and 21.7% as Hispanic. The mean BMI of those with available data was 26.5 (considered overweight), 36.1% of the participants were living with HIV, and approximately half (48.6%) met AUDIT criteria for hazardous drinking.

The subsample (i.e., participants with one high-quality, single drinking episode days) used in analysis evaluating correspondence between TAC features and self-reported number of drinks included 29 participants (Table 1).

Aim 1 results

Using our criterion of TAC value $>5 \mu\text{g/L}$, the TASMAL identified 783 drinking episodes over 364 days. Of these days, 50 were missing self-reported data, resulting in 314 TAC-defined drinking days with accompanying self-report data included in analyses. Coders subsequently coded just over one quarter of the 783 episodes ($n = 223$, 28%) as drinking episodes over 182 days (with 20 days of missing self-reported data, resulting in 162 coder-defined drinking days included in analyses).

In general, episodes that coders determined were not drinking episodes (i.e., coder-defined “false positives”) had greater mean peak TACs, absorption rates, elimination rates, and lower mean TAC-AUCs and durations than coder-defined drinking episodes (Table 2).

Aim 2 results

Using TAC-defined drinking days, the Skyn had a sensitivity of 78.1%, a specificity of 55.2%, a PPV of 68.2%, and an NPV of 67.2%. Using coder-defined drinking days, the Skyn had a sensitivity of 54.7%, a specificity of 94.6%, a PPV of 92.6%, and an NPV of 63.0% (Table 3).

Aim 3 results

The final sample size for these analyses was 71 coder-defined drinking days from 29 participants (Figure 3). This subsample of participants is similar to the full sample in terms of demographic and clinical characteristics (Table 1).

Among days with a single, high-quality drinking episode, the average peak TAC per coder-defined drinking day was 85.4 $\mu\text{g/L}$ and ranged from 10.0 to 110.3 $\mu\text{g/L}$. There was a strong significant correlation between peak TAC and self-reported number of drinks ($r = 0.65$, $p < 0.0001$). In the unadjusted GEE model, log self-reported number of drinks was significantly associated with log peak TAC ($\beta = 0.92$, $p < 0.0001$). The average TAC-AUC per coder-defined drinking day was 19,574.4 $\text{h}\cdot\mu\text{g/L}$ and ranged from 193.2 to 181,286.1 $\text{h}\cdot\mu\text{g/L}$ (TAC-AUC is an area, it is presented as the product of time (hours \times TAC level ($\mu\text{g/L}$)), calculated using the sum of the area of trapezoid method). There was a strong significant correlation between TAC-AUC and self-reported number of drinks ($r = 0.63$, $p < 0.0001$). In the unadjusted GEE model, log self-reported number of drinks was significantly associated with log TAC-AUC ($\beta = 1.6$, $p < 0.0001$). Ad hoc correlation analyses using all single-episode drinking days produced similar correlation coefficients (peak TAC and self-reported number of drinks: $r = 0.67$, $p < 0.0001$; TAC-AUC and self-reported number of drinks: $r = 0.63$, $p < 0.0001$).

Aim 4 results

Female sex was associated with greater log peak TAC compared to male sex when controlling for log self-reported number of drinks ($\beta = 0.36$, $p = 0.042$). Age, race/ethnicity, BMI, HIV status, and hazardous drinking were not significantly associated with log peak TAC. Similarly, female sex was associated with a greater log TAC-AUC compared to male sex when controlling for log self-reported number of drinks ($\beta = 0.67$, $p = 0.038$), whereas the other individual level factors were not (Table 4).

DISCUSSION

This study evaluated the concordance between self-reported drinking using daily self-report through EMA and detected drinking by the Skyn, a new-generation, wrist-worn transdermal alcohol biosensor, in a natural environment. Sensitivity, specificity, PPV, and NPV differed based on whether a TAC-defined drinking day (based on minimum TAC alone) or a coder-defined drinking day (based on minimum TAC plus coder review) was used. With

a subset of high-quality, single drinking episode days, the number of self-reported drinks was significantly associated with peak TAC and TAC-AUC. Female sex was significantly associated with greater peak TACs and TAC-AUCs when controlling for number of self-reported drinks.

