Alcohol xxx (xxxx) xxx

Contents lists available at ScienceDirect

Alcohol

journal homepage: http://www.alcoholjournal.org/

# Acceptability and validity of using the BACtrack skyn wrist-worn transdermal alcohol concentration sensor to capture alcohol use across 28 days under naturalistic conditions – A pilot study

Jimikaye B. Courtney*, Michael A. Russell, David E. Conroy

*College of Health and Human Development, Pennsylvania State University, University Park, Pennsylvania, 16802, United States*

ARTICLE INFO

*Article history:*
Received 23 June 2022
Received in revised form
9 November 2022
Accepted 30 November 2022

*Keywords:*
acceptability
ecological momentary assessments
self-reported drinking
transdermal alcohol concentration (TAC)
sensors
user burden
validity

ABSTRACT

Wrist-worn transdermal alcohol concentration (TAC) sensors have the potential to provide detailed information about day-level features of alcohol use but have rarely been used in field-based research or in early adulthood (i.e., 26–40 years) alcohol users. This pilot study assessed the acceptability, user burden, and validity of using the BACtrack Skyn across 28 days in individuals' natural settings. Adults aged 26–37 (N = 11, $M_{age}$ = 31.2, 55% female, 73% non-Hispanic white) participated in a study including retrospective surveys, a 28-day field protocol wearing Skyn and SCRAM sensors and completing ecological momentary assessments (EMA) of alcohol use and duration (daily morning reports and participant-initiated start/stop drinking EMAs), and follow-up interviews. Day-level features of alcohol use extracted from self-reports and/or sensors included drinks consumed, estimated Blood Alcohol Concentration (eBAC), drinking duration, peak TAC, area under the curve (AUC), rise rate, and fall rate. Repeated-measures correlations ($r_{rm}$) tested within-person associations between day-level features of alcohol use from the Skyn versus self-report or the SCRAM. Participants preferred wearing the Skyn over the SCRAM [$t(10) = -6.79$, $p < .001$, $d = 2.74$]. Skyn data were available for 5614 (74.2%) out of 7566 h, with 20.7% of data lost due to syncing/charging issues and 5.1% lost due to device removal. Skyn agreement for detecting drinking days was 55.5% and 70.3% when compared to self-report and the SCRAM, respectively. Correlations for drinking intensity between self-report and the Skyn were 0.35 for peak TAC, 0.52 for AUC, and 0.30 for eBAC, which were smaller than correlations between self-report and SCRAM, at 0.78 for peak TAC, 0.79 for AUC, and 0.61 for eBAC. Correlations for drinking duration were larger when comparing self-report to the Skyn ($r_{rm} = 0.36$) versus comparing self-report to the SCRAM ($r_{rm} = 0.31$). The Skyn showed moderate-to-large, significant correlations with the SCRAM for peak TAC ($r_{rm} = 0.54$), AUC ($r_{rm} = 0.80$), and drinking duration ($r_{rm} = 0.63$). Our findings support the acceptability and validity of using the Skyn for assessing alcohol use across an extended time frame (i.e., 28 days) in individuals' natural settings, and for providing useful information about day-level features of alcohol use.



## Introduction

Over half of U.S. adults consumed some alcohol in the past month (Substance Abuse and Mental Health Services Administration, 2019). Approximately 88,000 people in the U.S. die annually from alcohol-related causes, and 31% of all U.S. driving-related fatalities are due to alcohol-impaired driving (National Institute on Alcohol Abuse and Alcoholism, 2021). Short-term effects of alcohol use, such as injuries, occur during or soon after drinking (Clapp et al., 2018; Clapp, Madden, Mooney, & Dahlquist, 2017). Long-term effects, such as risk of disease or alcohol dependence, result from frequent (e.g., daily) and repeated alcohol exposure over time (Clapp et al., 2018; Holder, 2006; Rehm et al., 2009). Alcohol exposure varies across different age ranges, with emerging adulthood (18–25 years) characterized by increases in alcohol use (e.g., heavy drinking) and abuse that typically either become entrenched or resolve after age 25 (White & Jackson, 2004). After age 25, individuals' alcohol use patterns are typically well-established; therefore, examining alcohol use behaviors after age 25 (i.e., 26–40 years, "early adulthood") provides a more

* Corresponding author. University of North Carolina at Chapel Hill, G412 Fetzer Hall CB#8700, Chapel Hill, NC, 27599, USA. Tel.: 919-445-1520.
*E-mail address:* jimikaye@unc.edu (J.B. Courtney).

https://doi.org/10.1016/j.alcohol.2022.11.004
0741-8329/

accurate representation of what we would expect to characterize an individual's typical alcohol use behavior, and thus their exposure that would impact disease risk (Substance Abuse and Mental Health Services Administration, 2019; White & Jackson, 2004). Despite the significant public health burden associated with the short- and long-term effects of alcohol use, researchers currently have a limited understanding of individuals' daily alcohol use exposures across extended timeframes. Due to the complex, sporadic, and cyclical nature of alcohol use (Greenfield & Kerr, 2003), characterizing daily alcohol use exposure across weeks or months is necessary and will help inform understanding of the short- and long-term effects of alcohol use and risk for alcohol dependence.

To date, the majority of research characterizing daily alcohol use has relied on individuals self-reporting number of drinks consumed. Self-report is a straightforward, acceptable, valid, and reliable method (Piasecki, 2019; Simons, Wills, Emery, & Marks, 2015; Thornton et al., 2022). However, it is limited by the difficulty of accurately estimating drink volume and ethanol content (Bond, Greenfield, Patterson, & Kerr, 2014) and by the neurocognitive effects of alcohol (Hultgren, Scaglione, Buben, & Turrisi, 2020). Self-report may over- or under-estimate alcohol use (Alessi, Barnett, & Petry, 2019; Merrill, Fan, Wray, & Miranda, 2020), is subject to social desirability biases and stigma associated with alcohol use (Davis, Thake, & Vilhena, 2010), and is limited by participant burden and the inability to capture information about alcohol exposure after consumption stops (Piasecki, 2019). Daily self-reports cannot capture many meaningful characteristics of daily alcohol use, such as the biological alcohol concentration, duration, and pattern of exposure, and an individual's proximal risk (Hultgren et al., 2020; Piasecki, 2019). These limitations of self-report have increased interest in developing feasible, acceptable, and reliable methods for capturing sensor-based biological measures of daily alcohol use exposure in individuals' natural settings (Wang, Fridberg, Leeman, Cook, & Porges, 2019).

Several methods already exist for capturing biological measures of alcohol use exposure, including sensors that measure blood and breath alcohol concentration (BAC and BrAC, respectively) and transdermal alcohol concentration (TAC). Sensors measuring BAC or BrAC only provide information about recent alcohol consumption due to the rapid metabolism of alcohol, thereby requiring more frequent repeated measures that may be too burdensome or invasive for feasibly capturing alcohol use in natural settings (Campbell, Kim, & Wang, 2018; Fairbairn & Kang, 2019; Leffingwell et al., 2013; van Egmond, Wright, Livingston, & Kuntsche, 2020). In contrast, TAC sensors provide a mostly passive and non-invasive approach for continuously monitoring alcohol use. TAC sensors are worn on the ankle or wrist and measure the 1% of alcohol that is eliminated through the skin via insensible perspiration or sweat, such that there is a temporal delay (30 min–5 h) between alcohol consumption and detection via TAC sensors (Barnett, 2015; Leffingwell et al., 2013; Piasecki, 2019; Sakai, Mikulich-Gilbertson, Long, & Crowley, 2006; Swift, 2003). TAC is not quantitatively equivalent to BAC or BrAC and there are not currently reliable methods for translating TAC to BAC due to the variety of factors that can impact individual differences in TAC (e.g., sex, age, skin temperature, motion) (Fairbairn & Bosch, 2021; Luczak et al., 2018; Luczak & Ramchandani, 2019; Swift, 2000, 2003; Wang et al., 2019). However, TAC still provides meaningful information about an individual's relative, rather than absolute, alcohol exposure and offers comprehensive information that can be used to richly characterize daily biological alcohol exposure (Piasecki, 2019; Wang et al., 2019).

One of the most widely used TAC sensors with the strongest evidence of validity and reliability is the Secure Continuous Remote Alcohol Monitor (SCRAM), which is worn on the ankle and samples TAC every 30 min (Barnett, Meade, & Glynn, 2014; Karns-Wright et al., 2017; Marques & McKnight, 2009). The SCRAM demonstrates moderate-to-large positive correlations with self-reported alcohol use and BrAC (Alessi et al., 2019; Karns-Wright et al., 2017; Leffingwell et al., 2013; Russell, Turrisi, & Smyth, 2022; Sakai et al., 2006), and it has established rules for cleaning data and detecting alcohol use (Roache et al., 2019). The SCRAM has been used to capture alcohol use in natural settings across a single day (Clapp et al., 2017) and for up to 16 weeks (Dougherty et al., 2015). Yet the SCRAM is bulky, about the size of a card deck and weighing 164.4 g/5.8 oz., which can cause participant discomfort and interfere with engaging in activities like running or contact sports (Alessi, Barnett, & Petry, 2017; Ash et al., 2022; Barnett et al., 2017; Barnett, Tidey, Murphy, Swift, & Colby, 2011; Sakai et al., 2006). Participants' most common complaint about wearing the SCRAM is embarrassment due to its similarity to house arrest monitors, with some individuals refusing to participate due to social stigma (Alessi et al., 2017; Ash et al., 2022; Barnett et al., 2011). The SCRAM's quasi-continuous sampling and large sampling interval limit its ability to capture the dynamics of alcohol exposure (Piasecki, 2019; Wang et al., 2019), a challenge that is amplified by SCRAM TAC values lagging 1–4 h behind BrAC (Fairbairn & Kang, 2019; Leffingwell et al., 2013; Marques & McKnight, 2009; Sakai et al., 2006; van Egmond et al., 2020). Recent advances have led to the development of wrist-worn alcohol biosensors that address many limitations of the SCRAM.

One promising wrist-worn sensor is the BACtrack Skyn. The Skyn is a small, lightweight, and unobtrusive device, which may increase participant acceptability for daily use (Ash et al., 2022; Campbell et al., 2018; Fairbairn & Kang, 2019; van Egmond et al., 2020; Wang et al., 2019). The Skyn provides more frequent measures than the SCRAM, sampling TAC every 20 s (Roache et al., 2019; Rosenberg et al., 2021). A few studies have tested the Skyn under controlled laboratory conditions. One found that Skyn TAC correlated strongly with BrAC, and the Skyn had a significantly shorter lag in detecting peak TAC versus peak BrAC when compared to the SCRAM's detection of peak TAC, with a mean difference of 66 min (Fairbairn & Kang, 2019). However, 18–28% of Skyn prototypes failed (depending on the metric), whereas only 2% of SCRAMs failed (Fairbairn & Kang, 2019). In another study, researchers employing machine-learning algorithms successfully estimated BAC using Skyn data; however, 16–24% of Skyn devices failed (Fairbairn, Kang, & Bosch, 2020). A third study found good agreement between Skyn and SCRAM TAC curves, although Skyn data were considerably "noisier" than SCRAM data, likely due to the Skyn's higher sampling rate (Wang et al., 2021). These studies provide preliminary evidence for the validity of Skyn prototype TAC data under controlled laboratory conditions; however, it is also necessary to test the Skyn in natural settings for extended periods of time as it is often infeasible and unethical to dose participants in laboratory settings to the level they dose themselves in natural settings, and the characteristics of alcohol use may be more varied and/or severe in natural environments.