The Skyn had a much higher sensitivity using TAC-defined drinking days than coder-defined drinking days. That is, 78% of self-reported drinking days were detected using our TAC definition compared to 55% of self-reported drinking days using coder-defined drinking days. However, specificity and PPV were much lower when using the TAC-defined drinking days than when using coder-defined drinking days. With the TAC definition, the 5 $\mu\text{g/L}$ threshold identified only 55% of self-reported non-drinking days, compared to 95% of self-reported non-drinking days following coder review. Further, 68% of TAC-defined drinking days were “true” drinking days (32% false positive), whereas 93% of coder-defined drinking days were “true” drinking days (7% false positive).

The NPV was relatively similar between the two definitions, with close to 65% (63% to 67%) of TAC-defined or coder-defined non-drinking days being “true” non-drinking days. Overall, using a low TAC threshold resulted in fairly high detection of drinking episodes, but also overestimation of drinking episodes (false positives). Adding human coders reduced false positives substantially, but also led to missing more drinking episodes (false negatives). Both improvements in the hardware (i.e., the Skyn) and the software (i.e., the TASMACH) may help improve the sensitivity and specificity of the Skyn.

Using the SCRAM with self-reported data via EMA as the comparison measure, when using the TAC definition, we observed a higher sensitivity (78.1% vs. 61.1%) but a lower specificity (55.2% vs. 84.5%), but when using the coder definition, we observed the opposite, with a lower sensitivity (54.7%) and higher specificity (94.6%; Mun et al., 2021). Both the TAC and coder definitions produced equivalent or higher sensitivities than reported in another study examining the Skyn compared to daily diary data, which reported sensitivities between 40% and 69% for detecting any drinking while using minimum TAC-AUC cutoffs of 56 to 374 h- $\mu\text{g/L}$ to attempt to achieve 90% to 100% specificity (Ash et al., 2022). While the coder definition resulted in a similar specificity to Ash et al.’ (2022) 70% to 100% specificity, the TAC definition was much lower.

The purpose for monitoring and the populations which are being monitored may dictate priorities for favoring a more sensitive set of rules versus a more specific set when detecting alcohol use, thus dictating the criteria used. For example, among persons trying to reduce their drinking on their own, it may be more important to detect drinking when it occurs in real-time. Whereas for probational populations or abstinence-focused trials, accurately depicting no drinking may be more important. As sensitivity increases, specificity will decrease, and vice versa, thus determining which aspect is more important is critical. It is also important to note that PPV and NPV are affected by prevalence, thus as the prevalence of alcohol use increases—the PPV will increase but the NPV will decrease.

The discrepancy in results between TAC-defined and coder-defined drinking days highlights the necessity of continued investigation to further develop criteria to accurately distinguish

between drinking episodes and environmental exposures or other “noise” in the TAC data. TAC readings obtained in the field are more complex and more difficult to interpret due to interferences (e.g., products containing alcohol such as hand sanitizer and hair spray) from a less controlled environment and unobserved behaviors (e.g., variability in compliance), which may also explain the high proportion of low-quality drinking days identified. Coders in our study relied heavily on their prior experience with TAC curves derived from the Skyn TAC in the laboratory setting. However, given the lack of established criteria for drinking episode detection, like the ones established for SCRAM-based TAC readings, we were not able to provide a definite set of rules but rather relied on prior experience with the Skyn TAC data and some general guidelines mentioned in the “Methods” section. Further, manual review of drinking episodes is time consuming and prone to human errors/biases, and reliance on manual review would prevent the Skyn from being used as a real-time intervention tool. As more data on true versus false positive drinking episodes are collected from future studies, researchers may be able to develop a set of criteria like what has been done with the SCRAM-CAM to detect drinking episodes objectively (e.g., minimum TAC = 0.02 g/dl, plus some minimal absorption/elimination rates; Barnett et al., 2011; Roache et al., 2019). Also, previous research has demonstrated the potential of using machine learning to estimate BrAC, and similar algorithms might be developed to detect drinking episodes (Fairbairn et al., 2020). Our findings suggest that minimum TAC, peak TAC, TAC-AUC, episode duration, absorption and elimination rates, and sex should be considered in such approaches.

It is possible that self-reported drinking days were misreported, resulting in self-reported and Skyn-detected drinking days not being aligned, which could explain some of the false positives and false negatives. In some cases, drinking episodes spanned over multiple days, and we assigned such episodes to the day that the peak TAC occurred. There were also many instances of days with multiple detected drinking episodes (i.e., 783 detected episodes over 497 days), possibly due to multiple “real” drinking episodes with enough time in between them to reach baseline TAC levels or due to potential environmental exposures in addition to “real” drinking episodes. The lag time between when drinking occurred and TAC could also explain some false positives and negatives, as drinking may have occurred on 1 day, but a positive TAC or the peak TAC was not detected until the following day. However, the time lag between TAC and BrAC for the Skyn is shorter than that of the SCRAM-CAM, averaging 24 min compared to 69 min, respectively (Fairbairn & Kang, 2019), and may not have had a significant effect. We also considered both self-reported drinking days and Skyn-detected drinking days over a social day, from 10 a.m. to 9:59 a.m., and it is possible that participants did not always report drinking for a social day, rather they may have reported drinking for a standard day (12 a.m. to 11:59 p.m.).

Using either high-quality only or all single episode drinking days, correlation coefficients in the 0.6s were observed. While these represent strong correlations, it is clear that the relationship between self-reported number of drinks and TAC is not perfect. It is reassuring, however, that even when including low-quality drinking episodes where peak TAC or TAC-AUC may be inaccurate, strong relationships between the variables were maintained. While the sensitivity/specificity data indicate that Skyn TAC data alone may not

be sufficient for distinguishing between drinking and nondrinking days, this finding suggests that distinguishing between high versus low quality drinking episodes may not be necessary.

Similar to the findings based on the SCRAM-CAM ankle biosensor (Richards et al., 2021), most individual-level factors were not associated with peak TAC or TAC-AUC captured by the Skyn, after controlling for alcohol use level. Female sex was associated with larger peak TACs and larger TAC-AUCs, consistent with literature examining this relationship using the SCRAM-CAM (Barnett et al., 2014; Hill-Kapturczak et al., 2015; Richards et al., 2021). In future model development, sex should be considered. However, it should be noted that our sample was small and may be underpowered to detect other group differences.

Strengths/limitations and conclusions

This research is among the first to examine the performance of the Skyn under naturalistic drinking conditions. Findings from this work demonstrate the potential of incorporating the wrist-worn biosensor for remote data collection but highlight the need for further investigation in terms of drinking episode detection. Use of EMA to collect self-reported alcohol use is a strength of this study, as number of drinking episodes and number of drinks per drinking episode are less prone to recall bias than with other retrospective methods (Piasecki, 2019).

Using the TASMACH allowed us to import large quantities of biosensor data to obtain TAC measurements, but the criteria for detecting drinking episodes using TAC from the Skyn are not established, which is why we conducted manual review of each detected drinking episode. An important next step of this work is to determine stricter guidelines for distinguishing whether a detected drinking episode is real or not. A limitation of this work is that of the 121 single episode drinking days identified by both Skyn and EMA, over 40% were low quality, thus were not included in the GEE analyses (aims 3 and 4). Thus, the inclusion of only high-quality drinking days in our GEE models may have resulted in inflated correlations. However, post hoc correlation analyses were conducted that suggest similar relationships between TAC and self-reported number of drinks, regardless of data quality. Additionally, as with all new technologies, revisions to the devices may reduce the applicability of this research.

In sum, results of the current study demonstrate that Skyn-derived TAC can be used to measure alcohol use under naturalistic drinking conditions but also suggest Skyn TAC drinking episode data alone may not be fully trustworthy. Additional research to accurately identify drinking episodes based on Skyn TAC readings is warranted.

ACKNOWLEDGMENTS

This study was supported by the National Institute on Alcohol Abuse and Alcoholism (R21 AA027191) and by the National Institute on Drug Abuse (T32 DA017629).