To our knowledge, only four studies have tested the Skyn in natural settings. These studies included samples of five to 47 moderate-to-heavy drinking adults (age range: 18–38 years) with monitoring periods ranging from a single drinking episode to 14 days (Ash et al., 2022; Rosenberg et al., 2021; Wang et al., 2021). One study using the Skyn T10 model (a post-prototype model) showed Skyn sensitivity ranged from 40 to 69%, 67–89%, and 82–100% for detecting any drinking, moderate drinking, or heavy drinking, respectively, and specificity ranged from 70 to 100%, with differences depending on the cohort, sample-specific cut-offs used to detect drinking, and the age of the device (Ash et al., 2022). In the same sample, the SCRAM sensitivity ranged from 41 to 53%, 46–72%, and 50–91% for detecting any drinking, moderate drinking, or heavy drinking, respectively; however, the SCRAM consistently showed 100% specificity across cohorts (Ash et al., 2022). Other studies using post-prototype Skyn devices [model number(s) not specified] found it detected 12 out of 15 self-reported drinking events across 5 days (Rosenberg et al., 2021), and it failed to detect episodes with only one or two drinks across 5 or 14 days (Rosenberg et al., 2021; Wang et al., 2021). The Skyn showed a mean delay of 35.6 min in detecting self-reported drinking onset (Wang et al., 2021). There were strong correlations between Skyn-detected drinking start time and area under the curve and self-reported drinking start time and number of drinks, respectively (Rosenberg et al., 2021). It was difficult to distinguish TAC elevations due to biological versus environmental alcohol exposure (Wang et al., 2021), and drinking detection rules were not robust due to TAC thresholds differing widely between cohorts and device lots (Ash et al., 2022). Participants reported high acceptability for using the Skyn (Ash et al., 2022; Rosenberg et al., 2021), greater tolerability and lower burden for using the Skyn versus the SCRAM (Ash et al., 2022), and similar usability and acceptability ratings for the Skyn and self-report (Wang et al., 2021). However, participants noted it was difficult to upload Skyn data, which resulted in data loss (Wang et al., 2021). Participants wanted the device to be waterproof, have a longer battery life, and to include notifications about battery level and upload progress (Wang et al., 2021). The Skyn app now includes notifications about data upload progress. These studies provide support for the acceptability and feasibility of using the Skyn across 5 or 14 days in natural settings. However, additional research is needed to test acceptability, feasibility, and validity of using the Skyn across extended timeframes (e.g., 14 or more days) (Ash et al., 2022; Fairbairn & Bosch, 2021; van Egmond et al., 2020; Wang et al., 2019). Collecting Skyn data across extended timeframes is needed for providing sufficient coverage of alcohol use to observe risk processes, given that consumption patterns are complex, sporadic, and cycle weekly (e.g., social weekends versus weekdays) and seasonally (e.g., holidays) (Finlay, Ram, Maggs, & Caldwell, 2012; Greenfield & Kerr, 2003). Collecting Skyn data across extended timeframes during early adulthood is valuable for understanding alcohol exposure and its risks, given that early adulthood is the time when individuals' alcohol use patterns become established and more characteristic of typical alcohol use exposure (Substance Abuse and Mental Health Services Administration, 2019; White & Jackson, 2004). Research is also needed to examine the validity of daily measures of alcohol use exposure captured across extended timeframes via the Skyn, compared to self-report and other device-based measures of alcohol use (e.g., SCRAM) (Ash et al., 2022; Fairbairn & Bosch, 2021; van Egmond et al., 2020; Wang et al., 2019).

The primary purpose of this pilot study was to assess the acceptability, user burden, and validity of using the Skyn across 28 days in early adults' natural settings. This population was selected because individuals' alcohol use patterns are typically well-established after age 25 and are more characteristic of average alcohol exposure (Substance Abuse and Mental Health Services Administration, 2019; White & Jackson, 2004). We compared the acceptability of using the Skyn and SCRAM and evaluated perceived burden associated with using the Skyn. We assessed the proportion of alcohol use data available from the Skyn and we assessed validity based on the 1) correspondence in detecting drinking days between the Skyn versus self-report or the SCRAM, and 2) correspondence between the Skyn's ability to detect day-level features of alcohol use in the field compared to more established measures, including self-report and the SCRAM. Day-level alcohol use features were compared due to the small number of days with more than one SCRAM-defined drinking episode, differential time lags between Skyn and SCRAM drinking detection, the lag between drinking initiation and TAC detection (Fairbairn & Kang, 2019; Piasecki, 2019; Wang et al., 2019), and due to self-reports being measured daily. This study extends the existing literature by examining the use of the Skyn in naturalistic settings across the longest time frame to date (28 days), and replicating real-world conditions of Skyn use, such that participants were permitted to indirectly expose the Skyn to environmental alcohol (e.g., via use of hand sanitizer) and to use the Skyn app on their personal smartphones to sync device data, emulating the typical approach used for other app-connected device-based measures of health behaviors (e.g., physical activity devices). Additionally, day-level features of alcohol use (e.g., peak TAC, drinking duration) detected via the Skyn (worn all 28 days) are compared to self-report (collected all 28 days) and to the SCRAM (worn for the first 14 days).

## Materials and methods

### *Participants, recruitment, and screening*

Participants were recruited throughout the state of Pennsylvania via email (e.g., University listservs), direct mail postcards, and word of mouth, and completed a screening survey prior to enrollment. Recruitment materials targeted individuals who consumed alcohol on at least two days per week. To be eligible, participants needed to 1) speak and read English fluently, 2) be 26–40 years of age, 3) live in Pennsylvania, 4) own and use an Apple phone, and 5) have consumed alcohol on at least two days per week over the prior month. Participants were excluded if they 1) were pregnant or planning on becoming pregnant, 2) had ever been diagnosed with cancer, and 3) scored greater than 8[1] on the Alcohol Use Disorder Identification Test (AUDIT) (Babor, Higgins-Biddle, Saunders, & Monteiro, 2001), as we were interested in examining device performance among light and moderate drinkers. Participants ($n = 20$) completed the screening survey, and five did not qualify due to their AUDIT scores ($n = 3$), age ($n = 1$), or lack of an Apple device ($n = 1$). Of the 15 qualified participants, three did not respond to researchers after completing the screening survey. The remaining 12 qualified participants consented to participate; however, one withdrew after three days due to experiencing physical pain while wearing the SCRAM (the device rubbed against their ankle skin) despite several attempts to reduce/eliminate pain based on troubleshooting advice from research personnel, and their data were removed from the study, resulting in a final study sample of 11 participants.

### *Procedure*

The study lasted 30 days and included several phases: 1) baseline visit, 2) device-training visit, 3) 28-day field protocol with midterm survey collection, and 4) follow-up survey and interview. Data were collected between January 7, 2021, and May 28, 2021, and, due to the COVID-19 pandemic, all visits were conducted remotely. Participants were compensated up to $120. All procedures were approved by the university institutional review board (Study #00026481).

Participants completed their baseline visit remotely via Zoom, at which time they provided consent and completed a 30-min survey

[1] The screening score of greater than 8 (i.e., scores of 9 or greater) was used by mistake. The guidelines for the AUDIT indicate that participants should be screened out based on scores of 8 or greater. Including participants with AUDIT scores of 8 may have resulted in the sample including some heavy/problem drinkers in addition to the more moderate/social drinkers.

assessing demographics, drinking motives (Cooper, 1994), and substance use and physical activity behaviors.

Following the baseline visit, devices were mailed to participants with a paper wear log. Upon receipt of devices, participants completed a 1-h training visit via Zoom regarding how to wear the devices, use the apps (i.e., Skyn and Personal Analytics Companion or PACO app) (Evans, 2016, 2021), and complete the wear log. PACO is an open-source platform developed for researchers to collect survey-based behavioral health data through a smartphone application (downloaded to the participant's device) or through the PACO website (Evans, 2016, 2021).

Immediately following the training visit, participants began the 28-day field protocol, during which they wore the SCRAM, the Skyn, and a physical activity monitor (activPAL), and completed daily EMA surveys and event-initiated drinking surveys on their smartphones. On the morning of day 15, participants removed the SCRAM and completed the midterm survey. Throughout the 28-day field protocol, research personnel monitored the Skyn data to ensure participants were syncing their data. Participants were contacted after two consecutive days of missing Skyn data ($n = 3$) to increase compliance and data quality. Research personnel provided troubleshooting help throughout the study period for participants as needed. On the morning of day 29, participants removed the Skyn and completed the follow-up survey. Within 1 week after the 28-day field protocol, participants completed a 10-min follow-up interview. Data from the activPAL were not used in this manuscript.

***Skyn protocol.*** Participants wore the Skyn T10 model (rented from the manufacturer by the research team) on their non-dominant wrist continuously for the 28-day field protocol during waking and sleeping hours, only removing it to shower, bathe, swim, or charge the device, which required 60–90 min. Participants downloaded the app to their smartphones. Skyn data were synced via Bluetooth to participants' smartphones and uploaded to the company's cloud-based server. Alcohol use data were available to participants via the app, but the data were not pushed to participants and there was no record of whether participants sought behavioral feedback. Data synced automatically if the app was open, the device was connected, and there was internet access. If the app was closed or the device was not connected, participants had to manually connect and sync the data. Participants were reminded daily to sync their devices and upload data, because failure to sync resulted in data loss, as the Skyn can only hold 72 h of data. The Skyn holds its charge for up to 72 h, though not all batteries lasted that long, and data loss occurred if the battery was depleted. After 28 days, participants mailed the Skyn back to the lab using a pre-paid, pre-addressed envelope provided by the researchers.

***SCRAM protocol.*** Participants wore the SCRAM on their ankle continuously for the first 14 days of the field protocol during waking and sleeping hours, only removing it to bathe, swim, or participate in vigorous activity or contact sports. The SCRAM was only worn for 14 days to reduce participant burden given lower acceptability of the SCRAM compared to the Skyn (Ash et al., 2022). After 14 days, participants mailed the SCRAM back to the lab and data were uploaded to the SCRAMNet server, which houses TAC data, records TAC “positives”, and tracks device wear compliance through infrared voltage and skin temperature readings.

***Wear log protocol.*** Participants completed a daily wear log in which they reported whether they removed the Skyn or SCRAM devices. Participants reported the day of the week, the date, and the time a given device was removed and replaced, with the option to report multiple device removals throughout the day.

***EMA protocol.*** Participants used the PACO app to complete the EMA protocol, which consisted of two survey types. The first was the scheduled EMA survey, which prompted participants to complete a morning survey at 9:00 AM. Participants could respond immediately upon prompting or could self-initiate surveys. The morning survey required 1–2 min to complete and asked about the previous day's drinking and substance use behaviors and drinking-related influences (e.g., seeing others drink, having access to alcohol). At the end of the morning survey, participants were reminded to sync their Skyn data and to charge the Skyn device. The second EMA type was the event-initiated drinking surveys based on Piasecki's (2019) episode bracketing approach, in which participants initiated EMA surveys when they started consuming their first drink of the day and when they stopped consuming their last drink of the day.

On the morning of day 29, participants completed the follow-up survey. Within one week of completing the 28-day field protocol, participants completed a 10-min, one-on-one semi-structured interview via Zoom, and provided information about their experiences using the Skyn device and app.

*Measures*

***Device acceptability.*** During midterm and follow-up, participants reported acceptability of using the SCRAM (midterm only) and Skyn devices by rating their agreement with 10 items from the Measuring User Acceptance of Wearable Symbiotic Devices survey (Spagnolli, Guardigli, Orso, Varotto, & Gamberini, 2014). We used the perceived comfort (five items, e.g. “I think the SCRAM/Skyn was comfortable”), facilitating conditions (two items, e.g., “The SCRAM/Skyn limits the way in which I like to perform my daily activities”), effort expectancy (two items, e.g., “It seems easy to learn how to use the SCRAM/Skyn”), and perceived privacy (one item, i.e., “I think the SCRAM/Skyn threatens my privacy”) subscales. Agreement ratings were on a 5-point Likert scale ranging from Strongly Disagree (1) to Strongly Agree (5). We calculated average acceptability for each subscale and the entire survey, with higher scores indicating greater acceptability ($\alpha = 0.84$–$0.89$). Participant interview responses about likes/dislikes related to using the Skyn and its app provided additional context about Skyn acceptability.