Funding information

National Institute on Alcohol Abuse and Alcoholism, Grant/Award Number: R21AA027191; National Institute on Drug Abuse, Grant/Award Number: T32 DA017629

REFERENCES

- Alessi SM, Barnett NP & Petry NM (2017) Experiences with SCRAMx alcohol monitoring technology in 100 alcohol treatment outpatients. *Drug and Alcohol Dependence*, 178, 417–424. [PubMed: 28709081]
- 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. [PubMed: 30710610]
- Ash GI, Gueorguieva R, Barnett NP, Wang W, Robledo DS, DeMartini KS et al. (2022) Sensitivity, specificity, and tolerability of the BACTrack *Skyn* compared to other alcohol monitoring approaches among young adults in a field-based setting. *Alcoholism, Clinical and Experimental Research*, 46, 783–796. [PubMed: 35567595]
- Barnett NP (2015) Alcohol sensors and their potential for improving clinical care: editorial. *Addiction*, 110, 1–3.
- 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. [PubMed: 28107772]
- Barnett NP, Meade EB & Glynn TR (2014) Predictors of detection of alcohol use episodes using a transdermal alcohol sensor. *Experimental and Clinical Psychopharmacology*, 22, 86–96. [PubMed: 24490713]
- Barnett NP, Souza T, Rosen I, Luczak SE, Glynn TR & Swift R (2022) transdermal alcohol sensor data macro. <https://www.brown.edu/academics/public-health/research/alcohol-addiction-studies/tasmac/>
- 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 and Alcohol Dependence*, 118, 391–399. [PubMed: 21665385]
- 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. *Experimental and Clinical Psychopharmacology*, 20, 373–381. [PubMed: 22708608]
- Dougherty DM, Hill-Kapturczak N, Liang Y, Karns TE, Cates SE, Lake SL et al. (2014) Use of continuous transdermal alcohol monitoring during a contingency management procedure to reduce excessive alcohol use. *Drug and Alcohol Dependence*, 142, 301–306. [PubMed: 25064019]
- Dougherty DM, Hill-Kapturczak N, Liang Y, Karns TE, Lake SL, Cates SE et al. (2015) The potential clinical utility of transdermal alcohol monitoring data to estimate the number of alcoholic drinks consumed. *Addictive Disorders & Their Treatment*, 14, 124–130. [PubMed: 26500459]
- Dougherty DM, Karns TE, Mullen J, Liang Y, Lake SL, Roache JD et al. (2015) Transdermal alcohol concentration data collected during a contingency management program to reduce at-risk drinking. *Drug and Alcohol Dependence*, 148, 77–84. [PubMed: 25582388]
- Fairbairn CE & Bosch N (2021) A new generation of transdermal alcohol biosensing technology: practical applications, machine-learning analytics and questions for future research. *Addiction*, 116, 2912–2920. [PubMed: 33908674]
- Fairbairn CE & Kang D (2019) Temporal dynamics of transdermal alcohol concentration measured via new-generation wrist-worn biosensor. *Alcoholism, Clinical and Experimental Research*, 43, 2060–2069. [PubMed: 31469451]
- Fairbairn CE, Kang D & Bosch N (2020) Using machine learning for real-time BAC estimation from a new-generation transdermal biosensor in the laboratory. *Drug and Alcohol Dependence*, 216, 108205. [PubMed: 32853998]
- First MB, Williams JBW, Karg RS & Spitzer RL (2016) Structured clinical interview for DSM-5 disorders: scid-5-cv: clinician version. Arlington, VA: American Psychiatric Association Publishing.
- 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. [PubMed: 24923985]

- Kilmer B, Nicosia N, Heaton P & Midgette G (2013) Efficacy of frequent monitoring with swift, certain, and modest sanctions for violations: insights from South Dakota's 24/7 sobriety project. *American Journal of Public Health*, 103, e37–e43. [PubMed: 23153129]
- Liang K-Y & Zeger SL (1986) Longitudinal data analysis using generalized linear models. *Biometrika*, 73, 13–22.
- Litten RZ, Bradley AM & Moss HB (2010) Alcohol biomarkers in applied settings: recent advances and future research opportunities: alcohol biomarkers in applied settings. *Alcoholism, Clinical and Experimental Research*, 34, 955–967. [PubMed: 20374219]
- Mathias CW, Hill-Kapturczak N, Karns-Wright TE, Mullen J, Roache JD, Fell JC et al. (2018) Translating transdermal alcohol monitoring procedures for contingency management among adults recently arrested for DWI. *Addictive Behaviors*, 83, 56–63. [PubMed: 29397211]
- Merrill J, Gunn R, Neary A, Souza T & Barnett N (2022) Feasibility and acceptability of using a wrist-worn transdermal alcohol biosensor to collect data in the field. In: *Proceedings of the 55th Hawaii international conference on system sciences*. Honolulu, HI: University of Hawai'i.
- Mun E, Li X, Businelle MS, Hébert ET, Tan Z, Barnett NP et al. (2021) Ecological momentary assessment of alcohol consumption and its concordance with transdermal alcohol detection and timeline follow-back self-report among adults experiencing homelessness. *Alcoholism, Clinical and Experimental Research*, 45, 864–876. [PubMed: 33583057]
- Norman T, Peacock A, Ferguson SG, Kuntsche E & Bruno R (2021) Combining transdermal and breath alcohol assessments, real-time drink logs and retrospective self-reports to measure alcohol consumption and intoxication across a multi-day music festival. *Drug and Alcohol Review*, 40, 1112–1121. [PubMed: 33174260]
- Northcote J & Livingston M (2011) Accuracy of self-reported drinking: observational verification of 'last occasion' drink estimates of young adults. *Alcohol*, 46, 709–713.
- Piasecki TM (2019) Assessment of alcohol use in the natural environment. *Alcoholism, Clinical and Experimental Research*, 43, 564–577. [PubMed: 30748019]
- Richards VL, Liu Y, Orr J, Leeman RF, Barnett NP, Bryant K et al. (2021) Sociodemographic and clinical factors associated with transdermal alcohol concentration from the SCRAM biosensor among persons living with and without HIV. *Alcoholism, Clinical and Experimental Research*, 45, 1804–1811. [PubMed: 34342009]
- 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. [PubMed: 30179708]
- Rosenberg M, Ludema C, Kianersi S, Luetke M, Jozkowski K, Guerra-Reyes L et al. (2021) Wearable alcohol monitors for alcohol use data collection among college students: feasibility and acceptability in a pilot study. *Epidemiology*. Available from: 10.1101/2021.02.17.21251959. Preprint.
- Russell MA, Turrise RJ & Smyth JM (2022) Transdermal sensor features correlate with ecological momentary assessment drinking reports and predict alcohol-related consequences in young adults' natural settings. *Alcoholism, Clinical and Experimental Research*, 46, 100–113. [PubMed: 35066894]
- 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. [PubMed: 8329970]
- Sullivan JT, Sykora K, Schneiderman J, Naranjo CA & Sellers EM (1989) Assessment of alcohol withdrawal: the revised clinical institute withdrawal assessment for alcohol scale (CIWA-Ar). *Addiction*, 84, 1353–1357.
- Swift R (2003) Direct measurement of alcohol and its metabolites. *Addiction*, 98(Suppl 2), 73–80. [PubMed: 14984244]
- Villalba K, Cook C, Dévieux JG, Ibanez GE, Oghogho E, Neira C et al. (2020) Facilitators and barriers to a contingency management alcohol intervention involving a transdermal alcohol sensor. *Heliyon*, 6, e03612. [PubMed: 32258468]

Wang Y, Fridberg DJ, Shortell DD, Leeman RF, Barnett NP, Cook RL et al. (2021) Wrist-worn alcohol biosensors: applications and usability in behavioral research. *Alcohol*, 92, 25–34. [PubMed: 33609635]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