***User burden.*** During the follow-up survey, participants reported burden associated with using the Skyn by rating either the frequency or degree of burden based on 18 items from the User Burden Scale (Suh, Shahriaree, Hekler, & Kientz, 2016), which refers to burden associated with using both the Skyn device and app (i.e., “BACtrack tools”). We used the difficulty of use (four items, e.g., “The BACtrack tools were hard to learn”), physical burden (three items, e.g., “The BACtrack tools made me feel physical pain”), time and social burden (four items, e.g., “I spent too much time using the BACtrack tools”), mental and emotional burden (four items, e.g., “The BACtrack tools required me to remember too much information”), and privacy (three items, e.g., “The BACtrack tools' policies about privacy were not trustworthy”) subscales. Ratings were on a 5-point Likert scale, with frequency responses ranging from Never (1) to All of the Time (5) and degree of burden responses ranging from Not at All (1) to Extremely (5). We calculated average levels of burden for each subscale and for the entire survey, with higher scores indicating greater user burden. Average scores were interpreted using Suh et al.‘s guidelines (2016) ($\alpha = 0.82$), such that subscales received grades of A (top 15% of scores; 86–100%), B (next 30% of scores; 56–85%), C (next 40% of scores; 16–55%), D (next 10% of scores; 6–15%), or F (bottom 5% of scores; 0–5%) based on the following average scores, with ranges varying for each subscale (for example, an average score of 2 is a “C” on the time and social subscale and a “D” on the difficulty of use subscale): Average scores of 0 received an “A”, average scores ranging from 0 to 0.5 received a “B”, average scores ranging from 0.33 to 2.33 received a “C”, average scores ranging from 1 to 3.33 received a "D", and average scores ranging from 1.67 to 4 received an "F" (Suh et al., 2016). Participant interview responses about difficulties using the Skyn and Skyn app provided additional context about user burden.

***Data Quality and Validity.*** Data quality was assessed based on the proportion of alcohol use data available via self-report (i.e., number of daily surveys) and the Skyn and SCRAM devices (i.e., relative to the number of days/hours useable data were collected). Validity was assessed based on: 1) correspondence in detecting drinking days between the Skyn versus self-report or the SCRAM, and 2) correspondence between the Skyn's ability to detect day-level features of alcohol use in the field compared to more established measures, including self-report and the SCRAM, though neither is a gold standard for field studies of drinking (Piasecki, 2019).

***Self-reported day-level alcohol use features.*** During the morning EMA survey, participants reported the total number of standard drinks of alcohol they consumed "yesterday" (*drinking intensity* [*c*]), with responses ranging from 0 to 15 (or more, coded at 15). A standard serving was defined as 12 ounces (oz.) of 5% beer, 8–9 oz. of 7% beer, 4–5 oz. of wine, a 1.5 oz. shot of liquor (straight or in a mixed drink), or 12 oz. of hard seltzer; the survey was accompanied by a picture (National Institute on Alcohol Abuse and Alcoholism, 2022). Daily self-reports were selected because they show stronger correlations with SCRAM features than other methods, such as timeline follow-back (Russell et al., 2022). Participants also reported the times when they started and stopped drinking "yesterday". During the event-initiated surveys, participants reported when they started or stopped drinking, and survey time stamps were used to identify drinking start/stop times. If participants reported having consumed one or more drinks on the morning survey it was classified as a drinking day. If participants reported not drinking "yesterday" (e.g., Day 2) on the morning survey but they did initiate a start and/or stop drinking survey on the same day referenced by the morning survey (e.g., Day 2), then it was considered a drinking day ($n = 6$). *Drinking duration* was calculated based on the time elapsed between when participants reported starting and stopping drinking. There were two separate drinking duration variables, one based on the morning EMA ($t_{\text{morn}}$) and the second based on the event-initiated surveys ($t_{\text{event}}$), both of which were used in subsequent analyses. Estimated Blood Alcohol Concentration (eBAC) values were calculated using the following formula: $\text{eBAC} = [(c/2) \times (\text{GC}/w)] - (\beta_{60} \times t)$ (Matthews & Miller, 1979). Here eBAC is expressed in g/dL, $c$ is the number of standard drinks reported, GC is a gender constant (9.0 for females and 7.5 for males), $w$ is weight in pounds, $\beta_{60}$ is the metabolism rate of alcohol per hour (0.017 g/dL), and $t$ is the number of hours spent drinking ($t_{\text{morn}}$ or $t_{\text{event}}$). There were two separate eBAC variables, one based on the morning EMA drinking time ($\text{eBAC}_{\text{morn}}$) and the second based on the event-initiated surveys ($\text{eBAC}_{\text{event}}$). Negative values for $\text{eBAC}_{\text{morn}}$ and $\text{eBAC}_{\text{event}}$ were replaced with 0. *Drinking days, drinking intensity, drinking duration,* and *eBAC* from the morning EMA surveys were temporally aligned with event-initiated drinking surveys and Skyn and SCRAM sensor data.

***Segmenting Skyn and SCRAM TAC data into "social days".*** Given that drinking often extends past midnight, we assumed that participants' EMA morning reports of "yesterday's" drinking included the hours between midnight and the morning report. Therefore, Skyn and SCRAM TAC data were segmented into "social days" (Russell et al., 2022), using 9:00 AM as the boundary because it was the time of the morning EMA. If participants had at least one drinking episode on a social day, it was classified as a drinking day. If a drinking episode spanned multiple days, day-level features were calculated separately for each day, and if a day contained multiple episodes, day-level features were calculated using all data for that social day.

***Skyn day-level alcohol use features.*** The Skyn uses fuel-cell technology to assess TAC every 20 s in μg/L air and captures skin temperature (°C) to help identify device wear compliance and motion (*g*'s). Skyn TAC data do not have any zero values or any 'true-zero' in the sense that there is no value output by the Skyn that can, at face value, be interpreted as representing a complete lack of alcohol (i.e., 0 alcohol) being excreted, Skyn data also include negative values, baseline values vary between individuals, and there are no established guidelines for identifying drinking episodes (Fairbairn & Kang, 2019; Rosenberg et al., 2021; Wang et al., 2019). Therefore, we identified episodes by iteratively examining the data using rules that were informed by consultations with the Skyn manufacturer (Personal Communication – Neely, 2021) and by Roache et al.'s (2019) rules for cleaning SCRAM data. Although Roache et al.'s (2019) rules could not be directly applied to the Skyn data due to differences between the SCRAM and Skyn TAC scales, the rules were used to inform the parameters investigated in the Skyn data to identify drinking episodes, including steep point-to-point rises or falls in TAC slopes (Roache et al., 2019; Rule 3), removing extremely short TAC events (Roache et al., 2019; Rule 5), and removing short TAC events with steep point-to-point rises in TAC slopes (Roache et al., 2019; Rule 6). Table 1 defines the steps used for processing Skyn data and identifying episodes. Briefly, TAC readings taken when the temperature was <29 °C were considered non-wear and were recoded to missing based on the manufacturer's recommendation. Fifteen-minute rolling averages of minute-level Skyn data were used to smooth the TAC data (Rosenberg et al., 2021; Wang et al., 2019) and any observation with a 15-min rolling TAC average ≥ 15 μg/L air was identified as a potential alcohol use episode based on the manufacturer's recommendation ($n = 414$). Potential Skyn alcohol use episodes were removed if they had any of the following characteristics: 1) ≤15 min (similar to Roache et al.'s [2019]; Rules 5 and 6); 2) ≤60 min and peak TAC ≥400 μg/L air; 3) Rise rate ≥2500 (similar to Roache et al.'s [2019] Rule 3); 4) ≤60 min and rise rate ≥1000; or 5) Rise rate = 0. All Skyn TAC values that were not part of an episode were replaced with 0 μg/L air as those TAC values were considered indicative of non-drinking alcohol exposure (i.e., environmental contamination). This introduced new 0 μg/L air values into the Skyn data because, as noted above, Skyn TAC data do not include any zero values or any 'true-zero' values. A total of 260 episodes were removed, resulting in 154 episodes. Of these 154 episodes, 53 occurred during days 1–14 from participants who also had SCRAM data. Five characteristics were extracted from Skyn drinking episodes to establish day-level features of alcohol use based on the social day (9:00 AM to 9:00 AM) (Russell et al., 2022), including *drinking intensity* (peak TAC and area under the curve [AUC]), *drinking duration* (hours with TAC >0 μg/L air), *rise rate* (average rate of all ascending point-to-point TAC values), and *fall rate* (average rate of all descending point-to-point TAC values).

***SCRAM day-level alcohol use features.*** The SCRAM uses self-generated air flow to capture transdermal ethanol evaporation every 30 min in g/1470 L of air, and TAC is determined using fuel-cell technology (Fairbairn & Kang, 2019; van Egmond et al., 2020). For the SCRAM, if a participant self-reported removing the sensor, all TAC values during the time frame it was removed were replaced with missing values. SCRAM devices failed for four participants, resulting in seven participants with SCRAM data. Based on Roache et al.'s (2019) validated guidelines, those seven participants recorded 43 SCRAM drinking episodes across days 1–14. Five characteristics were extracted from SCRAM drinking episodes to establish day-level features of alcohol use: *drinking intensity* (peak TAC and AUC), *drinking duration* (hours with TAC >0 g/1470 L air), *rise rate*, and *fall rate*.

**Table 1**
Steps for Processing Skyn Data to Identify Drinking Episodes

| Step Number | Step name | Description |
|---|---|---|
| 1 | Recoding negative TAC values to zero | All Skyn TAC values that were negative (less than zero) were replaced with zero. This step was recommended by BACtrack Skyn personnel due to the fact that there are no zero values in Skyn data[a] and negative TAC values are physiologically implausible. |
| 2 | Recoding TAC values with temperature <29 °C to missing[b] | All Skyn TAC values that occurred when the temperature reading was <29 °C were considered missing data and were recoded as missing. This step was recommended by BACtrack Skyn personnel because the Skyn monitor measures body temperature (°Celsius) as a means to detect non-wear time. Personnel suggested that any temperature reading <29 °C was indicative of participant non-wear. |
| 3 | Calculating 15-minute rolling averages for TAC[c] | 15-minute rolling averages for TAC (μg/L air) were calculated that included the 7 minutes of data prior to and after a given minute of data. This step was recommended by BACtrack Skyn personnel to smooth out TAC data. |
| 4 | Identifying minutes with rolling TAC ≥15 μg/L air | Any row of data with a 15-minute rolling average for TAC ≥15 μg/L air was identified as potentially representing the device being exposed to alcohol, either via the participant's skin, which would be indicative of a participant consuming alcohol, or via environmental contamination (i.e., hand sanitizer, cologne, alcohol-based products). BACtrack Skyn personnel identified 15 μg/L air as a cut-off TAC value (from the rolling average) that could potentially represent true alcohol exposure. This step was needed due to the lack of zero values in Skyn data, such that a person could have a zero for blood alcohol concentration but still show positive TAC readings from the Skyn device. |
| 5 | Bracket individual alcohol exposure episodes[d] | Individual alcohol exposure episodes were bracketed separately from one another, such that consecutive strings of data with rolling TAC averages ≥15 μg/L air were bracketed into an 'alcohol exposure episode' to indicate that the device was exposed to non-negligible amounts of alcohol. |
| 6 | Define characteristics of alcohol exposure episodes | Characteristics of the alcohol exposure episodes were defined, including the peak TAC, AUC, episode duration, average ascending TAC slope, and average descending TAC slope. This step was used as a starting point for identifying episode characteristics that were indicative of actual alcohol use (i.e., actual drinking episodes) versus characteristics that were indicative of non-drinking alcohol exposure (i.e., environmental contamination). |
| 7 | Calculating characteristics for Skyn alcohol exposure episodes | Descriptive statistics for each of the characteristics defined in Step 6 were calculated for each Skyn alcohol exposure episode. This step was used to identify minimum, maximum, and mean values of episode characteristics (e.g., peak TAC, duration) to indicate values that were likely representative of actual alcohol use versus non-drinking alcohol exposure, such as environmental contamination. |
| 8 | Removing Skyn alcohol exposure episodes ≤15 minutes in duration[e] | Skyn alcohol exposure episodes ≤15 minutes in duration were removed. Episodes ≤15 minutes were considered to be indicative of non-drinking alcohol exposure based on descriptive characteristics calculated during Step 7 and based on the physiological improbability that an individual would metabolize all of the alcohol they consumed rapidly enough to make it undetectable by the Skyn (i.e., rolling average <15 μg/L air) in ≤15 minutes. This step is similar to Roache et al.'s Rules 5 and 6 (2019). |
| 9 | Removing Skyn alcohol exposure episodes ≤60 minutes in duration and with a peak TAC ≥400 μg/L air[f] | Skyn alcohol exposure episodes ≤60 minutes in duration and with a peak TAC ≥ 400 μg/L air were removed. Episodes with these characteristics were considered to be indicative of non-drinking alcohol exposure based on Step 7, as the combination of a high peak TAC with a short episode duration reflected the Skyn being exposed to non-drinking-related alcohol, like hand sanitizer, resulting in a higher peak TAC and shorter duration than would be likely if an individual consumed enough alcohol to reach such a high peak TAC. If an individual consumed that much alcohol, it would likely take longer than 60 minutes to metabolize all of the alcohol rapidly enough to make it undetectable by the Skyn in ≤60 minutes. |
| 10 | Removing Skyn alcohol exposure episodes with a rise rate ≥2500[g] | Skyn alcohol exposure episodes with a rise rate ≥2500 were removed. Episodes with this characteristic were considered to be indicative of non-drinking alcohol exposure based on Step 7, as such large rise rates were considered implausible representations of true drinking, as the TAC increased too quickly over too short of time period for true drinking, and likely only occurred due to environmental contamination, such as exposure to high concentrations of alcohol (i.e., hand sanitizer) at the Skyn sensor. This step is similar to Roache et al.'s Rule 3 for processing SCRAM data (2019). |
| 11 | Removing Skyn alcohol exposure episodes ≤60 minutes in duration and with a rise rate ≥1000[h] | Skyn alcohol exposure episodes ≤60 minutes in duration and with a rise rate ≥1000 were removed. Episodes with these characteristics were considered to be indicative of non-drinking alcohol exposure based on Step 7, as such short episodes combined with such high rise rates were considered implausible representations of true drinking, as the TAC increased too quickly over too short a time period and the episode was too short in duration for an individual to metabolize the amount of alcohol that could account for such a rapid increase in TAC over that short time period of ≤60 minutes. |
| 12 | Removing Skyn alcohol exposure episodes with a rise rate = 0[i,j] | Skyn alcohol exposure episodes with a rise rate = 0 were removed. Episodes with this characteristic were considered to be indicative of non-drinking alcohol exposure based on Step 7, as actual alcohol consumption should be reflected by increases (ascending limb/slope) and decreases (descending limb/slope) in TAC paralleling changes in blood alcohol concentration after alcohol is consumed, metabolized, and eliminated from the body, and a rise rate = 0 reflects no increase in TAC across the episode. |
| 13 | Replacing non-zero TAC values that were not part of an alcohol exposure episode with 0 μg/L air | All non-zero TAC values that were not part of an alcohol exposure episode, including negligible non-zero TAC values, as well as TAC values that were originally part of an alcohol exposure episode, but that were later removed based on Steps 8−12, were replaced with 0 μg/L air, as the original non-zero values were deemed to not represent drinking-related alcohol exposure. They were not replaced with missing values because the data were not missing *per se*, given that the Skyn had TAC measures that were taken when the device was being worn (the temperature was ≥29 °C). These values were |