FIGURE 1.
BACtrack Skyn. Views of the alcohol biosensor used in this study

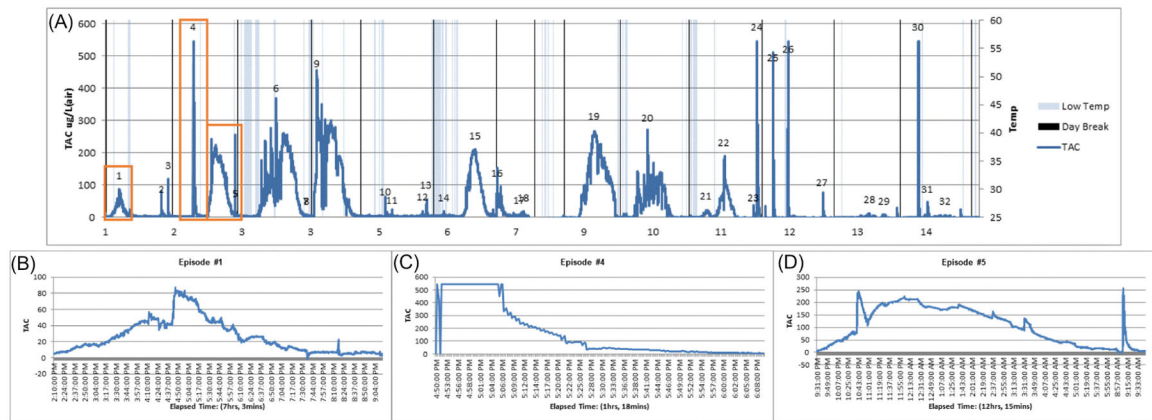
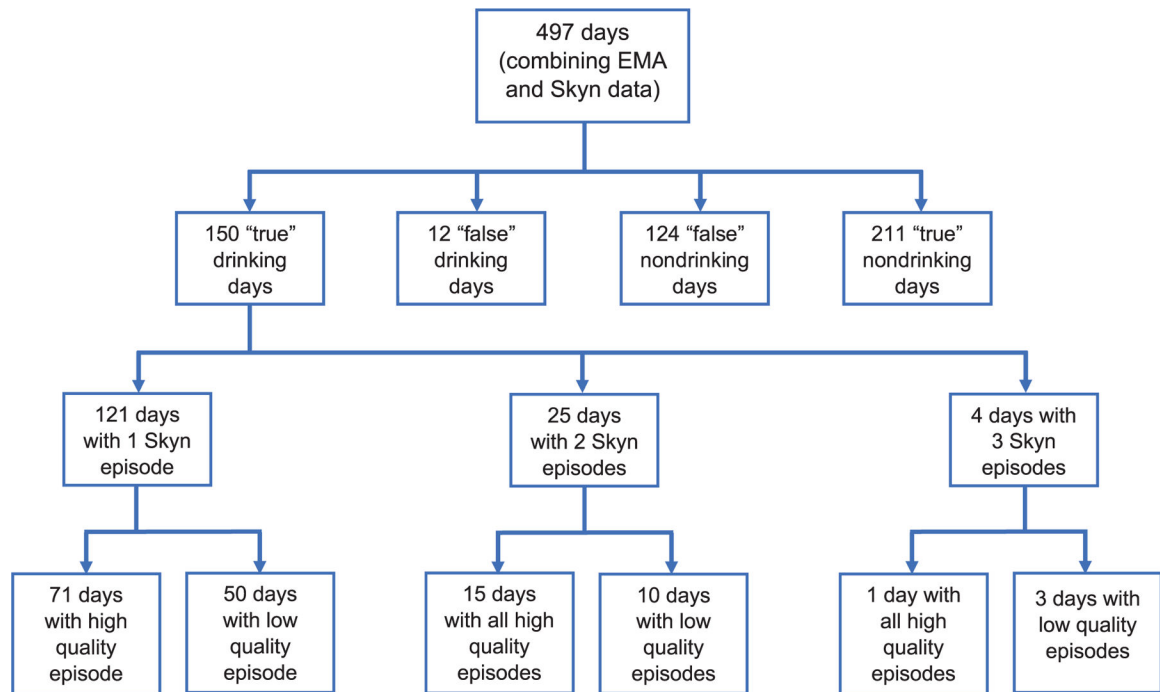


FIGURE 2.

Drinking episode graphs. (A) Transdermal alcohol concentration (TAC) graph for a single participant over 14 days produced by the TASMAC (transdermal alcohol sensor data macro) software. Using researcher-defined criterion of $5 \mu\text{g/L}$ (TAC-defined), 32 drinking episodes were identified, but not all TAC-defined drinking episodes reflect consumed alcohol. (B) Episode 1 was coder-defined to be a true, high-quality drinking episode. (C) Episode 4 was coded to be an environmental exposure, thus was not considered a drinking episode. This episode was determined to be an environmental exposure because of the very high absorption rate and relatively quick decline to TAC-defined baseline. (D) Episode 5 was coded to be a true, low-quality drinking episode; it was considered low-quality because of two sharp peaks (10:43 p.m. and 9:08 a.m.) that were coded as not reflecting consumed alcohol.

**FIGURE 3.**

Skyn and EMA drinking day selection. After combining the EMA and Skyn data, there were 497 observations (including drinking and non-drinking days). For aim 2 (comparing the sensitivity, specificity, PPV, and NPV), all observations from Skyn and EMA were used. “True” refers to coder-defined drinking and nondrinking days where Skyn and EMA agree. “False” refers to coder-defined drinking and nondrinking days where Skyn and EMA do not agree. For aims 3 and 4 (determining the level of correspondence between quantitative TAC values and self-reported number of drinks by EMA; exploring what individual-level factors are associated with quantitative TAC values when controlling for self-reported drinks), only the coder-defined, single-Skyn-episode days with high quality data were used.

TABLE 1

Sample characteristics

	Full sample (<i>n</i> = 36) ^a	Subsample (<i>n</i> = 29) ^b
Sex		
Male	22 (61%)	18 (62%)
Female	14 (39%)	11 (38%)
Age		
Mean (SD)	34.3 (14.8)	34.7 (15.4)
Race/Ethnicity ^c		
Non-Hispanic, White	18 (51%)	14 (50%)
Other	17 (49%)	14 (50%)
Body mass index		
Mean (SD)	26.5 (5.7) ^d	25.4 (4.8) ^e
HIV status		
Negative	23 (64%)	19 (66%)
Positive	13 (36%)	10 (34%)
AUDIT category ^{c,f}		
Nonhazardous drinking	18 (51%)	14 (50%)
Hazardous drinking	17 (49%)	14 (50%)

Abbreviation: AUDIT, Alcohol Use Disorder Identification Test.

^aParticipants included in sensitivity/specificity analyses.

^bParticipants included in generalized estimating equations models, includes participants with high-quality, single drinking episode days.

^cData missing for 1 participant.

^dData missing for 12 participants.

^eData missing for 10 participants.

^fAUDIT scores <8 were categorized as nonhazardous drinking, AUDIT scores ≥ 8 were categorized as hazardous drinking.

TABLE 2

Mean TAC values for coder-defined nondrinking and drinking episodes

	Mean (SD)	Median (range)
Coder-defined nondrinking episodes (<i>n</i> = 560)		
Absorption rate (TAC/h)	4585.1 (8247.0)	493.8 (1.2 to 36,962.4)
Elimination rate (TAC/h)	477.0 (692.3)	198.7 (0.0 ^a to 4930.8)
Peak TAC (μg/L)	173.1 (203.0)	67.8 (0.0 ^a to 616.2)
TAC-AUC (h·μg/L)	2711.7 (10,284.1)	449.3 (5.1 to 134,823.9)
Duration (h)	1.1 (2.5)	0.4 (0.02 to 30.5)
Coder-defined drinking episodes (<i>n</i> = 223) ^b		
Absorption rate (TAC/h)	63.3 (71.8)	35.2 (2.3 to 430.3)
Elimination rate (TAC/h)	57.4 (135.8)	29.5 (1.7 to 1798.2)
Peak TAC (μg/L)	95.8 (113.7)	42.8 (10.0 to 615.5)
TAC-AUC (h·μg/L)	19,908.6 (35,640.3)	3925.3 (143.6 to 209,730.7)
Duration (h)	5.7 (5.9)	3.4 (0.3 to 28.9)
Coder-defined high-quality drinking episodes (<i>n</i> = 142)		
Absorption rate (TAC/h)	60.4 (61.1)	38.3 (5.6 to 331.8)
Elimination rate (TAC/h)	42.3 (42.4)	27.5 (5.8 to 250.7)
Peak TAC (μg/L)	66.8 (75.0)	35.3 (10.0 to 415.9)
TAC-AUC (h·μg/L)	12,182.6 (25,538.3)	3085.9 (143.6 to 181,332.0)
Duration (h)	4.2 (4.4)	2.5 (0.3 to 22.3)
Coder-defined low-quality drinking episodes (<i>n</i> = 81)		
Absorption rate (TAC/h)	68.3 (87.6)	33.7 (2.3 to 430.3)
Elimination rate (TAC/h)	83.9 (216.5)	35.5 (1.7 to 1798.2)
Peak TAC (μg/L)	146.6 (147.8)	88.1 (13.4 to 615.5)
TAC-AUC (h·μg/L)	33,452.9 (45,647.5)	10,440.5 (226.1 to 209,730.7)
Duration (h)	8.4 (7.2)	6.0 (0.5 to 28.9)