**Table 1** (*continued* )

| Step Number | Step name | Description |
|---|---|---|
| | | replaced with 0 so that we could calculate day-level characteristics of alcohol use from the Skyn TAC data with non-drinking related TAC values removed from the day-level data. |

[a] Raw Skyn data did not include any values equal to zero.
[b] Descriptive statistics of TAC readings that occurred when the temperature was <29 °C revealed that all readings were negligible, non-zero values, increasing confidence that the cut-off of 29 °C was appropriate and likely did not inadvertently result in removing drinking episodes from the data.
[c] Smoothing of TAC data through the use of rolling or moving averages is an approach that has been used by previous researchers (Rosenberg et al., 2021; Wang et al., 2019).
[d] This first round of bracketing alcohol exposure episodes resulted in identifying 414 episodes across the 11 participants' 308 study days, with a range of 1–63 episodes per participant.
[e] 108 episodes violated the rule of being ≤15 minutes in duration.
[f] 70 episodes violated the rule of being ≤60 minutes in duration and having a peak TAC ≥ 400 μg/L air.
[g] 118 episodes violated the rule of having a rise rate ≥2500.
[h] 165 episodes violated the rule of being ≤60 minutes in duration and having a rise rate ≥1000.
[i] 21 episodes violated the rule of having a rise rate = 0.
[j] After removing a total of 260 episodes based on rules outlined in steps 10–14, there were 154 episodes across 10 participants (one participant had zero episodes).

### *Statistical analyses*

Paired samples *t* tests were used to compare Skyn versus SCRAM acceptability at midterm, and to compare Skyn acceptability at midterm versus follow-up. Cohen's *d* was used to estimate effect sizes. Detection rates of drinking days were compared between the Skyn versus self-report or the SCRAM, and the SCRAM versus self-report. Rates of true positives/negatives, false positives/negatives, sensitivity, specificity, and positive and negative predictive value were calculated by comparing Skyn versus self-report and SCRAM versus self-report (see Table 4 for formulas). When comparing Skyn versus SCRAM, we examined agreement and disagreement between devices in detecting days as drinking versus non-drinking days and the likelihood of days being detected as drinking days by both devices or non-drinking days by both devices.

Repeated-measures correlations with bootstrapping were calculated using the *rmcorr* package in R (Bakdash & Marusich, 2017) to examine within-person associations between day-level features of alcohol use based on Skyn and SCRAM data and self-reported EMA surveys. Repeated-measures correlations account for the non-independence of observations through an atypical application of ANCOVA with a fixed effect of ID, thus removing between-person variance from the predictor and estimating the within-person association between the predictor and outcome (i.e., day-level features of alcohol use measured via the Skyn versus self-report or the SCRAM) (Allison, 2019; Bakdash & Marusich, 2017). This was an appropriate method given the focus on testing within-person comparisons of day-level features of alcohol use. Day-level features examined included drinking intensity, drinking duration, eBAC, peak TAC, AUC, rise rate, and fall rate. We used R version 4.0.1 for all analyses. Statistical significance was set at $p < .05$.

## Results

Table 2 includes participant demographics and descriptive statistics for drinking behaviors across the 28 days. The sample (N = 11) was 55% female and 73% non-Hispanic white with a mean age of 31.2 ± 3.3 years. The majority of participants (64%) worked full time and 36% were graduate students. Participants self-reported drinking alcohol an average of 15.4 ± 7.4 days (range = 5–27 days), consuming a mean of 39.9 ± 31.2 drinks across the 28 days (range: 6 to 107 drinks), and a mean of 2.8 ± 2.0 drinks per drinking day (range = 1 to 10 drinks/drinking day). Using CDC definitions, participants were categorized as light (≤3 drinks/week), moderate (>3 to 14 and > 3 to 7 drinks/week for men and women, respectively), or heavy (≥15 and ≥ 8 drinks/week for men and women, respectively) drinkers, such that 18.2% (n = 2) were light drinkers, 54.5% (n = 6) were moderate drinkers, and 27.3% (n = 3) were heavy drinkers (Centers for Disease Control and Prevention, 2019). Based on self-report, 63.6% (n = 7) of participants engaged in binge drinking (4+/5+ drinks/day for women/men) on at least one day across the 28-day study. Across all participants, the mean number of binge drinking days was 2.7 ± 3.7 days (range = 0–12 days), with those who binge-drank engaging in binge drinking on a mean of 4.3 ± 3.8 days (range = 1–12 days).

### *Acceptability and user burden*

One participant dropped the study after three days due to discomfort wearing the SCRAM, which implies low device acceptability. However, they dropped prior to completing the midterm surveys for SCRAM and Skyn acceptability, and the subsequent comparisons are based on the 11 participants who provided acceptability data. At midterm (study day 15), Skyn acceptability was 4.4 ± 0.6 (SD) and SCRAM acceptability was 2.5 ± 0.8 out of 5. Skyn acceptability ratings were significantly higher than SCRAM

**Table 2**
Participant demographics and descriptive statistics for drinking behaviors

| Demographics | N=11 |
|---|---|
| Age in years (Mean ± SD) | 31.2 ± 3.0 |
| Sex (n (%)) | |
| Male | 5 (45.5%) |
| Female | 6 (54.5%) |
| Race, non-hispanic ethnicity [n (%)][a] | |
| White | 8 (72.7%) |
| Asian | 1 (9.1%) |
| Black | 2 (18.2%) |
| Work status (n (%)) | |
| Full-time | 7 (63.6%) |
| Undergraduate/Graduate student – working | 1 (9.1%) |
| Undergraduate/Graduate student – not working | 3 (27.3%) |
| Drinking descriptive statistics | **N=11** |
| Days drink per week (Mean ± SD) | 3.8 ± 1.8 |
| Drinks per week (Mean ± SD) | 10.0 ± 7.5 |
| Binge drinkers [n (%)][b] | 7 (63.6%) |
| Days per month – All participants (Mean ± SD) | 2.7 ± 3.7 |
| Days per month – Binge drinkers (Mean ± SD)[c] | 4.3 ± 3.8 |
| Drinker type [n (%)][d] | |
| Light | 2 (18.2%) |
| Moderate | 6 (54.5%) |
| Heavy | 3 (27.3%) |

Notes: SD = standard deviation.
[a] All participants were non-Hispanic.
[b] Binge drinking was defined based on consuming 4+ or 5+ drinks per day for women or men, respectively.
[c] Mean days per week with binge drinking among people who engaged in binge drinking (n=7).
[d] Drinkers were categorized using CDC definitions. Light = ≤3 drinks/week, moderate = >3 to 14 and >3 to 7 drinks/week for men and women, respectively, and heavy = ≥15 and ≥8 drinks/week for men and women, respectively (Centers for Disease Control, 2019).

ratings overall [$t$ (10) = −6.79, $p$ < .001, $d$ = −2.74] and for the perceived comfort [$t$ (10) = −6.66, $p$ < .001, $d$ = −2.77], facilitating conditions [$t$ (10) = −8.32, $p$ < .001, $d$ = −3.00], and effort expectancy [$t$ (10) = −4.34, $p$ < .01, $d$ = −1.33) subscales. There were no differences in Skyn versus SCRAM acceptability ratings for the perceived privacy subscale [$t$ (10) = −0.99, $p$ = .37, $d$ = −0.41]. As shown in Supplemental Table 1, several themes emerged related to Skyn acceptability in the follow-up interviews, including participants liking the feel and look of the Skyn versus the discomfort and social stigma of the SCRAM and the low effort for using the Skyn app.

At follow-up, Skyn acceptability was 4.5 ± 0.7, which was the same as at midterm [$t$ (10) = −0.78, $p$ = .46, $d$ = −0.16]. Mean user burden for the Skyn was 0.2 ± 0.2, and it received A's (top 15% of scores, ranging from 86 to 100%) for the time and social (0.2 ± 0.2), mental and emotional (0.2 ± 0.3), and privacy (0.1 ± 0.2) subscales, received a B (next 30% of scores, ranging from 56 to 85%) for the difficulty of use (0.3 ± 0.4) subscale, and received a C (next 40% of scores, ranging from 16 to 55%) for the physical burden (0.4 ± 0.6) subscale (Suh et al., 2016). The 'B's/C's' for difficulty of use and physical burden were mirrored by interview responses, with themes related to issues the participants experienced due to skin discomfort, the fragility and limited durability of the Skyn, and difficulties related to battery life (reported by all 11 participants), charging, and syncing the device (see Supplemental Table 1 for details).

### *Data Quality and Validity*

***Alcohol use data quality and availability.*** One of 10 Skyn devices (10%) failed to hold a charge and was replaced halfway through the field protocol. Participants reported wearing the Skyn for 89% of the 28 days. Overall, data were available for a total of 5617 h (74.2%) out of 7566 h. Across the 11 participants, data availability ranged from 180.6 to 589.2 h/person (27.1%–85.4% of the 28-day field protocol). However, 1563 h (20.7%) of data were lost due to participant failure to sync the data, charge the device, or incomplete syncing (range: 61.0–440.8 h/person; 8.8%–66.2%). Another 386.2 h (5.1%) of data were lost due to device removal (based on Skyn temperature readings <29 °C) (range: 7.4–55.3 h/person; 1.1%–8.0%). In total, 1949 h (25.8%) of data were missing across all 11 participants across the 28-day field protocol (range: 100.8–485.4 h/person; 14.6%–72.9%). The proportion of missing data did not vary by day of the week (i.e., weekdays versus weekends, social weekdays versus social weekends) or any sociodemographic variables ($p$ > .05).

Of the seven SCRAM devices used for the study, three (43%) failed to record any TAC data despite recording temperature and infrared voltage and despite functioning correctly and recording TAC data in previous studies (Russell et al., 2022; West, Bomysoad, Russell, & Conroy, 2022). The three devices that failed were worn by four of the 11 participants (36%), resulting in a sample of seven participants with SCRAM data. Among those seven participants, SCRAM data were available for a total of 2363 h (97.8%) out of 2416 h across days 0–14. SCRAM data availability ranged from 320.5 to 347.0 h/person (93.0%–100.0% of days 0–14). Four participants self-reported removing the devices briefly (~60–120 min) to participate in exercise, such that a total of 53.5 h (2.2%) of data were lost due to device removal (range: 0.0–24.0 h/person; 0.0%–7.0%). Missingness did not vary by day of the week (i.e., weekdays versus weekends, social weekdays versus social weekends) or any sociodemographic variables ($p$ > .05).

Participants completed 97% of the assigned EMA morning reports. Participants initiated start and stop drinking reports on 88% and 74% of morning-reported drinking days, respectively.

***Detecting drinking days.*** Table 3 provides descriptive statistics for the type of drinking day and day-level features of alcohol use by measurement type. Across the 28-day field protocol, we identified 107 drinking days and 185 non-drinking days with the Skyn and participants self-reporting 169 drinking days and 142 non-drinking days. Across the first 14 days, and among participants whose SCRAM devices functioned, we identified 38 drinking days and 57 non-drinking days with the Skyn, 40 drinking days and 65 non-drinking days with the SCRAM, and participants self-reported 56 drinking days and 47 non-drinking days.

Table 4 provides detection rates of drinking and non-drinking days, and rates for true positives/negatives, false positives/negatives, sensitivity, specificity, and positive/negative predictive values for Skyn and SCRAM versus self-report. When compared to self-report, the Skyn and SCRAM had similar false positive (13.6% and 10.6%, respectively) and false negative rates (44.5% and 37.5%, respectively). Approximately half of the false negatives for the Skyn and SCRAM (44% and 57%, respectively) occurred on days when participants self-reported consuming one alcoholic beverage. For the false negatives occurring on days when participants consumed more than one alcoholic beverage, participants self-reported 3.5 ± 1.8 drinks (range: 2–8) for days with Skyn false negatives and self-reported 2.1 ± 0.3 drinks (range 2–3) for days with SCRAM false negatives. When compared to self-report, the Skyn and SCRAM had similar levels of sensitivity (i.e., the proportion of self-reported drinking days that the device detected) and positive predictive value (i.e., the likelihood that a device-detected drinking day was also a self-reported drinking day). Specifically, Skyn and SCRAM sensitivity were 55.5% and 62.5%, respectively, and positive predictive values were 85.0% and 87.5%, respectively.

When comparing the Skyn to the SCRAM there was some disagreement in detecting drinking days. Specifically, 20.7% of drinking days detected by the Skyn were not detected by the SCRAM, and 29.7% of drinking days detected by SCRAM were not detected by the Skyn. However, the Skyn and SCRAM also showed agreement in that 70.3% of drinking days detected by the Skyn were detected by the SCRAM, and 79.3% of non-drinking days detected by the Skyn were detected by the SCRAM. The likelihood of a Skyn-detected drinking day also being a SCRAM-detected drinking day was 68.4%. The likelihood of a Skyn-detected non-drinking day also being a SCRAM-detected non-drinking day was 80.7%.

***Correlations between day-level features of alcohol use.*** Table 5 provides repeated-measures correlations of day-level features of alcohol use (e.g., drinking intensity, duration, estimated BAC) for all three measurement types. Skyn and SCRAM day-level alcohol use features showed significant within-person correlations with all self-reported alcohol use features. However, the correlations between Skyn-based measures of drinking intensity (e.g., peak TAC, AUC) and self-reported measures of drinking intensity (i.e., drinks consumed) were smaller than the correlations between SCRAM-based measures of drinking intensity (e.g., peak TAC, AUC) and self-reported measures of drinking intensity. Specifically, the correlations between Skyn peak TAC and AUC with self-reported number of drinks were 0.35 and 0.52, respectively, whereas the correlations between SCRAM peak TAC and AUC with self-reported number of drinks were 0.78 and 0.79, respectively. Similarly, the correlations between Skyn peak TAC and AUC with morning report-based eBAC were 0.30 and 0.48, respectively, whereas the correlations between SCRAM peak TAC and AUC with morning report-based eBAC were 0.61 and 0.59, respectively. In contrast, the Skyn correlations with self-reported drinking duration were marginally larger than the SCRAM correlations with self-reported drinking duration. Specifically, the correlations between Skyn drinking duration with morning EMA and event-contingent EMA drinking duration were 0.36 and 0.37, respectively, whereas the correlations

**Table 3**
Day-level alcohol use characteristics by measurement source[a]

| Study days 1–14 | Skyn | SCRAM | Self-report |
|---|---|---|---|
| Type of day | | | |
| Drinking day [n (%)] | 38 (36.2%) | 40 (24.2%) | 56 (53.3%) |
| Non-drinking day [n (%)] | 57 (54.3%) | 65 (39.4%) | 47 (44.8%) |
| Missing [n (%)] | 10 (9.5%) | 60 (36.4%) | 2 (2.4%) |
| Drinking intensity | | | |
| Number of drinks (*c*) (Mean ± SD)[b] | – | – | 1.45 ± 1.05 |
| Peak TAC (Mean ± SD)[c,d] | 53.66 ± 39.58 | 0.018 ± 0.012 | – |
| Area under the curve (Mean ± SD)[e,f] | 82.64 ± 76.02 | 0.08 ± 0.06 | – |
| **Drinking Duration (hours/day)** | | | |
| Morning EMA ($t_{morn}$) (Mean ± SD)[g] | – | – | 1.59 ± 1.37 |
| Event contingent EMA ($t_{event}$) (Mean ± SD)[h] | – | – | 2.14 ± 1.78 |
| Time with TAC > 0 (Mean ± SD)[i,j] | 1.40 ± 1.20 | 2.75 ± 1.43 | – |
| **Estimated BAC (g/dL)**[k] | | | |
| Morning EMA ($eBAC_{morn}$) (Mean ± SD)[g] | | | 0.007 ± 0.005 |
| Event Contingent EMA ($eBAC_{event}$) (Mean ± SD)[h] | | | 0.005 ± 0.006 |
| Average Ascending TAC Slope (Mean ± SD)[k,l] | 129.29 ± 118.70 | 0.005 ± 0.002 | – |
| Average Descending TAC Slope (Mean ± SD)[m,n] | −72.14 ± 44.50 | −0.006 ± 0.003 | – |
| **Study days 1–28** | **Skyn** | **SCRAM** | **Self-report** |
| **Type of day** | | | |
| Drinking day [n (%)] | 107 (32.5%) | 40 (12.2%) | 169 (51.4%) |
| Non-drinking day [n (%)] | 185 (56.2%) | 65 (19.8%) | 142 (43.2%) |
| Missing [n (%)] | 37 (11.2%) | 224 (68.1%) | 18 (5.5%) |
| **Drinking intensity** | | | |
| Number of drinks (Mean ± SD)[b] | – | – | 1.41 ± 1.02 |
| Peak tAC (Mean ± SD)[c,d] | 42.85 ± 25.07 | 0.018 ± 0.012 | – |
| Area under the curve (Mean ± SD)[e,f] | 64.56 ± 53.47 | 0.08 ± 0.06 | – |
| **Drinking duration (hours/day)** | | | |
| Morning EMA (Mean ± SD)[g] | – | – | 1.88 ± 1.27 |
| Event contingent EMA (Mean ± SD)[h] | – | – | 1.95 ± 1.51 |
| Time with TAC > 0 (Mean ± SD)[i,j] | 1.08 ± 0.83 | 2.75 ± 1.43 | – |
| **Estimated BAC (g/dL)** | | | |
| Morning EMA ($eBAC_{morn}$) (Mean ± SD)[g] | – | – | 0.014 ± 0.007 |
| Event contingent EMA ($eBAC_{event}$) (Mean ± SD)[h] | – | – | 0.012 ± 0.008 |

Notes: SCRAM = Secure Continuous Remote Alcohol Monitor; TAC = Transdermal alcohol concentration; EMA = Ecological momentary assessment; SD = standard deviation
[a] Values are reported based on all days and are not limited to drinking days.
[b] 3.6% of days 1–14 were missing data. 7.3% of days 1–28 were missing data.
[c] Peak TAC for the Skyn is in μg/L air. 7.9% of days 1–14 were missing data. 11.2% of days 1–28 were missing data.
[d] Peak TAC for the SCRAM is in g/dL. 36.4% of days 1–14 were missing data across all 11 participants. 0% of days 1–14 were missing data for the 7 participants with functional SCRAM devices.
[e] 7.9% of days 1–14 were missing data. 11.2% of days 1–28 were missing data.
[f] 36.4% of days 1–14 were missing data across all 11 participants. 0% of days 1–14 were missing data for the 7 participants with functional SCRAM devices.
[g] 3.6% of days 1–14 were missing data. 7.3% of days 1–28 were missing data.
[h] 14.5% of days 1–14 were missing data. 18.8% of days 1–28 were missing data.
[i] 7.9% of days 1–14 were missing data. 11.2% of days 1–28 were missing data.
[j] 36.4% of days 1–14 were missing data across all 11 participants. 0% of days 1–14 were missing data for the 7 participants with functional SCRAM devices.
[k] 7.9% of days 1–14 were missing data.
[l] 38.2% of days 1–14 were missing data across all 11 participants. 1.8% of days 1–14 were missing data for the 7 participants with functional SCRAM devices.
[m] 7.9% of days 1–14 were missing data.
[n] 36.7% of days 1–14 were missing data across all 11 participants. 0% of days 1–14 were missing data for the 7 participants with functional SCRAM devices.

between SCRAM drinking duration with morning EMA and event-contingent EMA drinking duration were 0.31 and 0.29, respectively The Skyn demonstrated large, significant correlations with SCRAM peak TAC ($r_{rm} = 0.54$), AUC ($r_{rm} = 0.80$), and drinking duration ($r_{rm} = 0.63$), which were larger than the Skyn correlations with self-reported drinking intensity, duration, and eBAC. The Skyn showed null or small correlations with SCRAM rise rate ($r_{rm} = 0.19$) and fall rate ($r_{rm} = 0.29$).

***Sensitivity analyses.*** Due to the fact that there are currently no rules for identifying drinking episodes using Skyn data (van Egmond et al., 2020; Wang et al., 2021), we investigated the possibility that the cut-off value used for identifying drinking episodes (TAC ≥15 μg/L air) impacted the findings. To examine this possibility, and given prior research finding that Skyn TAC values stayed below 15 μg/L air in a participant who consumed one standard drink (Wang et al., 2021), sensitivity analyses explored using a lower cut-off of TAC ≥10 μg/L air for identifying drinking episodes. This lower cut-off increased the Skyn's false positive rate for detecting self-reported drinking from 12.9% with the ≥15 μg/L air cut-off to 24.2% with the ≥10 μg/L air cut-off and decreased the false negative rate from 42.4% with the ≥15 μg/L air cut-off to 35.4% with the ≥10 μg/L air cut-off. Lowering the cut-off did not improve sensitivity to detecting lighter drinking. Specifically, the ≥15 μg/L air cut-off resulted in 67 false negative days, whereas the ≥10 μg/L air cut-off resulted in 56 false negative days. Despite the decrease in the number of false negative days with the lower cut-off, the same proportion of false negatives (46%) occurred on days with self-reports of consuming one drink. Correlations between Skyn and self-report values decreased with the lower cut-off, as did correlations between Skyn and SCRAM values, suggesting that this lower cut-off value did not improve the ability to detect drinking versus not drinking, or day-level characteristics of drinking, with the Skyn.

## Discussion

This pilot study examined the acceptability and validity of using the Skyn across 28 days in individuals' natural settings and compared day-level features of alcohol use captured via the Skyn to

**Table 4**
Skyn detection of drinking and non-drinking days compared to SCRAM and self-report

| Detection rates | Skyn versus self-report | SCRAM versus self-report[a] |
|---|---|---|
| True positive [n (%)][b] | 91 (32.3%) | 35 (34.0%) |
| False positive [n (%)][c] | 16 (5.7%) | 5 (4.9%) |
| True negative [n (%)][d] | 102 (36.2%) | 42 (40.8%) |
| False negative [n (%)][e] | 73 (25.9%) | 21 (20.4%) |
| False positive rate (%)[c] | 13.6% | 10.6% |
| False negative rate (%)[e] | 44.5% | 37.5% |
| Sensitivity (%)[f] | 55.5% | 62.5% |
| Specificity (%)[g] | 86.4% | 89.4% |
| Positive predictive value (%)[h] | 85.0% | 87.5% |
| Negative predictive value (%)[i] | 58.3% | 66.7% |

[a] For comparisons with SCRAM data, participants with missing SCRAM data due to device malfunction were excluded ($n = 4$), such that, for self-report, there were 56 drinking days, 47 non-drinking days, and 2 days with missing data.

[b] True positives (TP) occur when the test measure (e.g., Skyn) detects a drinking day that corresponds with a drinking day based on the reference measure (i.e., self-report), excluding missing data.

[c] False positives (FP) occur when the test measure detects a drinking day that corresponds with a non-drinking day based on the reference measure, excluding missing data. The false positive rate is the proportion of days incorrectly identified as drinking days by the test measure versus the reference measure and was calculated as: FP/(FP + TN).

[d] True negatives (TN) occur when the test measure detects a non-drinking day that corresponds with a non-drinking day based on the reference measure, excluding missing data.

[e] False negatives (FN) occur when the test measures detect a non-drinking day that corresponds with a drinking-day based on the reference measure, excluding missing data. The false negative rate is the proportion of days incorrectly identified as non-drinking days by the test measure versus the reference measure and was calculated as: FN/(FN + TP).

[f] Sensitivity is the proportion of test-measured drinking days detected by the reference measure and was calculated as: TP / (TP + FN).

[g] Specificity is the proportion of test-measured non-drinking days detected by the reference measure and was calculated as: TN / (TN + FP).

[h] Positive predictive value is the likelihood that a day classified as a drinking day by the test measure was a drinking day as defined by the reference measure and was calculated as: TP/(TP + FP).

[i] Negative predictive value is the likelihood that a day classified as a non-drinking day by the test measure was a non-drinking day as defined by the reference measure and was calculated as: TN/(TN + FN).

self-report (across 28 days) and the SCRAM (across 14 days). Participants reported high acceptability for using the Skyn and preferred it to the SCRAM. It was feasible to collect Skyn TAC data; however, charging and battery life issues contributed significantly to missing data. The Skyn demonstrated reasonable correspondence with self-reported and SCRAM-defined drinking days and provided useful information for characterizing day-level features of alcohol use that corresponded with self-report and the SCRAM.

Our study contributes to the literature by comparing the acceptability of using the Skyn to the SCRAM across 14 days in natural settings and under conditions emulating the typical use of app-connected device-based measures of health behaviors (e.g., physical activity monitors), though results should be interpreted with caution given the high failure rate of the SCRAMs. Consistent with previous research (Ash et al., 2022), participants rated the Skyn as more acceptable than the SCRAM. Skyn acceptability ratings remained high across 28 days, supporting the ability to use the Skyn across extended timeframes. Participants' high acceptability ratings and assessments of the Skyn as comfortable, stylish, and unobtrusive aligned with previous research (Ash et al., 2022; Rosenberg et al., 2021; Wang et al., 2019, 2021), and addressed two SCRAM limitations – discomfort and social stigma (Alessi et al., 2017; Ash et al., 2022; Barnett et al., 2011, 2017; Sakai et al., 2006), which resulted in one participant dropping out of the study after three days. Also similar to previous research (Wang et al., 2021), participants found the Skyn app easy to use, even for technical novices. Perceived privacy concerns were similar for the Skyn and SCRAM, which was somewhat unexpected given that SCRAM data are secured within the device, whereas Skyn data are available in the app and uploaded to a cloud-based server, information that was communicated to participants during the informed consent process. However, this was the first study examining privacy concerns related to the Skyn, which should be assessed in other sociodemographic groups, as it is possible that privacy concerns could vary by age or experience with cloud-based storage (Auxier et al., 2019; Kunst, 2017).

Despite high acceptability, the Skyn received a "C" rating (scores ranging from 16 to 55%) for physical burden and a "B" rating (scores ranging from 56 to 85%) for difficulty of use. Several participants experienced skin irritation from the Skyn's mesh strap, which was unanticipated as other researchers did not report adverse skin reactions, though that may reflect the shorter time frame of previous studies (Ash et al., 2022; Rosenberg et al., 2021; Wang et al., 2021). Researchers should be aware of the potential for adverse skin reactions. Our participants experienced similar difficulties to those in previous studies related to trouble uploading data, charging the device, and maintaining the battery life, all of which resulted in data loss (Wang et al., 2021). Participants wanted the Skyn to have a longer battery life and for the device and/or app to provide notifications about remaining battery life (Wang et al., 2019, 2021). In this sample, the battery lasted less than 48 h, required daily charging, and one device needing to be replaced after only 28 days in use, suggesting the devices may have a short lifespan, findings that are consistent with Ash et al.'s study in which the Skyn T10 model's batteries typically failed prior to 72 h of field use and sensitivity decreased three months after device shipment (Ash et al., 2022). These difficulties were exacerbated by participant confusion regarding charging the device, as the Skyn T10 model needs to be powered off for the 'currently charging' light to turn on. This resulted in data loss, as several participants forgot to power the device back on after it was charged. Given consistent concerns regarding battery life and charging (Ash et al., 2022; Wang et al., 2019, 2021), the manufacturer updated the Skyn device, with the new T15 model touting a ~10-day battery life and the updated app including a battery life indicator, both of which should help address the aforementioned concerns and reduce participant burden and the risk of data loss related to the Skyn's battery life. Importantly, it is difficult to make direct comparisons between the outcomes of our study, which used the T10 model, with the outcomes of older studies using Skyn prototypes or future studies which may use the T15 (or newer) models.

Participants expressed concerns about the Skyn's lack of waterproofing or sufficient water resistance, which could limit the ability to use it across extended periods in field-based studies due to the risk of device damage (Ash et al., 2022; Fairbairn & Bosch, 2021; Wang et al., 2021), although the new Skyn T15 model has greater water resistance than the T10 model used in this study. Exposure to environmental alcohol could also impact the Skyn membrane's permeability and decrease sensor sensitivity, suggesting that, similar to the SCRAM, the Skyn membrane likely needs replaced periodically due to natural wear and tear (Ash et al., 2022; Wang et al., 2019). Participants questioned the Skyn's durability, which is notable given previously mentioned expectations that newer wrist-worn TAC sensors could expand opportunities for assessing alcohol use across longer timeframes (Ash et al., 2022; Fairbairn & Bosch, 2021; Wang et al., 2021). These battery life and durability issues give rise to practical concerns related to the administrative burden and costs associated with using the Skyn, particularly given that it is still a relatively expensive device (Fairbairn & Bosch, 2021). Overall, participants expressed high levels of acceptability for using the Skyn across 28 days in natural settings and they preferred it to the SCRAM. Recent improvements to the Skyn T15 model, including increased battery life and water resistance, should reduce user burden. Additional improvements addressing device durability could further enhance the ability to use the Skyn across extended time frames in natural settings.

Our study tested the Skyn in natural settings and it differed from Ash et al.'s study (2022) by permitting participants to indirectly expose the Skyn to environmental alcohol (e.g., via use of hand sanitizer) and by having participants use the Skyn app on their personal smartphones to sync data. It was feasible to collect Skyn data, with 74% of data available across the 28-day field protocol. This was comparable to a previous study using the WrisTAS sensor (Bond et al., 2014), but was lower than previous studies using the Skyn in which data were available for 90–96% of study days (Ash et al., 2022; Rosenberg et al., 2021). However, our study covered a longer time frame than previous Skyn studies [28 versus 5 days for Rosenberg et al. (2021) and 14 days for Ash et al. (2022), respectively], and we accounted for missing data from device removal, failed syncing, and battery failure. Notably, despite low overall missing data in Ash et al.'s study, battery failure was the primary cause of missing data in their study (2022). Although previous studies mentioned low temperature readings suggesting that participants may have removed their devices (Ash et al., 2022; Wang et al., 2021), ours is the first study providing detailed information about the proportion of Skyn data that were missing due to device removal (based on skin temperature readings) versus syncing or lost data issues. Our finding that a small proportion of data (5.1%) were lost due to device removal was an improvement compared to a study in which participants removed the WrisTAS for 10% of days (Simons et al., 2015). A larger proportion (20.7%) of data were lost due to failure to sync the data, charge the device, or incomplete syncing, which may be avoidable with the newer Skyn T15 model due to its longer battery life and the app's new battery life and connection status indicators. Including a battery life and charging indicator on the device itself would be useful for researchers and participants interested in using the Skyn across extended time frames in natural settings (Ash et al., 2022; Wang et al., 2021).

Despite missing data concerns, only 10% of Skyn devices ($n = 1$) failed, with the battery life lasting less than 48 h upon first use and shortening to less than 8 h with continued use. This was lower than a previous study in which 18–38% of Skyn prototypes failed (Fairbairn & Kang, 2019) and was lower than the WrisTAS, which failed on 18% of study days (Simons et al., 2015). In contrast, our SCRAM devices had an unusually high 43% failure rate. The SCRAM devices failed to record TAC, despite skin temperature and infrared voltage data indicating that the devices were properly attached. This was abnormally high, with previous pooled analysis finding only 5% of study days were impacted by SCRAM devices malfunctioning (Barnett et al., 2014). It is unclear why failure rates were so high, as these devices functioned well in detecting TAC in previous studies (Russell et al., 2022; West et al., 2021). It is possible that the SCRAM devices failed due to the number of days in service, as greater days in service predicts lower likelihood of detecting alcohol with the SCRAM (Marques & McKnight, 2009); however, the devices had been used for an average of 109 ± 27 days (range: 81–165), and there was no relationship between prior days of service and device failure ($p = .33$), suggesting that days of service does not account for high device failure rates. Despite these high failure rates, the SCRAM still holds some advantages over the Skyn. The SCRAM is sufficiently water resistant to be worn while

**Table 5**
Day-level repeated-measures correlations between Skyn, SCRAM, and Self-Reported drinking variables

| Comparisons | Drinking intensity (drinks consumed vs. peak TAC or peak TAC vs. peak TAC)[a] | | Drinking intensity (drinks consumed vs. AUC or AUC vs. AUC)[b] | | Drinking duration (morning EMA vs. time with TAC > 0)[c] | |
|---|---|---|---|---|---|---|
| | $r_{rm}$ (df) | [95% CI] | $r_{rm}$ (df) | [95% CI] | $r_{rm}$ (df) | [95% CI] |
| Skyn versus self-report | 0.35 (259) | [0.23, 0.47]*** | 0.52 (259) | [0.43, 0.62]*** | 0.36 (259) | [0.25, 0.47]*** |
| SCRAM versus self-report | 0.78 (67) | [0.64, 0.87]*** | 0.79 (67) | [0.66, 0.87]*** | 0.31 (95) | [0.12, 0.56]** |
| Skyn versus SCRAM | 0.51 (68) | [0.19, 0.75]*** | 0.80 (68) | [0.59, 0.92]*** | 0.63 (87) | [0.51, 0.78]*** |
| | **Drinking duration** (Event-contingent EMA vs. Time with TAC > 0) [d,e] | | **Estimated BAC** (Morning EMA vs. Peak TAC) [f] | | **Estimated BAC** (Event-contingent EMA vs. Peak TAC) [f] | |
| **Comparisons** | **$r_{rm}$ (df)** | **[95% CI]** | **$r_{rm}$ (df)** | **[95% CI]** | **$r_{rm}$ (df)** | **[95% CI]** |
| Skyn versus self-report | 0.37 (230) | [0.23, 0.52]*** | 0.30 (259) | [0.15, 0.41]*** | 0.23 (222) | [0.10, 0.39]*** |
| SCRAM versus self-report | 0.29 (89) | [0.12, 0.50]** | 0.61 (67) | [0.41, 0.77]*** | 0.42 (61) | [0.16, 0.68]*** |
| | **Estimated BAC** (Morning EMA vs. AUC) [f] | | **Estimated BAC** (Event-contingent EMA vs. AUC) [f] | | | |
| **Comparisons** | **$r_{rm}$ (df)** | **[95% CI]** | **$r_{rm}$ (df)** | **[95% CI]** | | |
| Skyn versus self-report | 0.48 (259) | [0.35, 0.60]*** | 0.42 (222) | [0.16, 0.62]*** | | |
| SCRAM versus self-report | 0.59 (67) | [0.42, 0.72]*** | 0.48 (61) | [0.10, 0.71]*** | | |
| | **Rise rate** [g] | | **Fall rate** [g] | | | |
| **Comparisons** | **$r_{rm}$ (df)** | **[95% CI]** | **$r_{rm}$ (df)** | **[95% CI]** | | |
| Skyn versus SCRAM | 0.18 (65) | [-0.10, 0.50] | 0.25 (68) | [0.02, 0.44]* | | |

Notes: SCRAM = Secure Continuous Remote Alcohol Monitor; TAC = Transdermal Alcohol Concentration; AUC = Area Under the Curve; EMA = Ecological Momentary Assessment; df = Degrees of freedom; CI = Confidence Interval

*** $p < .001$; ** $p < .01$; * $p < .05$

[a] For Skyn and SCRAM versus self-report, this is the number of self-reported drinks versus peak TAC. For Skyn versus SCRAM, this is comparisons of day-level peak TAC measured by both devices.

[b] For Skyn and SCRAM versus self-report, this is the number of self-reported drinks versus AUC. For Skyn versus SCRAM, this is comparisons of day-level AUC measured by both devices.

[c] For Skyn and SCRAM versus self-report, this is the amount of time spent drinking based on the morning EMA report. For Skyn versus SCRAM, this is the total amount of time with TAC > 0 μg/L air or TAC > 0 g/dL, respectively.

[d] For Skyn and SCRAM versus self-report, this is the amount of time spent drinking based on the event-contingent EMA versus time with TAC > 0 μg/L air (Skyn) or TAC > 0 g/dL (SCRAM).

[e] There are no values reported for Skyn versus SCRAM because the comparison for drinking duration is reported in the column labeled "Drinking Duration (Morning EMA vs. Time with TAC > 0).

[f] Estimated BAC was not compared for Skyn versus SCRAM because there is not a validated approach for estimating BAC from Skyn data.

[g] There are no values comparing Skyn versus self-report or SCRAM versus self-report because the self-report data did not permit calculations of rise and fall rates.

showering, the battery lasts for months, and it can be locked onto a participant's ankle, preventing device removal (Barnett et al., 2014; Karns-Wright et al., 2017; Marques & McKnight, 2009; van Egmond et al., 2020). Lastly, for both the Skyn and SCRAM there were no differences in missing data across social weekends (Thursday—Saturday) versus weekdays, which is meaningful given that the majority of drinking occurs on social weekends (Finlay et al., 2012). Overall, it was feasible to collect sufficient TAC data using the Skyn.

The Skyn showed a similar probability to the SCRAM for detecting self-reported drinking days (i.e., sensitivity). The Skyn's sensitivity rate (55.5%) was comparable to Ash et al.'s study (2022), in which sensitivity to detecting any drinking ranged from 40 to 60%; however, it was somewhat lower than studies in which sensitivity rates for wrist-worn TAC sensors ranged from 72.4% to 85.6% (Bond et al., 2014; Rosenberg et al., 2021; Simons et al., 2015). This may be due to differences between our study and previous studies regarding the methods used to identify drinking episodes or the types of participants, as participants in previous studies were heavier drinkers than our participants (Ash et al., 2022; Rosenberg et al., 2021). Ash et al.'s study of treatment-seeking heavy-drinking adults (2022) found that the Skyn showed 70—100% sensitivity for detecting heavy drinking and 67—89% sensitivity for detecting moderate drinking. In contrast, sensitivity for detecting any drinking in Ash et al.'s study (2022) was much lower (40—60%), supporting previous assertions that the Skyn may not be able to consistently detect low-level drinking (Rosenberg et al., 2021; Wang et al., 2021). Indeed, the Skyn often failed to detect self-reported drinking when participants consumed only one alcoholic beverage, even after testing a lower TAC cut-off. This was similar to previous studies in which some real-world drinking episodes of one to two drinks (Rosenberg et al., 2021; Wang et al., 2021) were detected by the Skyn, whereas some were not. This limitation could be addressed by refining data cleaning and processing rules for the Skyn, similar to how Roache et al. (2019) increased the SCRAM's sensitivity for detecting low-level drinking from 39.9% (based on the manufacturer's criteria) to 68.5% (Ash et al., 2022; van Egmond et al., 2020).

The Skyn showed greater agreement in detecting drinking days when compared to the SCRAM rather than self-report. This is likely because the Skyn and SCRAM assess the same phenomenon (i.e., TAC) and share the limited ability to detect low-level drinking (Ash et al., 2022; Roache et al., 2019; Rosenberg et al., 2021; van Egmond et al., 2020; Wang et al., 2019, 2021), though neither is a gold-standard measure. However, the Skyn sometimes disagreed with the SCRAM, such that 20.7% of days detected as drinking days by the Skyn were detected as non-drinking days by the SCRAM. This disagreement highlights a disadvantage of wrist-versus ankle-worn TAC sensors; namely, wrist-worn sensors are more susceptible to environmental contamination (Ash et al., 2022; van Egmond et al., 2020; Wang et al., 2021), such that researchers may consider instructing participants to avoid the use of alcohol-based cleaners or hand sanitizers while wearing the Skyn (Ash et al., 2022). This limitation could be addressed by developing rules for identifying and removing environmental contamination from Skyn TAC data (van Egmond et al., 2020; Wang et al., 2021).

This study is unique in comparing day-level features of alcohol use detected via the Skyn to both self-report and the SCRAM. The Skyn showed larger correlations with self-report than the SCRAM with regard to drinking duration, a finding that could be due to the Skyn assessing TAC more frequently than the SCRAM, which ultimately results in a smaller lag in TAC detection after drinking onset for the Skyn versus the SCRAM (Fairbairn & Kang, 2019). The Skyn's correlation between AUC and self-reported number of drinks ($r_{rm} = 0.52$) was lower than previous studies using the Skyn ($r = 0.72$) and WrisTAS ($r = 0.62$), but this may reflect statistical and methodological differences between our study and previous studies (Bond et al., 2014; Rosenberg et al., 2021). For instance, Bond et al. (2014) adjusted self-reported number of drinks based on ethanol content, resulting in a new variable that aligned better conceptually with TAC than number of drinks *per se* (Bond et al., 2014), potentially increasing the strength of the correlation between the WrisTAS and self-report.

Similarly, the Skyn's correlations with SCRAM-defined day-level features of alcohol use were larger than its correlations with self-report, although results should be interpreted with caution due to the small sample size with SCRAM data (n = 7). The Skyn and SCRAM assess the same phenomenon, TAC, whereas self-report does not account for within-person factors impacting alcohol metabolism (van Egmond et al., 2020; Wang et al., 2019), may not accurately reflect ethanol content of drinks (Bond et al., 2014), and is subject to recall biases (Alessi et al., 2019; Merrill et al., 2020). Skyn and SCRAM drinking duration represent the duration of biological alcohol exposure, whereas self-report represents the length of time individuals are actively drinking and does not account for the descending limb of alcohol exposure, such that only 25% of self-reported post-drinking assessments are 'descending limb moments' (Piasecki, 2019; Piasecki et al., 2014; Piasecki, Wood, Shiffman, Sher, & Heath, 2012). This highlights the value of researchers using alcohol sensors when they are specifically interested in examining the duration of biological alcohol exposure. The Skyn showed poor correlations with SCRAM rise and fall rates, which may reflect environmental contamination resulting in rapid increases and decreases in Skyn TAC data (Wang et al., 2021). Indeed, despite our attempts to remove environmental contamination from episodes based on rapid increases in Skyn TAC (rules 10 and 11), we were unable to remove all potential environmental contamination because it often overlapped with actual drinking episodes. Regardless of these challenges, Skyn TAC values appear to be useful in characterizing day-level features of alcohol use.

The following limitations should be considered. Despite the large volume of data, our small sample size and focus on early adults aged 26—40 years limit the generalizability of findings. Recruitment criteria permitted the inclusion of light or moderate drinkers, as we intended to test Skyn feasibility and validity across most types of drinkers, rather than limiting the study to the heavy-drinking populations typically used in previous Skyn studies (Ash et al., 2022; Wang et al., 2019). This resulted in the study including light (18%), moderate (55%), and heavy (27%) drinkers who consumed a mean of 2.8 drinks per drinking day, with 63.6% of participants engaging in binge drinking on at least one day of the study. We also excluded participants whose AUDIT scores were greater than 8, which may have resulted in the sample including both moderate/social drinkers and some heavy/problem drinkers, limiting generalizability to problem drinkers. However, the study still provides useful information about the Skyn's feasibility and validity across early adult participants with a range of drinking behaviors. Participants wore the SCRAM for 14 of the 28 days to reduce participant burden (Ash et al., 2022) and were permitted to remove the SCRAM, with 2.2% of data missing due to device removal. These features of the data collection, combined with the high SCRAM failure rates, limit the ability to draw strong conclusions about the comparability of real-world Skyn versus SCRAM data. Additional research is needed to fully compare real-world Skyn and SCRAM data and acceptability across extended periods, which would also enable assessing changes in Skyn versus SCRAM acceptability over time. It is possible that participants changed their drinking behaviors in response to using the Skyn and seeing their alcohol use data in the app; however, participants reported that seeing data did not impact their drinking, and previous research found that wearing the WrisTAS did not impact alcohol consumption or abstinence (Simons et al., 2015). Notably, the newest version of the Skyn app does not display TAC data; nonetheless, future studies examining participant reactivity to wearing the Skyn device are warranted, as reactivity could differ across sociodemographic groups or among clinical versus non-clinical populations (König, Allmeta, Christlein, Van Emmenis, & Sutton, 2022). Our method for processing Skyn data may have impacted findings, and previous studies have employed a variety of approaches for processing Skyn data (Ash et al., 2022; Fairbairn et al., 2020; Fairbairn & Kang, 2019; Rosenberg et al., 2021; Wang et al., 2019, 2021), which should be considered when interpreting our findings.

## Conclusions

Our findings support the acceptability and validity for using the Skyn to assess alcohol use across an extended timeframe (i.e., 28 days) in individuals' natural settings. Participants were willing to wear the Skyn across extended timeframes, and the Skyn provided useful information about day-level features of alcohol use that corresponded fairly well with other commonly used approaches for collecting alcohol use data in natural settings, including self-report and the SCRAM. Recent improvements to the Skyn's battery life and charging should enhance usability across extended time frames in natural settings. The development of methods for cleaning and processing Skyn data to detect low-level drinking and remove environmental contamination would enhance the utility of the Skyn, as well as the ability to compare findings across research studies.

## CRediT author statement

**Jimikaye Courtney:** Conceptualization, methodology, formal analysis, investigation, data curation, writing – original draft. **Michael Russell:** Conceptualization, resources, data curation, writing – review & editing. **David Conroy:** Conceptualization, resources, writing – review & editing, funding acquisition

## Acknowledgments

Jimikaye Courtney's time was supported by the Prevention and Methodology Training Program (T32 DA017629; MPIs: J. Maggs & S. Lanza) with funding from the National Institute on Drug Abuse. Study data were collected and managed using Research Electronic Data Capture (REDCap) tools hosted at Penn State Health Milton S. Hershey Medical Center and Penn State College of Medicine (Harris et al., 2009). In using REDCap, the project described was supported by the National Center for Advancing Translational Sciences, National Institutes of Health, through Grant UL1 TR002014 and Grant UL1 TR00045. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute on Drug Abuse or the National Institutes of Health. The authors have no conflicts of interest to declare.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.alcohol.2022.11.004.

## References

Alessi, S. M., Barnett, N. P., & Petry, N. M. (2017). Experiences with SCRAMx alcohol monitoring technology in 100 alcohol treatment outpatients. *Drug and Alcohol Dependence, 178*, 417–424.

Alessi, S. M., Barnett, N. P., & Petry, N. M. (2019). Objective continuous monitoring of alcohol consumption for three months among alcohol use disorder treatment outpatients. *Alcohol, 81*, 131–138.

Allison, P. D. (2019). *Fixed effects regression methods for longitudinal data using SAS*. SAS Institute.

Ash, G. I., Gueorguieva, R., Barnett, N. P., Wang, W., Robledo, D. S., DeMartini, K. S., 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*(5), 783–796.

Auxier, B., Rainie, L., Anderson, M., Perrin, A., Kumar, M., & Turner, E. (2019). Americans and privacy: Concerned, confused and feeling lack of Control over their personal information. Pew research center. Retrieved from https://www.pewresearch.org/internet/2019/11/15/americans-and-privacy-concerned-confused-and-feeling-lack-of-control-over-their-personal-information/.

Babor, T. F., Higgins-Biddle, J. C., Saunders, J. B., & Monteiro, M. G. (2001). *The alcohol use disorders identification test: Guidelines for use in primary care*. World Health Organization. Retrieved from. http://www.psiholocator.com/images/who_msd_msb_016a.pdf.

Bakdash, J. Z., & Marusich, L. R. (2017). Repeated measures correlation. *Frontiers in Psychology, 8*, 456.

Barnett, N. P. (2015). Alcohol sensors and their potential for improving clinical care. *Addiction, 110*(1), 1–3.

Barnett, N. P., Celio, M. A., Tidey, J. W., Murphy, J. G., Colby, S. M., & Swift, R. M. (2017). A preliminary randomized controlled trial of contingency management for alcohol use reduction using a transdermal alcohol sensor. *Addiction, 112*(6), 1025–1035.

Barnett, N. P., Meade, E. B., & Glynn, T. R. (2014). Predictors of detection of alcohol use episodes using a transdermal alcohol sensor. *Experimental and Clinical Psychopharmacology, 22*(1), 86–96.

Barnett, N. P., Tidey, J., Murphy, J. G., Swift, R., & Colby, S. M. (2011). Contingency management for alcohol use reduction: A pilot study using a transdermal alcohol sensor. *Drug and Alcohol Dependence, 118*(2–3), 391–399.

Bond, J. C., Greenfield, T. K., Patterson, D., & Kerr, W. C. (2014). Adjustments for drink size and ethanol content: New results from a self-report diary and transdermal sensor validation study. *Alcoholism: Clinical and Experimental Research, 38*(12), 3060–3067.

Campbell, A. S., Kim, J., & Wang, J. (2018). Wearable electrochemical alcohol biosensors. *Current Opinion in Neuropharmacology, 10*, 126–135.

Centers for Disease Control and Prevention. (2019). Glossary – alcohol. Retrieved from https://www.cdc.gov/nchs/nhis/alcohol/alcohol_glossary.htm.

Clapp, J. D., Madden, D. R., Gonzalez Villasanti, H., Giraldo, L. F., Passino, K. M., Reed, M. B., et al. (2018). A system dynamic model of drinking events: Multi-level ecological approach: A system dynamic model of drinking events. *Systems Research and Behavioral Science, 35*(3), 265–281.

Clapp, J. D., Madden, D. R., Mooney, D. D., & Dahlquist, K. E. (2017). Examining the social ecology of a bar-crawl: An exploratory pilot study. *PLoS One, 12*(9). e0185238.

Cooper, M. L. (1994). Motivations for alcohol use among adolescents: Development and validation of a four-factor model. *Psychological Assessment, 6*(2), 117–128.

Davis, C. G., Thake, J., & Vilhena, N. (2010). Social desirability biases in self-reported alcohol consumption and harms. *Addictive Behaviors, 35*(4), 302–311.

Dougherty, D. M., Lake, S. L., Hill-Kapturczak, N., Liang, Y., Karns, T. E., Mullen, J., et al. (2015). Using contingency management procedures to reduce at-risk drinking in heavy drinkers. *Alcoholism: Clinical and Experimental Research, 39*(4), 743–751.

van Egmond, K., Wright, C. J. C., Livingston, M., & Kuntsche, E. (2020). Wearable transdermal alcohol monitors: A systematic review of detection validity, and relationship between transdermal and breath alcohol concentration and influencing factors. *Alcoholism: Clinical and Experimental Research, 44*(10), 1918–1932.

Evans, B. (2016). Paco – applying computational methods to scale qualitative methods. *Ethnographic Praxis in Industry - Conference Proceedings, 1*, 348–368.

Evans, B. (2021). Paco: The personal analytics companion [WWW document]. PACO. Available at: https://github.com/google/paco .

Fairbairn, C. E., & Bosch, N. (2021). A new generation of transdermal alcohol biosensing technology: Practical applications, machine-learning analytics and questions for future research. *Addiction, 116*(10), 2912–2920.

Fairbairn, C. E., & Kang, D. (2019). Temporal dynamics of transdermal alcohol concentration measured via new-generation wrist-worn biosensor. *Alcoholism: Clinical and Experimental Research, 43*(10), 2060–2069.

Fairbairn, C. E., 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.

Finlay, A. K., Ram, N., Maggs, J. L., & Caldwell, L. L. (2012). Leisure activities, the social weekend, and alcohol use: Evidence from a daily study of first-year college students. *Journal of Studies on Alcohol and Drugs, 73*(2), 250–259.

Greenfield, T. K., & Kerr, W. C. (2003). Tracking alcohol consumption over time. *Alcohol Research & Health, 27*(1), 30–38.

Harris, P. A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., & Conde, J. G. (2009). Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Informatics, 42*(2), 377–381.

Holder, H. D. (2006). *Alcohol and the community: A Systems approach to prevention (International research monographs in the addictions)*. Cambridge: Cambridge University Press.

Hultgren, B. A., Scaglione, N. M., Buben, A., & Turrisi, R. (2020). Examining protocol compliance and self-report congruence between daily diaries and event-contingent ecological momentary assessments of college student drinking. *Addictive Behaviors, 110*, 106471.

Karns-Wright, T. E., Roache, J. D., Hill-Kapturczak, N., Liang, Y., Mullen, J., & Dougherty, D. M. (2017). Time delays in transdermal alcohol concentrations relative to breath alcohol concentrations. *Alcohol and Alcoholism, 52*(1), 35–41.

König, L. M., Allmeta, A., Christlein, N., Van Emmenis, M., & Sutton, S. (2022). A systematic review and meta-analysis of studies of reactivity to digital in-the-moment measurement of health behaviour. *Health Psychology Review, 16*(4), 551–575.

Kunst, A. (2022, February 28). Usage of cloud storage services for data by U.S. consumers 2017, by age group. Statistica. https://www.statista.com/statistics/714163/us-usage-cloud-storage-services-age-group/

Leffingwell, T. R., Cooney, N. J., Murphy, J. G., Luczak, S., Rosen, G., Dougherty, D. M., et al. (2013). Continuous objective monitoring of alcohol use: Twenty-first century measurement using transdermal sensors. *Alcoholism: Clinical and Experimental Research, 37*(1), 16–22.

Luczak, S. E., Hawkins, A. L., Dai, Z., Wichmann, R., Wang, C., & Rosen, I. G. (2018). Obtaining continuous BrAC/BAC estimates in the field: A hybrid system integrating transdermal alcohol biosensor, intellidrink smartphone app, and BrAC estimator software tools. *Addictive Behaviors, 83*, 48–55.

Luczak, S. E., & Ramchandani, V. A. (2019). Special issue on alcohol biosensors: Development, use, and state of the field: Summary, conclusions, and future directions. *Alcohol, 81*, 161–165.

Marques, P. R., & McKnight, A. S. (2009). Field and laboratory alcohol detection with 2 types of transdermal devices. *Alcoholism: Clinical and Experimental Research, 33*(4), 703–711.

Matthews, D. B., & Miller, W. R. (1979). Estimating blood alcohol concentration: Two computer programs and their applications in therapy and research. *Addictive Behaviors, 4*(1), 55–60.

Merrill, J. E., Fan, P., Wray, T. B., & Miranda, R., Jr. (2020). Assessment of alcohol use and consequences: Comparison of data collected via timeline followback interview and daily reports. *Journal of Studies on Alcohol and Drugs, 81*(2), 212–219.

National Institute on Alcohol Abuse and Alcoholism. (2021). Alcohol facts and statistics. https://www.niaaa.nih.gov/publications/brochures-and-fact-sheets/alcohol-facts-and-statistics.

National Institute on Alcohol Abuse and Alcoholism. (2022). Rethinking drinking — alcohol & your health. Retrieved from https://www.rethinkingdrinking.niaaa.nih.gov/.

Piasecki, T. M. (2019). Assessment of alcohol use in the natural environment. *Alcoholism: Clinical and Experimental Research, 43*(4), 564–577.

Piasecki, T. M., Cooper, M. L., Wood, P. K., Sher, K. J., Shiffman, S., & Heath, A. C. (2014). Dispositional drinking motives: Associations with appraised alcohol effects and alcohol consumption in an ecological momentary assessment investigation. *Psychological Assessment, 26*(2), 363–369.

Piasecki, T. M., Wood, P. K., Shiffman, S., Sher, K. J., & Heath, A. C. (2012). Responses to alcohol and cigarette use during ecologically assessed drinking episodes. *Psychopharmacology (Berl), 223*(3), 331–344.

Rehm, J., Mathers, C., Popova, S., Thavorncharoensap, M., Teerawattananon, Y., & Patra, J. (2009). Global burden of disease and injury and economic cost attributable to alcohol use and alcohol-use disorders. *Lancet, 373*(9682), 2223–2233.

Roache, J. D., Karns-Wright, T. E., Goros, M., Hill-Kapturczak, N., Mathias, C. W., & Dougherty, D. M. (2019). Processing transdermal alcohol concentration (TAC) data to detect low-level drinking. *Alcohol, 81*, 101–110.

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 (preprint). *Epidemiology*.

Russell, M. A., Turrisi, R. J., & Smyth, J. M. (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*(1), 100–113.

Sakai, J. T., Mikulich-Gilbertson, S. K., Long, R. J., & Crowley, T. J. (2006). Validity of transdermal alcohol monitoring: Fixed and self-regulated dosing. *Alcoholism: Clinical and Experimental Research, 30*(1), 26–33.

Simons, J. S., Wills, T. A., Emery, N. N., & Marks, R. M. (2015). Quantifying alcohol consumption: Self-report, transdermal assessment, and prediction of dependence symptoms. *Addictive Behaviors, 50*, 205–212.

Spagnolli, A., Guardigli, E., Orso, V., Varotto, A., & Gamberini, L. (2014). Measuring user acceptance of wearable Symbiotic devices: Validation study across application scenarios. In G. Jacucci, L. Gamberini, J. Freeman, & A. Spagnolli (Eds.), *Symbiotic interaction, lecture notes in computer science*. Springer international publishing.

Substance Abuse and Mental Health Services Administration. (2022), 2018 NSDUH Detailed Tables https://www.samhsa.gov/data/report/2018-nsduh-detailed-tables.

Suh, H., Shahriaree, N., Hekler, E. B., & Kientz, J. A. (2016). Developing and Validating the user burden scale: A Tool for assessing user Burden in computing systems. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 3988–3999.

Swift, R. (2000). Transdermal alcohol measurement for estimation of blood alcohol concentration. *Alcoholism: Clinical and Experimental Research, 24*(4), 422–423.

Swift, R. (2003). Direct measurement of alcohol and its metabolites. Addiction. *2*(98 Suppl), 73–80.

Thornton, L., Osman, B., Champion, K., Green, O., Wescott, A. B., Gardner, L. A., et al. (2022). Measurement properties of smartphone approaches to assess diet, alcohol use, and tobacco use: Systematic review. *JMIR mHealth and uHealth, 10*(2), e27337.

Wang, Y., Fridberg, D. J., Leeman, R. F., Cook, R. L., & Porges, E. C. (2019). Wrist-worn alcohol biosensors: Strengths, limitations, and future directions. *Alcohol, 81*, 83–92.

Wang, Y., Fridberg, D. J., Shortell, D. D., Leeman, R. F., Barnett, N. P., Cook, R. L., et al. (2021). Wrist-worn alcohol biosensors: Applications and usability in behavioral research. *Alcohol, 92*, 25–34.

West, A. B., Bomysoad, R. N., Russell, M. A., & Conroy, D. E. (2022). Daily physical activity, sedentary behavior and alcohol use in at-risk college students. *Annals of Behavioral Medicine, 56*(7), 712–725.

White, H. R., & Jackson, K. (2004). Social and psychological influences on emerging adult drinking behavior. *Alcohol Research & Health, 28*(4), 182–190.