Note: There was 90.9% coder agreement for nondrinking versus drinking episodes and 79.4% coder agreement for high-quality versus low-quality drinking episodes.

Abbreviations: AUC, area under the curve; TAC, Transdermal alcohol concentration.

^aValues = 0.0 indicate inability to calculate the value.

includes high-quality and low-quality drinking episodes.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

TABLE 3Drinking days by self-report and Skyn biosensor ($n = 36$)

	Self-report			
	Yes	No		
TAC-defined drinking days				
Skyn				
Yes	214	100	314	
No	60	123	183	
	274	223	497	
Sensitivity = 78.1%				
Specificity = 55.2%				
PPV = 68.2%				
NPV = 67.2%				
Coder-defined drinking days				
Skyn				
Yes	150	12	162	
No	124	211	335	
	274	223	497	
Sensitivity = 54.7%				
Specificity = 94.6%				
PPV = 92.6%				
NPV = 63.0%				

Abbreviations: NPV, negative predictive value; PPV, positive predictive value.

TABLE 4

Individual models of the associations between individual-level factors and log-TAC^a (*n* = 71 drinking days)

	Beta	<i>p</i> -value
Peak TAC and sex		
Number of self-reported drinks	0.94	<0.0001
Female sex ^b	0.36	0.0418
Peak TAC and age		
Number of self-reported drinks	0.92	<0.0001
Age	0.005	0.4159
Peak TAC and race/ethnicity		
Number of self-reported drinks	0.95	<0.0001
Other race/ethnicity ^c	-0.31	0.1025
Peak TAC and BMI		
Number of self-reported drinks	0.85	<0.0001
BMI	-0.03	0.1608
Peak TAC and HIV status		
Number of self-reported drinks	0.93	<0.0001
Positive HIV status ^d	-0.17	0.3963
Peak TAC and hazardous drinking		
Number of self-reported drinks	0.94	<0.0001
Hazardous drinking ^e	0.10	0.6342
TAC-AUC and sex		
Number of self-reported drinks	1.58	<0.0001
Female sex ^b	0.67	0.0383
TAC-AUC and age		
Number of self-reported drinks	1.56	<0.0001
Age	0.01	0.3497
TAC-AUC and race/ethnicity		
Number of self-reported drinks	1.57	<0.0001
Other race/ethnicity ^c	-0.28	0.4557
TAC-AUC and BMI		
Number of self-reported drinks	1.41	0.0002
BMI	-0.04	0.3870
TAC-AUC and HIV-status		
Number of self-reported drinks	1.57	<0.0001
Positive HIV status ^d	0.07	0.8580
TAC-AUC and hazardous drinking		
Number of self-reported drinks	1.58	<0.0001
Hazardous drinking ^e	0.28	0.4856

Abbreviations: AUC, area under the curve; TAC, transdermal alcohol concentration area under the curve.

^aAll TAC values and number of self-reported drinks were log-transformed.

^bReference group = males.

^cReference group = non-Hispanic White; Other refers to Non-Hispanic Black and Hispanic individuals.

^dReference group = HIV negative.

^eReference group = non-hazardous drinking as measured by the Alcohol Use Disorder Identification Test (AUDIT); hazardous drinking defined by a score ≥ 8 on the AUDIT.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript