

RESEARCH ARTICLE



Sensitivity, specificity, and tolerability of the BACTrack Skyn compared to other alcohol monitoring approaches among young adults in a field-based setting

Garrett I. Ash^{1,2} | Ralitzia Gueorguieva³ | Nancy P. Barnett⁴ | Wuyi Wang¹ | David S. Robledo¹ | Kelly S. DeMartini¹ | Brian Pittman¹ | Nancy S. Redeker^{1,5} | Stephanie S. O'Malley¹ | Lisa M. Fucito^{1,6,7}

¹Yale School of Medicine, New Haven, Connecticut, USA

²Veterans Affairs Connecticut Healthcare System, West Haven, Connecticut, USA

³Yale School of Public Health, New Haven, Connecticut, USA

⁴Brown School of Public Health, Providence, Rhode Island, USA

⁵Yale School of Nursing, Orange, Connecticut, USA

⁶Yale Cancer Center, New Haven, Connecticut, USA

⁷Smilow Cancer Hospital, Yale-New Haven Hospital, New Haven, Connecticut, USA

Correspondence

Garrett I. Ash, Veterans Affairs Connecticut Healthcare System (11-ACSLG), Building 35A, Rm 235, 950 Campbell Avenue, West Haven, CT 06516, USA.

Email: garrett.ash@yale.edu

Funding information

National Institute on Alcohol Abuse and Alcoholism, Grant/Award Number: R34AA026021, R34AA026021-01A1S1 and 1R21AA028886; American Heart Association, Grant/Award Number: #852679/Garrett Ash/2021-2024; Robert E. Leet and Clara Guthrie Patterson Trust Mentored Research Award, Bank of America, N.A., Trustee; Fellowship from Office of Academic Affiliations at the United States Veterans Health Administration; National Center for Advancing Translational Sciences, Grant/Award Number: UL1TR000142

Abstract

Background: There is a need for novel alcohol biosensors that are accurate, able to detect alcohol concentration close in time to consumption, and feasible and acceptable for many clinical and research applications. We evaluated the field accuracy and tolerability of novel (BACTrack Skyn) and established (Alcohol Monitoring Systems SCRAM CAM) alcohol biosensors.

Methods: The sensor and diary data were collected in a larger study of a biofeedback intervention and compared observationally in the present sub-study. Participants (high-risk drinkers, 40% female; median age 21) wore both Skyn and SCRAM CAM sensors for 1–6 days and were instructed to drink as usual. Data from the first cohort of participants ($N = 27$; 101 person-days) were used to find threshold values of transdermal alcohol that classified each day as meeting or not meeting defined levels of drinking (heavy, above-moderate, any). These values were used to develop scoring metrics that were subsequently tested using the second cohort ($N = 20$; 57 person-days). Data from both biosensors were compared to mobile diary self-report to evaluate sensitivity and specificity in relation to *a priori* standards established in the literature.

Results: Skyn classification rules for Cohort #1 within 3 months of device shipment showed excellent sensitivity for heavy drinking (94%) and exceeded expectations for above-moderate and any drinking (78% and 69%, respectively), while specificity met expectations (91%). However, classification worsened when Cohort #1 devices ≥ 3 months from shipment were tested (area under curve for receiver operator characteristic 0.87 vs. 0.79) and the derived classification threshold when applied to Cohort #2 was inadequately specific (70%). Skyn tolerability metrics were excellent and exceeded the SCRAM CAM ($p \leq 0.001$).

Conclusions: Skyn tolerability was favorable and accuracy rules were internally derivable but did not yield useful scoring metrics going forward across device lots and months of usage.

KEYWORDS

alcohol use, biosensors, self-monitoring, transdermal alcohol sensors, wearable alcohol sensors

INTRODUCTION

Alcohol use monitoring has many benefits, including identifying candidates for early intervention to promote moderation and abstinence, self-awareness, self-regulation (Barnett et al., 2017; Davis-Martin et al., 2021; Dougherty et al., 2015), and legally enforced sobriety to protect the public. Mobile alcohol self-management applications have ≥2.7 million downloads (Hoepfner et al., 2017) and judicial-ordered monitoring has enrolled ~1 million clients (Alcohol Monitoring Systems, 2021). Besides frequent breathalyzer operation, the best-validated monitoring is self-report, but it suffers from biases, especially underreporting and misestimation of drinking timing (Alessi et al., 2019; Hultgren et al., 2020; Merrill et al., 2020). These problems have been partially addressed by the Alcohol Monitoring Systems SCRAM CAM (Littleton, CO) ankle bracelet which measures transdermal alcohol content (TAC) but has limitations. Its size, 6-ounce weight, ankle location, and noise/vibration can create physical and social discomfort and disrupt daily activities like exercise and sleep. Another limitation of the SCRAM CAM is that data upload occurs by physical proximity to a home or office base station rather than in real-time. These issues challenge long-term feasibility and ecological validity of the use of the SCRAM CAM (Barnett et al., 2017) and limit its application for a range of clinical and research questions. Other limitations shared by other TAC sensors are that the TAC curve has a substantial and variable time delay from measured BAC (Leffingwell et al., 2013) and sensitivity to sub-heavy drinking levels (i.e., 1–3 standard drinks) is poor to fair: 60%–73% for 90% specificity (Karns-Wright et al., 2018; Roache et al., 2019).

Through recent advances in alcohol biosensor technology, a new generation of TAC sensors have emerged with the potential to expand the opportunities for real-time, objective monitoring of alcohol consumption. Among these new sensors, the most promising is the Skyn by BACtrack (San Francisco, CA) (Davis-Martin et al., 2021; Fairbairn & Bosch, 2021; Wang et al., 2019). Compared to the SCRAM CAM, the Skyn is wrist-worn, less invasive (Wang et al., 2019, 2021), samples alcohol exposure more frequently, releases data immediately by Bluetooth, and detects drinking with less time delay (Fairbairn & Kang, 2019; Fairbairn et al., 2020). With these features, the Skyn is similar to commercially available health/fitness biosensors, which guide individuals to passively quantify, visualize, and cogitate biometrics for the sake of motivating and supporting self-management decisions (Ash et al., 2020a, 2021; Sim, 2019). The data from TAC sensors like the Skyn could allow researchers/clinicians to devise behavior change strategies that leverage real-time, personalized feedback about alcohol-health associations (Davis-Martin et al., 2021; Wang et al., 2021). For instance, there has been an explosion in the importance of sleep (Worley, 2018) and technology options for improving it (Choi et al., 2018; Goldman, 2017) and we previously demonstrated that young adults are interested in feedback about alcohol's effects on sleep and the utility of sleep for engaging them about their drinking (Ash et al., 2020b; Fucito et al., 2015, 2017, 2021). Older alcohol biosensor technology such as the SCRAM CAM

can disrupt sleep for some participants due to discomfort (Alessi et al., 2017; Barnett et al., 2017), but the Skyn and new-generation devices likely have minimal impact on sleep, thereby increasing the opportunity for these novel alcohol-health associations and personalized feedback applications. Besides sleep, the SCRAM CAM's weight and appearance disrupt other aspects of daily living for some participants including exercise, clothing choice, social comfort, and physical comfort (Barnett et al., 2017), which could impact drinking behaviors and confound alcohol-health associations. Thus, an important question is whether new TAC technology like the Skyn avoids such disruptions. To date, prior research has focused on the user experience of viewing Skyn data on its mobile app (Wang et al., 2021). No studies have compared the Skyn or similar new biosensors with the SCRAM CAM or other alcohol self-monitoring tools (e.g., daily diaries) on measures of disruption to daily living.

The Skyn also requires further study for its accuracy. Previous investigators issued the Skyn to predominantly non-treatment-seeking, social drinkers who volunteered to drink at a standardized time, to test the time delay between drinking and TAC elevation (Fairbairn & Kang, 2019; Fairbairn et al., 2020; Wang et al., 2021). However, they did not extend this quantitative analysis to Skyn TAC results in a field setting where participants could freely choose to drink or not and the timing of their drinking. In the real world, these drinking choices typically vary day-to-day.

One study (Wang et al., 2021) conducted preliminary research in a field setting by issuing the Skyn to non-treatment-seeking social drinkers across the adult lifespan (21+ years old) living with and without HIV who were instructed to freely drink during daily living. The authors presented a subjective description of visual results from 1 (subjectively chosen) of the 12 participants. They noted that the Skyn produced 8 TAC elevations over 14 days that looked like those observed in their laboratory experiments, however, five of these eight elevations did not correspond to a mobile diary self-report (i.e., potential false positives). There were also two instances with no noticeable elevation despite positive diary self-report (i.e., potential false negatives). In summary, subjective visual inspection for drinking events yielded 60% sensitivity and 50% specificity for one selected participant. The other 11 participants were presented as supplemental data. We downloaded those files and found that compared to the one case example presented in the main text, many had even more visual anomalies (e.g., 1 or more TAC elevations daily from a sample of participants supposedly identified as moderate drinkers, and shape of TAC elevations often very different than results from controlled drinking), and furthermore no self-report data were included to tabulate false positives or negatives. Thus, visual inspection of these data was clearly inadequate to detect drinking, and furthermore, visual inspection could not be translated into future automated scoring rules. The authors stressed these limitations and called for further studies objectively analyzing Skyn readings in a field setting with unknown drinking timing, to objectively calculate accuracy and rules going forward.

Furthermore, drinking classification accuracy may vary by drinking quantity: when such accuracy has been tested for the SCRAM

CAM, sensitivity was 93% for heavy drinking but 60%–73% for lower drinking levels (Barnett et al., 2014; Karns-Wright et al., 2018; Roache et al., 2015, 2019). Even among these lower levels, two to three standard drinks were more easily detected than one standard drink (Roache et al., 2015). Finally, TAC accuracy might vary with field conditions such as temperature, contaminants, and physical perturbations. Thus, there is a need to test how accurately the Skyn can distinguish drinking from non-drinking without *a priori* knowledge of drinking timing, whether this accuracy varies by drinking quantity, and whether useful drinking detection metrics can be developed that are robust to varying field conditions.

To address these gaps and advance the science, we conducted a field-based study of the Skyn with heavy-drinking young adults. We sought to determine the broad acceptability (i.e., tolerability) of field-based Skyn use, test its accuracy to detect drinking, and determine whether automated scoring metrics to detect drinking could be used going forward. We hypothesized that the Skyn would compare favorably on tolerability metrics to the SCRAM CAM and a mobile diary report method, detect drinking with similar accuracy as the SCRAM CAM worn by the same individuals in the present study and reported from prior studies, and lead to the development of drinking detection rules that could be used across multiple cohorts. The sample was young adults who reported recent heavy drinking and tested positive on a validated alcohol screener (DeMartini & Carey, 2012) in a naturalistic setting.

MATERIAL AND METHODS

Reporting follows STARD guidelines (Bossuyt et al., 2015; Appendix S1).

Study design

This investigation was a planned sub-study of a parent trial in which participants completed 14 days of monitoring with the SCRAM CAM and a sleep biosensor and daily smartphone diaries under instructions to drink as usual (NCT03658954 [Fucito et al., 2021]). They returned to the office weekly for psychoeducation and feedback from the devices on drinking and sleep behaviors (Figure 1). For the current sub-study, a subset of participants was fit for a Skyn Generation

2 watch on their dominant wrist as a third biosensor at the intake visit which was synced and recharged at the day 8 visit. To avoid bias, no field charging/syncing was permitted so one battery life duration per week was collected after which participants removed it.

The first 65 self-reported drinking episodes almost always ($k = 62$, 94%) started after 12:00 pm and often ($k = 35$, 54%) continued past 12:00 am, so we demarcated the data by “social days” (12:00 pm–12:00 pm) rather than calendar days. A total of 158 social days were recorded which were truncated on the front end for study start times after 12:00 pm (latest 12:15 am) or the back end if the Skyn battery died before 12:00 pm the next day (earliest 2:50 am). On average, these days contained 21.8 ($SD = 2.8$) hours of Skyn data. Initially, 171 days were collected but five (2.9%) were removed because the SCRAM CAM detected a manual tamper or major alcohol contaminant (defined in Sect. SCRAM CAM, *Drinking classification*), four (2.3%) because the SCRAM CAM had a technical malfunction, three (1.8%) because research staff missed a Skyn download, and one (0.6%) because drinking self-report was not completed. This extremely low rate of self-report non-completion was reflective of the same metric in the parent trial (<4%). In summary, the average participant in the 14-day study (total 658 person-days) provided 3.4 days of Skyn data (158 person-days), with 97.4% of the non-recorded days attributable to battery life limitations.

The Yale University Institutional Review Board approved the study. All volunteers signed the informed consent document. Participants were compensated up to US \$311 for completing procedures for the parent trial including up to \$3 per day for adherence to sensors and diaries.

Participants

We recruited participants via paid social media advertising and flyers posted in the community (Ash et al., 2020b). Enrolled participants were heavy drinkers (≥ 3 heavy drinking occasions [≥ 4 standard drinks for women, ≥ 5 for men; Corbin et al., 2014] in the last 2 weeks and Alcohol Use Disorders Identification Test Consumption Questions scores ≥ 5 for women and ≥ 7 for men indicative of risk of harm from drinking [DeMartini & Carey, 2012]) aged 18–25 years who read and understood English, owned smartphones, and were willing/able to complete daily diaries and wear alcohol biosensors. Participants were excluded for: (1) current,

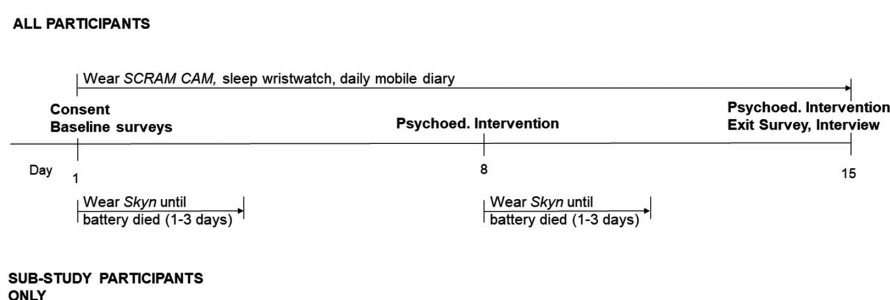


FIGURE 1 Single participant timeline. In some cases, participants only completed one of the two Skyn-wearing periods due to inventory

severe alcohol use disorder (e.g., withdrawal symptoms), (2) substance use disorder or positive urine drug screen for substances other than marijuana, (3) sleep disorders, (4) night or rotating shift work, (5) current enrollment in alcohol or sleep treatment, or (6) medical conditions preventing SCRAM CAM use (circulation problems, neuropathy, deep vein thrombosis, leg ulcers, tendonitis, diabetes, pregnancy, history of swelling, metal allergies, pacemaker, or any other implanted medical device).

The parent trial enrolled 120 participants who started the intervention (Figure 2). Among them, 69 enrolled when we could access Skyn wristwatch devices (between August 2, 2019 and February 24, 2021). Among this subset, participants were excluded as follows: three declined sustained wearing of the SCRAM CAM, 1 had a SCRAM CAM technical malfunction during the entire wear period, and one did not comply with diary self-report. We initially excluded a further 18 participants because their randomization group did not include daily diary self-report but reinstated 3 of them because they completed the parent trial timeline followback (TLFB) follow-up assessment at the day 15 visit which we have found typically has an excellent day-level agreement with the prior 14 days' diaries. The other 15 out of 18 participants in this randomization group were excluded because they did not complete the follow-up TLFB until the day 28 visit, at which point its day-level agreement with the day 1–14 diaries was not acceptable. After these exclusions, the remaining 47 participants were consecutively included in this sub-study.

We used data from the first sequential 27 participants (*Cohort #1*, 101 person-days collected August 2019–March 2020 before the COVID-19 pandemic) to explore and establish classification rules for the Skyn. Data from the remaining 20 participants (*Cohort #2*, 57 person-days collected September 2020–March 2021 during the COVID-19 pandemic) were used to test these Skyn rules and compare performance to the SCRAM CAM. COVID-19 pandemic restrictions limited the *Cohort #2* sample size.

Instructions and screening

Participants were instructed to arrive at office visits with 0.00% blood alcohol levels, confirmed by breathalyzer (*Cohort #1*) or self-report (*Cohort #2*). They were also instructed to avoid alcohol-containing personal hygiene and household cleaning products (e.g., mouthwash, medicinal alcohol) throughout the study. Participants were told to contact research staff if having any intolerable discomfort with devices so staff could travel to adjust the device or offer the option of removal, though no such removals were made.

The Structured Clinical Interview for the Diagnostic and Statistical Manual of Mental Disorders, 5th Edition assessed current and past substance use disorders and other current psychiatric diagnoses (First et al., 2014). Other criteria were assessed by self-report.

Self-reported alcohol use

Timeline followback (TLFB)

At baseline, the 30-day Timeline Followback Interview (TLFB) (Sobell et al., 2003) was administered by a research assistant to obtain alcohol use frequency and quantity data. Participants were taught standard drink definitions (14 g EtOH) using visual aids and with personalized guidance based upon concentration and container size of typical drinks.

Daily diary

Sensor readings and self-report data assessed the same drinking events during which no clinical interventions occurred. Participants received a daily prompt via smartphone text to a mobile-friendly webform (Qualtrics, Provo, UT) at their expected wake time each morning, with up to two reminder prompts sent as needed (0.5 h and 2.5 h later). When participants indicated they drank yesterday, they were presented questions from the standardized and validated Drinking Self-Monitoring Log (Sobell & Sobell, 1993): "(1) Yesterday, what time did you START drinking alcohol? (2) What time did you STOP drinking alcohol? (3) How many TOTAL standard alcoholic drinks did you have?" Standard drink definitions from the TLFB interview were reiterated. Diary responses were exported from Qualtrics into an SPSS master database where each response represented one case. We inserted an additional variable indicating the social day (i.e., 12 pm–12 pm) referenced by the response, which was calculated as the date stamp of the response minus 1. This validated diary was chosen as a reference standard based on literature precedent (Barnett et al., 2017; Karns-Wright et al., 2018; Roache et al., 2019) and its close temporal proximity to the sensor measurements.

Drinking classification

Using established national criteria, heavy drinking was defined as ≥ 4 standard drinks on 1 occasion for women, ≥ 5 standard drinks on 1 occasion for men (Corbin et al., 2014). Above-moderate drinking was considered > 1 standard drink on 1 occasion for women, > 2 standard drinks on one occasion for men (U.S. Department of Agriculture & U.S. Department of Health & Human Services., 2020). Any drinking was defined as ≥ 1 standard drink on one occasion.

SCRAM CAM

Inventory

We purchased 10 SCRAM CAM devices in November 2018 for use with both cohorts, though some were replaced due to device failure.

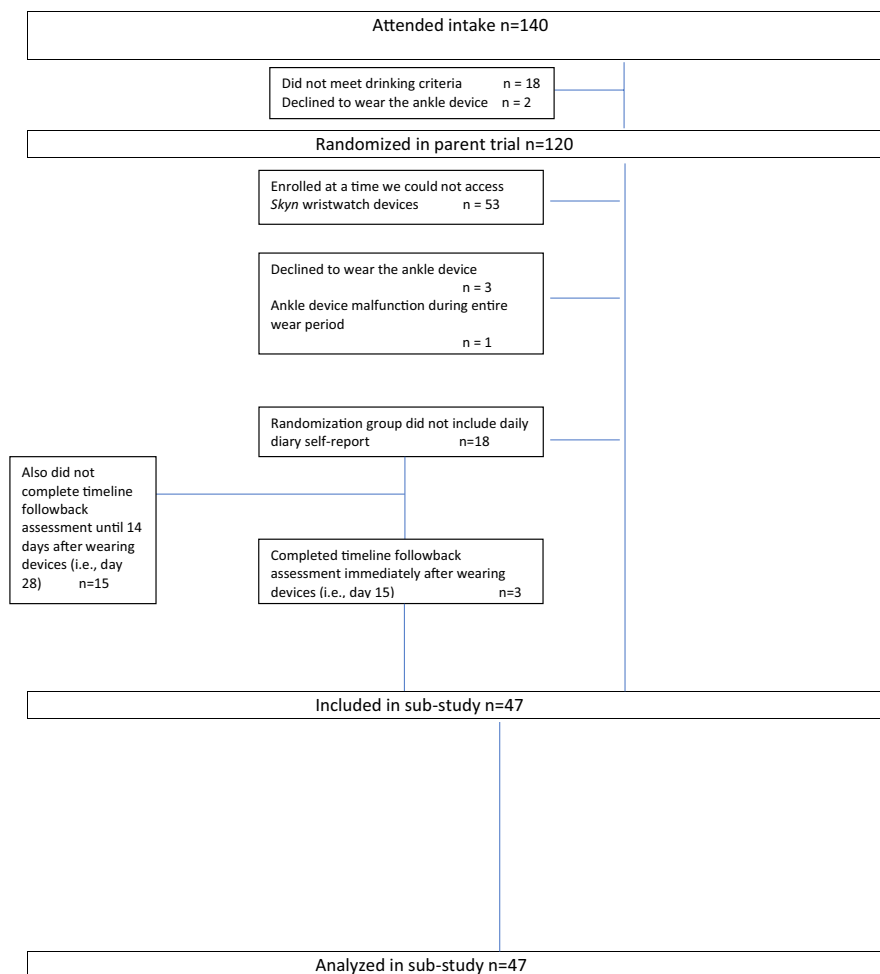


FIGURE 2 Flow of participants



FIGURE 3 SCRAM CAM (left), BACtrack Skyn (middle), and a US quarter (right, for size reference). Adapted with permission (Wang et al., 2019)

Two SCRAM CAMs were replaced before Cohort #1, two during Cohort #1, and three during Cohort #2.

Data collection

The SCRAM CAM is worn just above the ankle. Its compartment (at the back of Figure 3) contains a fuel cell that produces an electric current when exposed to alcohol, as well as a pump that pushes air vapor from the skin to the fuel cell every 30min. These items are protected by a metal faceplate that rests on the participant's skin and is replaced between participants.

The SCRAM CAM device was attached and adjusted for fit in the laboratory by the research staff. Participants were asked to not remove the device for any reason except a medical emergency. They were also instructed to avoid submerging the device in water, shower normally, use no products other than soap and water in the area around the bracelet, but clean around and under the device daily with mild soap and inspect for any redness, irritation, or bruising. The device was linked to the participant's anonymous ID number on the manufacturer's *Scramnet* cloud platform and uploaded weekly by magnet and USB-connected plastic receptacle. Participants were reminded their participation was not judicial-mandated, data would not be reported to any legal or other bodies, and there were no penalties for drinking.

Drinking classification

For each participant, a csv file containing the individual readings and demarcation of drinking episodes was downloaded from *Scramnet*. *Scramnet* defines drinking events as ≥ 3 transdermal alcohol concentration (TAC) points above the minimum value of 0.02 g/dl, rise from baseline to peak (i.e., absorption) < 0.05 g/dl, and fall from peak to baseline (i.e., elimination) < 0.025 g/dl (if peak < 0.15 g/dl) or < 0.015 (if peak > 0.15 g/dl). *Scramnet*'s subsequent analysis of skin

temperature and infrared light projection classified the event as (1) true drinking, (2) non-drinking events with minor environmental alcohol contamination, or (3) data invalidated by major environmental alcohol contamination, physical interferences such as clothing, device tampering, device non-wear, or technical malfunction. A column corresponding to the SCRAM CAM was added to SPSS database, in which social days containing a confirmed drinking episode were coded as a SCRAM CAM-registered drinking day, those containing any invalid readings were excluded (Sect. Study design), and all others were coded as SCRAM CAM-registered non-drinking days.

Skyn

Inventory

Cohort #1 used 6 Skyn Generation 2 devices shipped in August 2019. *Cohort #2* used 5 Skyn Generation 2 devices shipped in July 2020 and another 5 shipped in January 2021.

Data collection

The Skyn is worn on the wrist like a watch. Its small compartment (at the back of Figure 3) contains a fuel cell that produces an electric current when exposed to alcohol, similar to the SCRAM CAM but reliant on passive airflow through a thin semipermeable membrane that is recorded every 20 s, rather than active pumping across a metal faceplate. It reports μg of alcohol per L of air.

The Skyn was attached and adjusted for fit in the laboratory by the research staff. The Skyn is not waterproof so participants were asked to remove it for tasks including showering or other activities that would expose it to water. They were asked to maintain and monitor the area of wrist device placement similarly to the SCRAM CAM: clean daily with mild soap, inspect for any adverse reactions, and notify research staff in these instances. At the next office visit, the device was synced to the manufacturer's Skyn App from which we downloaded a csv file containing timestamped TAC readings. Unlike Scramnet, the Skyn App currently does not provide any automated scoring of TAC readings into drinking recognition.

Drinking classification

We established field-based Skyn device rules in *Cohort #1* ($n = 27$ participants, $k = 101$ person-days) and then validated them in *Cohort #2* ($n = 20$ participants, $k = 57$ person-days). The parameters used by Scramnet to define drinking events (peak, absorption, elimination; Sect 2.5.1.3) are hallmarks of distinguishing physiological TAC from confounders (Barnett et al., 2014; Karns-Wright et al., 2018; Roache et al., 2015, 2019). Therefore, we initially proposed to develop a Skyn drinking detection algorithm incorporating these parameters. However, unlike the SCRAM CAM which has a constant baseline at zero, Skyn readings constantly fluctuate even in the absence of TAC. Thus, choosing baseline values upon which to calculate absorption and elimination was equivocal (Appendix S2). We also considered machine learning leveraging the Skyn's higher sampling frequency, however, breath alcohol data to train machine learning (Fairbairn

et al., 2020) was unavailable in the field setting. Nonetheless, day-level analysis was possible from parameters including peak value (PEAKTAC) and area under the curve (AUCTAC), normalized to 24hr in the case of truncated days. PEAKTAC was subsequently excluded because AUCTAC had a higher area under the curve of the ROC (AUCROC) than PEAKTAC and rendered PEAKTAC insignificant in exploratory multiple variable logistic regression models. A threshold AUCTAC value was chosen from the ROC to achieve conditional optimization of sensitivity, the condition being at least ~90% specificity as per literature standards for non-judicial settings where up to 10% false positives are acceptable (Barnett et al., 2014; Karns-Wright et al., 2018; Roache et al., 2019). AUCTAC exceeding this threshold was established as the rule for future prediction of drinking events. Therefore, a column corresponding to the Skyn was added to the SPSS database, in which social days exceeding the chosen AUCTAC threshold were coded as a Skyn-registered drinking day, and all others as a Skyn-registered non-drinking day. Although the number of available days varied by participant (range 1 to 6, median 3, mean 3.4), the rules were not adjusted for within-participant correlations because we sought rules to be transferable across participants.

We then repeated the analysis (conditional optimization of sensitivity) but set the condition to 100% specificity rather than 90%. This condition aligned with the Scramnet algorithm (Sect. SCRAM CAM, *Drinking classification*) which targets 100% specificity as required in judicial contexts.

Finally, we plotted data from the Skyn's temperature sensor in a time series and inspected for abrupt, sustained decreases that may have represented episodes of device removal (Appendix S3).

Blinding

Analysis of *Cohort #1* was not blinded so that we could explore and establish classification rules. Analysis of *Cohort #2* was automated, so blinding was not applicable.

Survey of tolerability

At the end of monitoring on day 15, participants returned to remove the devices and complete a survey about life disruption caused by monitoring activities on a 5-point Likert scale ranging from "strongly disagree" to "strongly agree" that it was a disruption. Device questions were modified from a prior study and focused on discomfort, interference with daily activities, opportunity costs, physical adverse effects, and future willingness to use (Barnett et al., 2017). Diary questions were designed for this study and pertained to ease of completion, ease of remembering to complete, burdensomeness, opportunity costs, and future willingness to complete. Qualitative information was also obtained by open-ended written and verbal questions but not analyzed since quantitative results were conclusive and consistent with our visual inspection of qualitative responses.

Statistical analysis

Descriptive statistics

Demographic and self-reported drinking behaviors were compared between the cohorts by the Mann–Whitney U, chi-squared, or Fisher's exact test as appropriate.

Missing data

Within the scope of the sub-study design (Figure 1), the Skyn missed a very low percentage of the expected three readings per minute ($\leq 1\%$ for 154 days and $\leq 10\%$ for the remaining 4 days) compared to the 20%–30% that experts allow for more established biosensors (Battelino et al., 2019; O'Brien et al., 2013; Troiano et al., 2008). The few missing readings were intrinsically handled in analysis by the trapezoidal linear interpolation of AUCTAC calculations. Days missing SCRAM CAM data (Sect. SCRAM CAM, *Drinking classification*) were excluded from analysis since all but four of these days reflected environmental alcohol or participant non-compliance which likely confounded Skyn data as well.

Alcohol detection accuracy

Using self-report as the reference standard, sensitivity (true positives/[true positives + false negatives]) for each drinking level and specificity (true negatives/[true negatives + false positives]) for drinking versus non-drinking were calculated for each cohort. Their precision was estimated as a 95% confidence interval by the Clopper–Pearson method. Sensitivity and specificity results were judged in comparison to the best prior sensitivity and specificity demonstrated by the SCRAM CAM in the literature. These values were 93% sensitivity to heavy drinking, 73% sensitivity to above-moderate drinking, 65% sensitivity to light drinking, and 90% specificity for drinking versus non-drinking (Barnett et al., 2014; Karns-Wright et al., 2018; Roache et al., 2019).

Exploratory analysis of durability

Since devices were re-used and device performance could be a function of device age, data captured within versus outside 3 months of factory shipping (i.e., “durability”) were visually compared for sensitivity and specificity, as an exploratory analysis when permitted by sample size.

Survey of life disruption/acceptability

We compared the satisfaction survey items across SCRAM CAM, Skyn, and mobile diary, by a non-parametric approach (Friedman

test with post hoc pairwise Wilcoxon Signed Rank Tests) due to observed skewness. Items corresponding to a specific discomfort or interference (e.g., interference with exercise) were tabulated for the proportion answering “agree” or “strongly agree” and compared by McNemar's test. Significance was set at $p < 0.05$.

Sample size

The sample size was determined by consecutive sampling around logistical constraints (section 2.2, paragraph #2).

Software

AUCTAC and PEAKTAC were calculated using R Studio v1.4.1106 (MESS package). Group-level analyses were conducted using SPSS v26 (Chicago, IL) except 95% confidence intervals of proportions were calculated by R Studio (binom.test function).

RESULTS

Demographics

The sample had a slight majority of men and college enrollment, and a large majority of white race and non-Hispanic ethnicity (Table 1). Demographics and drinking behaviors were not statistically different by cohort ($p > .20$).

Self-report

The 158 person-days contained 89 days with any self-reported drinking (56%), of which 69 were days with “above-moderate” drinking (44%) though only 45 of those days were “heavy” drinking days (29%).

Skyn

For cohort #1, the AUCROC was fair at 0.70 which resulted in poor sensitivity for the different drinking levels (39% for any drinking up to 47% for heavy drinking) for the condition of 90% specificity. However, visual inspection of the SPSS database revealed that AUCROC was higher when restricted to datapoints within 3 months of device shipment (0.87) versus ≥ 3 months from shipment (0.79). We, therefore, based AUCTAC thresholds upon the former subset only, resulting in fair to excellent sensitivity for the different drinking levels (69%–94%) for the required specificity (91%) among Cohort #1. Testing these thresholds in Cohort #2 also yielded fair to excellent sensitivity for the different drinking levels (63%–100%) but specificity (70%) was far below the 90% condition (Table 2). We were, however, able to calculate a separate AUCTAC threshold within

TABLE 1 Descriptive statistics

	Cohort #1 (n = 27)	Cohort #2 (n = 20)	Difference	p
Age	21 (20, 23)	21 (20, 22)	U = 267.0	0.95
Female gender	9 (33%)	9 (45%)	$\chi^2 = 0.7$	0.42
Race				
White	21 (78%)	19 (95%)	Fisher's Exact	0.21
Asian	1 (4%)	0 (0%)		
Black	3 (11%)	0 (0%)		
Multiple	1 (4%)	0 (0%)		
Other	1 (4%)	1 (5%)		
Ethnicity (non-Hispanic)	23 (85%)	15 (75%)	$\chi^2 = 0.8$	0.38
Marital status (single)	27 (100%)	20 (100%)	Identical	Identical
Employment status				
Student	16 (59%)	15 (75%)	$\chi^2 = 1.3$	0.26
Employed full-time	5 (19%)	4 (20%)		
Employed part-time	3 (11%)	1 (5%)		
Unemployed	3 (11%)	0 (0%)		
Current college enrollment	16 (59%)	15 (75%)	$\chi^2 = 1.3$	0.26
Drinking during 14 days before baseline				
Average drinks per drinking day	5 (4, 7)	5 (4, 6)	U = 243.5	0.57
% of days with drinking	50 (43, 64)	64 (45, 71)	U = 317.5	0.30
% of days with heavy drinking	29 (29, 43)	36 (29, 48)	U = 287.0	0.70

Note: Given as median (interquartile range) or n (%).

Cohort #2 that had comparable specificity and sensitivity to above-moderate and heavy drinking as Cohort #1 (Table 2). In summary, internal classification thresholds meeting literature standards were derived within both cohorts but they varied widely between the two cohorts (56 vs. 173 h · μ g/L).

We then repeated the analysis with AUCTAC thresholds chosen to optimize sensitivity under the condition of 100% specificity, rather than 90% as in the first analysis. This second analysis produced a similar pattern as the first analysis: the threshold for Cohort #1 (106 h · μ g/L) yielded unacceptable specificity when tested in Cohort #2 (78%), but it was possible to achieve acceptable specificity in Cohort #2 by choosing a higher threshold (374 h · μ g/L) (Table 2). As expected, this second analysis that required higher specificity had lower sensitivity than the first analysis. Each cohort had 1 heavy drinking episode and 2–3 above-moderate drinking episodes that were detected in the first analysis but missed in the second analysis.

Inspecting the temperature time series in Cohort #1 (Appendix S3) revealed 12 potential non-wear episodes totaling 66.2 h (2.8% of total monitored hours). Cohort #2 featured 1 potential non-wear episode totaling 2.3 h (0.2% of total monitored hours).

SCRAM CAM

The SCRAM CAM had 100% specificity for both cohorts (Table 2). In Cohort #1 its sensitivity was poor for all drinking levels (41% for any drinking up to 50% for heavy drinking). In Cohort #2 its sensitivity

was again poor for any drinking (53%), but comparable to literature standards for above-moderate (72%) and heavy drinking (91%).

Survey of tolerability

Declining to wear the SCRAM CAM caused exclusion of 5 otherwise eligible participants (2 pre-randomization, 3 post-randomization; Figure 2) while nobody declined to wear the Skyn. Among the 47 included participants, interference with exercise, sleep, and clothing choice were each noted by a large fraction (32%–51%) for the SCRAM CAM but virtually none (0%–2%) for the Skyn (Table 3). Participants' average ratings for the burden, adverse effects, and aesthetic effects of wearing the biosensors were very low to low for the Skyn and low to moderate for the SCRAM CAM (Table 3). Participants perceived greater opportunity cost from wearing the SCRAM CAM than either wearing the Skyn or completing the mobile diary. Among all three alcohol use monitoring tools, the Skyn was rated better on overall burden and willingness to use in the future than the SCRAM CAM or diary.

DISCUSSION

Main findings

The purpose of this study was to determine the tolerability and accuracy of the Skyn alcohol biosensor in relation to the current standard SCRAM CAM device. We operationalized these criteria

TABLE 2 Accuracy results

	Skyn rules developed on Cohort #1 and tested on Cohort #1	Skyn rules developed on Cohort #1 and tested on Cohort #2	Skyn rules developed on Cohort #2 and tested on Cohort #2
Skyn for condition of 90% specificity ^a			
AUCTAC Cutoff (h · µg/L)	56	56 ^c	173
Sensitivity			
Any drinking	22/32, 69% [50%, 84%]	19/30, 63% [44%, 80%]	14/30, 47% [28%, 66%]
Above moderate drinking	22/28, 78% [59%, 92%]	16/18, 89% [65%, 99%]	14/18, 78% [52%, 94%]
Heavy drinking	17/18, 94% [73%, 100%]	11/11, 100% [72%, 100%]	10/11, 91% [59%, 100%]
Specificity for any drinking	20/22, 91% [71%, 99%]	19/27, 70% [50%, 86%]	24/27, 89% [71%, 98%]
Skyn for condition of 100% specificity ^a			
AUCTAC Cutoff (h · µg/L)	106	106 ^c	374
Sensitivity			
Any drinking	19/32, 59% [41%, 76%]	19/30, 63% [44%, 80%]	12/30, 40% [23%, 59%]
Above moderate drinking	19/28, 78% [48%, 84%]	16/18, 89% [65%, 99%]	12/18, 67% [41%, 87%]
Heavy drinking	16/18, 89% [65%, 99%]	11/11, 100% [72%, 100%]	9/11, 82% [48%, 98%]
Specificity for any drinking	22/22, 100% [85%, 100%]	21/27, 78% [58%, 91%]	27/27, 100% [87%, 100%]
SCRAM CAM rules from Alcohol Monitoring Systems ^b			
Sensitivity			
Any drinking	13/32, 41% [24%, 59%]	16/30, 53% [34%, 72%]	16/30, 53% [34%, 72%] ^d
Above moderate drinking	13/28, 46% [28%, 66%]	13/18, 72% [47%, 90%]	13/18, 72% [47%, 90%] ^d
Heavy drinking	9/18, 50% [26%, 74%]	10/11, 91% [59%, 100%]	10/11, 91% [59%, 100%] ^d
Specificity for any drinking	22/22, 100% [85%, 100%]	27/27, 100% [87%, 100%]	27/27, 100% [87%, 100%] ^d

Cohort #1 contained $k = 101$ person-days over 6 months, but visual inspection revealed reduced accuracy after the first 3 months so only the first $k = 54$ person-days were part of rule development.

Cohort #2 contained $k = 57$ person-days over 6 months.

AUCTAC, area under the transdermal alcohol concentration curve.

Heavy drinking, ≥ 4 standard drinks for women, ≥ 5 standard drinks for men.

Above-moderate drinking, >1 standard drink for women, >2 standard drinks for men.

Any drinking, ≥ 1 standard drink.

Bracketed numbers represent 95% confidence intervals.

^aSkyn rules classify drinking based on exceeding the AUCTAC cutoff developed from the indicated cohort (described in section 2.5.2.3).

^bSCRAM CAM rules classify drinking based on a combination of TAC peak value, absorption slope, and elimination slope (described in Section 2.5.1.3).

^cBoth this column and the previous column represent Skyn rules developed on the same cohort (i.e., Cohort #1) so they have the same AUCTAC cutoff.

^dBoth this column and the previous column represent the SCRAM CAM rules applied to the same cohort (i.e., Cohort #2) so they have the same results.

and then evaluated them among treatment-seeking young adult high-risk drinkers who freely consumed alcohol in their natural environment, since prior studies had already completed validation among social drinkers with standardized drinking timing. The Skyn's tolerability compared extremely favorably overall to the SCRAM CAM device and the alternative option of keeping a mobile daily diary. Furthermore, the Skyn was perceived as less disruptive

than the SCRAM CAM upon specific behavioral variables that are useful to monitor concurrently with drinking and may confound drinking if they are disturbed, such as sleep (Ash et al., 2020b; Fucito et al., 2015, 2017, 2021), exercise (Weinstock et al., 2016), and social comfort (Book & Randall, 2002).

As for Skyn accuracy, we were able to identify a TAC threshold (quantified as AUCTAC) in each cohort that was typically exceeded

TABLE 3 Tolerability Results (N = 47)

	Skyn	SCRAM CAM	Mobile Diary	Test Statistic	p
Life disruptions ^a					
Physical discomfort	2 (4%)	25 (53%)	N/A ^d	McNemar's test	<0.001
Social discomfort	0 (0%)	19 (40%)	N/A ^d	McNemar's test	<0.001
Interference with daily activities					
Work	0 (0%)	5 (11%)	N/A ^d	McNemar's test	0.06
Exercise	1 (2%) ^b	18 (38%)	N/A ^d	McNemar's test	<0.001
Sleep	1 (2%) ^b	15 (32%)	N/A ^d	McNemar's test	0.001
Ability to concentrate	0 (0%)	4 (9%)	N/A ^d	McNemar's test	.13
Clothing choice	0 (0%) ^b	24 (51%)	N/A ^d	McNemar's test	<0.001
Noticeable in any way	6 (13%) ^b	24 (51%)	N/A ^d	McNemar's test	<0.001
Opportunity cost (1–5 scale)	1.8 (1.1) ^b	3.0 (1.4)	1.7 (1.0) ^b	χ^2 (2) = 34.3	<0.001
Overall burden (1–5 scale) ^e	1.6 (0.5) ^{b,c}	2.8 (0.9)	2.7 (0.3)	χ^2 (2) = 59.4	<0.001
Adverse effects (1–10 scale)					
Itching	1.6 (1.4)	2.9 (2.1)	N/A	W = 3.4	0.001
Sweating	1.5 (1.1)	3.0 (2.0)	N/A	W = 4.3	<0.001
Skin Irritation	1.6 (1.5)	3.6 (2.8)	N/A	W = 4.0	<0.001
Marks on skin	7 (15%) ^b	22 (47%)	N/A	McNemar's test	<0.001
Willing to wear in the future (1–5 scale)	4.3 (0.9) ^{b,c}	3.7 (1.2)	3.7 (1.0)	χ^2 (2) = 17.3	<0.001

Note: Values given as mean (SD).

The possible inconvenience of removing the Skyn when interacting with water was not directly queried, but only 4/47 participants (9%) noted water as an inconvenience on free responses.

^aN (%) answering “agree” or “strongly agree”.

^bDifferent than SCRAM CAM ($p < 0.01$).

^cDifferent than diary ($p < 0.01$).

^dThese items were previously (Barnett et al., 2017) designed for tracking devices and often unapplicable to the diaries. For example, a diary would not cause physical discomfort. Therefore, diaries were evaluated by a separate set of questions we designed (ease of completion, ease of remembering to complete, enjoyability, burdensomeness).

^eCalculated as average of 5-point responses to prior rows and diary-specific questions mentioned in previous footnote.

on drinking days but not non-drinking days. The accuracy metrics achieved based upon this classification system were comparable to literature standards for TAC sensors: 91%–94% sensitivity to heavy drinking (literature standard 93%), 78% sensitivity to above-moderate drinking (literature standard 73%), 47%–69% sensitivity to any drinking (literature standard 65%), and 89%–91% specificity for drinking versus non-drinking (literature standard 90%). Even when we set a condition of 100% specificity to simulate judicial system requirements, sensitivity to heavy drinking (the typical level of drinking detectable by judicial system monitors) was still good at 82%–89%. Unfortunately, however, these findings did not translate into robust drinking detection rules because the TAC threshold differed widely between the two cohorts. Specifically, the threshold for cohort #1 was much lower than cohort #2, so application of the cohort #1 threshold to cohort #2 resulted in low specificity (i.e., frequent false positives). Furthermore, even within cohort #1 there was inconsistency that could only be resolved by discarding data collected more than 3 months after device shipment. Further refinement to the Skyn device could troubleshoot this wide variation of TAC thresholds, so that future studies can develop robust rules that can convert Skyn

TAC data to automated detection of drinking versus non-drinking. In addition, we demonstrated that the Skyn's high sampling frequency was robust to field conditions, which could be leveraged for more sophisticated analyses, such as TAC change over time after resolving fluctuation issues (Appendix S2) or machine-learning upon the rich time-series features (Fairbairn & Bosch, 2021; Fairbairn et al., 2020).

Such problems of inconsistent specificity did not arise with the SCRAM CAM. The AMS Scramnet scoring rules yielded 100% specificity in both cohorts. Although, the AMS Scramnet rules had in some cases poor sensitivity (e.g., just 50% to heavy drinking in cohort #1). Future studies applying more sophisticated analyses to Skyn data should also consider more sophisticated analyses that optimize SCRAM CAM sensitivity (Barnett et al., 2014; Roache et al., 2019). An incidental finding was that the SCRAM CAM sensitivity increased over the course of the study (i.e., higher for cohort #2 than cohort #1). We might speculate that the implementation of COVID-19 social distancing requirements (which occurred between cohort #1 and cohort #2) could have reduced potential confounders like physical movement and environmental alcohol at social gatherings. Physical movement could have been further reduced for cohort #2 versus

cohort #1 since habitual physical activity is lower in colder seasons (Belanger et al., 2009). Regardless of its explanation, the improvement of SCRAM CAM accuracy over time reinforces that the deterioration of accuracy over time was unique to the Skyn.

Thus, the present study reveals important findings about the longevity of new generation alcohol sensors. Manufacturers may consider further development before incurring the burden of sophisticated machine-learning analyses and the sample sizes that must complete field-based protocols to have adequately-sized training and testing cohorts (~70–100 participants) (e.g., Fairbairn et al., 2020; Suchting et al., 2019).

Comparison to previous research

Previous researchers issued the Skyn to predominantly non-treatment-seeking, social drinkers in a laboratory (Fairbairn & Kang, 2019; Fairbairn et al., 2020) or research staff members in a naturalistic setting (Wang et al., 2021) and asked them to drink at a standard, preset time. These designs enabled the measurement of the time delay between drinking start and TAC rise from baseline and demonstrated that TAC rise from baseline was automatically and reliably detected by MATLAB Changepoint or machine learning (Fairbairn & Kang, 2019; Fairbairn et al., 2020). However, in these prior studies, there was no need for rules to classify whether TAC rise represented drinking or artifact, since there was a priori knowledge that drinking occurred. One prior study conducted a preliminary inspection of Skyn readings in the field from individuals outside their research team, but encountered many anomalies visually and did not attempt a quantitative system for detecting drinking and inform automated scoring metrics going forward. All of these prior authors have flagged a need for systematic and robust rules that can distinguish whether Skyn TAC rise represents drinking or artifact under field conditions with varying drinking timing and environmental contaminants (Fairbairn & Kang, 2019; Fairbairn et al., 2020; Wang et al., 2021).

Our study responded to this call by testing the Skyn's accuracy to detect drinking by a systematic, objective method, then determining whether this method could lead to automated drinking classification rules going forward. In addition, our population was heavy-drinking young adults who have even wider day-to-day variation and drinking magnitude than the socially-drinking adults of various ages in the prior study (Wang et al., 2021). These objectives were previously pursued extensively for the SCRAM CAM (Barnett et al., 2014; Karns-Wright et al., 2018; Roache et al., 2015, 2019) and researchers derived rules that classify SCRAM CAM TAC departures from baseline as true versus false drinking. We sought to develop a similar classification ability for the Skyn, which was achieved using an AUCTAC threshold. This threshold could distinguish true versus false drinking, but was unstable over device usage and across manufacturing lots.

This study was not designed to assess reasons for such instability, but it is worthwhile to consider that the Skyn's thinner membrane

than the SCRAM CAM may be more susceptible to (1) contaminants that produce acute artifacts, (2) wear and tear over time, and/or (3) physical alterations by the manufacturer between device lots. Acute artifact readings were minimal since the Skyn could accurately classify drinking within the same manufacturing lot used over a short time window (~3 months). However, wear and tear over time is a possibility since devices were worn much longer ($115 \text{ wear hours/participant} \times 52 \text{ participants} \div 16 \text{ devices} = 374 \text{ wear hours/device}$) without direct observation of their handling compared to the largest prior study ($6 \text{ wear hours/participant} \times 110 \text{ participants} \div 5 \text{ devices} = 132 \text{ wear hours/device}$) (Fairbairn & Kang, 2019; Fairbairn et al., 2020). Although the lack of acute artifact readings is consistent with participants generally following our instructions to avoid exposing the Skyn to water and alcohol-containing household products, single exposure events such as applying sunblock may impede the Skyn membrane's permeability and make the sensor ineffective (Wang et al., 2019). Mechanical damage from dropping or hitting it is also possible. Lastly, even with proper handling, the Skyn membrane may wear down naturally and benefit from being replaced periodically. In our data, we see a decline in Skyn performance at 3 months which is the timepoint at which the SCRAM CAM requires users to replace its membrane. Data completeness from the Skyn was much improved from past reports. These reports were missing 30% of the data due to unknown reasons, human researcher errors related to the delicate prototype interface, or batteries dying within the 6-hour monitoring period (Fairbairn & Kang, 2019; Fairbairn et al., 2020). In contrast, the present study had almost no missing data until batteries died after 24–72 h of field use. Nonetheless, battery life requires further improvement before meeting the 72 h as advertised by the manufacturer.

Consistent with a prior study among heavy-drinking young adults (Barnett et al., 2017), our participants reported the SCRAM CAM was sometimes disruptive to physical and social comfort, sleep, exercise, clothing choice, and carried moderate overall opportunity cost and burden. In fact, it was reported as equally burdensome as active self-monitoring with a mobile daily diary. Likewise, in line with a prior study, the SCRAM CAM somewhat commonly caused adverse effects like itching (Barnett et al., 2017). Although prior studies indicate that not all participants experience these problems with the SCRAM CAM (Alessi et al., 2017; Barnett et al., 2017), we took the novel step of directly comparing the Skyn to the SCRAM CAM on these points and found the former was much less likely to cause life disruption in all these areas and adverse effects were much more minor and infrequent. Also, the proportions of our participants experiencing SCRAM CAM disruption to exercise, sleep, and clothing choice were substantial (32%–51%) and concerning when considering that monitoring multiple health behaviors drives engagement in our programs (Ash et al., 2020b; Fucito et al., 2015, 2017, 2021). The Skyn clearly offers potential to eliminate such confounding as it disrupted virtually nobody (0%–2%) in these areas.

Implications for practice

TAC sensors to date have been primarily used for judicial-ordered monitoring, but several research studies in the past decade have explored their use with other drinking populations to motivate and support alcohol self-management. For instance, in three studies, high-risk non-treatment-seeking adult drinkers wore the SCRAM CAM and received contingent financial reinforcement for reducing or abstaining from drinking; the results showed less frequent and severe drinking that persisted in some cases beyond removal of the reinforcements (Barnett et al., 2017; Dougherty et al., 2014). Similarly, we recently demonstrated that offering tracking of alcohol use in conjunction with sleep tracking via biosensors generated strong engagement among non-treatment-seeking heavy-drinking young adults relative to typical uptake rates for alcohol trials with this cohort (Ash et al., 2020b). TAC sensor technology also has the potential to support mobile behavioral strategies such as delivering tailored support messages timed to moments of vulnerability to relapse (i.e., just-in-time adaptive interventions) (Wang & Miller, 2020) and evaluating and refining alcohol treatment interventions based upon resultant long-term drinking changes measured objectively rather than self-report (Barnett et al., 2017; Dougherty et al., 2014). However, these novel intervention strategies necessitate longer-term, consistent uptake of these devices, which our findings suggest may be strong for the Skyn, but only moderate for the SCRAM CAM. Moreover, the SCRAM CAM was more disruptive to health behaviors (i.e., exercise, sleep) and activities of daily living, which limits its potential for tailored and adaptive alcohol interventions and personalized feedback strategies on alcohol-health associations (Ash et al., 2020b; Fucito et al., 2015, 2017, 2021; Weinstock et al., 2016).

Nevertheless, our results in combination with prior research findings suggest that passive alcohol use self-monitoring tools are appealing (Wang et al., 2021), effective at promoting alcohol behavior change (Barnett et al., 2017; Dougherty et al., 2014), and may be preferred over more active self-monitoring methods such as daily diaries. Given the better implementation potential of the Skyn and next-generation TAC sensors, more research is needed on these devices. To this aim, we took the novel step of testing Skyn accuracy to detect drinking at the day level. We developed a mathematical approach that could detect drinking, but unfortunately, the derived rules were not robust over time and across device lots. If this limitation could be addressed by improved hardware and analytics (c.f. Sect 4.1), future research should explore the Skyn for detecting not only the presence but also the timing of drinking to test its momentary relationship to critical events like sleep and driving.

Limitations

Some study limitations should be noted. The Skyn's availability and battery life limited the fraction of days in the parent trial that could be analyzed. However, there is no reason to suspect these factors introduced bias. The Skyn does not yet have a non-wear algorithm

to verify compliance. Unlike the SCRAM CAM the Skyn does not detect infrared light projection onto the skin. However, the good sensitivity to detect drinking when using rules established internal to the cohort suggests that accuracy issues were due to transferability of classification rules rather than non-wear; furthermore, temperature aberrancies suggestive of non-wear were extremely uncommon. Although our sample size was larger than the prior field-based study of the Skyn (Wang et al., 2021) it was still relatively small when divided into training (Cohort #1) and testing (Cohort #2), which could have influenced the robustness of the derived thresholds and confidence interval widths. A larger sample would have also permitted evaluation of within-participant correlations and assess inter-participant variability in classification thresholds, although such findings would further preclude the objective of finding rules transferable across participants. In addition, although demographic factors (sex, age, race, ethnicity, drinking risk level) were ruled out as a confounder of our analysis since they were not statistically different between cohorts, a larger and more diverse sample is needed for better determination of generalizability.

Main conclusions

Skyn feasibility and tolerability results were excellent compared to the SCRAM CAM and an alternative procedure, keeping a mobile drinking diary. These results provide support for Skyn and similar new TAC sensor technology to the behavioral self-management space currently enjoyed by commercially available health/fitness biosensors, if the Skyn could develop accuracy competitive with the SCRAM CAM and mobile drinking diary. Accuracy results were encouraging along these lines in that we derived an approach to validly detect drinking that agreed with mobile diaries to the same or greater extent as previously reported for the SCRAM CAM across a range of drinking levels, but it did not translate into scoring metrics that could be usefully applied across timeframes >3 months among different device lots and individuals. We recommend that improved hardware and analytics resolve these issues, at which time future studies could address time delay from drinking. Most wearable technology studies use carefully maintained devices, but studies like this one can reveal durability concerns and guide next steps in transferring findings to clinical and commercial practice.

ACKNOWLEDGMENTS

The authors would like to thank Kathleen Tackos, Sydney Cannon, Jack LaVigne, Megan Benlock, Haneen Alani, Marisa Rogers, Madielyn Augustine, Allison McNeilly, Natalia Roberts, Anthony Cifre, Momoko Ishii, Jenna Burton, and Asiya Way for their contributions to the implementation of the study. The authors would like to thank Dr. Walter Roberts for expert advice on analysis and interpretation of Skyn data. This work was supported by two grants (R21AA028886, R34AA026021) and one supplement (R34AA026021-01A1S1) from the National Institute on Alcohol Abuse and Alcoholism, a component of the National Institutes of Health (NIH). GA was supported

by a fellowship from the Office of Academic Affiliations at the United States Veterans Health Administration, a Robert E. Leet and Clara Guthrie Patterson Trust Mentored Research Award, Bank of America, N.A., Trustee, and American Heart American Grant #852679 (Garrett Ash, 2021-2024). RG was supported by Yale's Clinical and Translational Science Award UL1TR000142 from the National Center for Advancing Translational Science, also a component of the NIH. The content is solely the responsibility of the authors and does not necessarily represent the official views of the funding sources. The funding sources had no further role in the study design, in the collection, analysis and interpretation of the data, in the writing of the report, or in the decision to submit the paper for publication.

CONFLICT OF INTEREST

The authors have no conflict of interest to declare.

ORCID

Garrett I. Ash  <https://orcid.org/0000-0002-8655-7525>
 Ralitzia Gueorgieva  <https://orcid.org/0000-0003-0944-5973>
 Nancy P. Barnett  <https://orcid.org/0000-0001-9634-8287>
 Wuyi Wang  <https://orcid.org/0000-0001-7179-4416>
 David S. Robledo  <https://orcid.org/0000-0003-1548-1989>
 Kelly S. DeMartini  <https://orcid.org/0000-0003-0943-2507>
 Brian Pittman  <https://orcid.org/0000-0002-0353-5604>
 Nancy S. Redeker  <https://orcid.org/0000-0001-7817-2708>
 Stephanie S. O'Malley  <https://orcid.org/0000-0003-0988-5522>
 Lisa M. Fucito  <https://orcid.org/0000-0001-9410-2830>

REFERENCES

- Alcohol Monitoring Systems. (2021). SCRAM CAM continuous alcohol monitoring [Commercial Web site]. 2022. Available from: <https://www.scramsystems.com/monitoring/scr-am-continuous-alcohol-monitoring/>. [Accessed 10th January 2022].
- 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.
- Ash, G.I., Robledo, D.S., Ishii, M., Pittman, B., DeMartini, K.S., O'Malley, S.S. et al. (2020b) Using web-based social media to recruit heavy-drinking young adults for sleep intervention: Prospective observation. *Journal of Medical Internet Research*, 22(8), e17449.
- Ash, G.I., Stults-Kolehmainen, M., Busa, M.A., Gregory, R., Garber, C.E., Liu, J. et al. (2020a) Establishing a global standard for wearable devices in sport and fitness: perspectives from the new england chapter of the American college of sports medicine members. *Current Sports Medicine Reports*, 19, 45–49.
- Ash, G.I., Stults-Kolehmainen, M., Busa, M.A., Gaffey, A.E., Angeloudis, K., Muniz-Pardos, B. et al. (2021) Establishing a global standard for wearable devices in sport and exercise medicine: perspectives from academic and industry stakeholders. *Sports Medicine (Auckland, N. Z.)*, 51(11), 2237–2250.
- 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, 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, 86–96.
- Battellino, T., Danne, T., Bergenstal, R.M., Amiel, S.A., Beck, R., Biester, T. et al. (2019) Clinical targets for continuous glucose monitoring data interpretation: Recommendations from the international consensus on time in range. *Diabetes Care*, 42, 1593–1603.
- Belanger, M., Gray-Donald, K., O'Loughlin, J., Paradis, G. & Hanley, J. (2009) Influence of weather conditions and season on physical activity in adolescents. *Annals of Epidemiology*, 19(3), 180–186.
- Book, S.W. & Randall, C.L. (2002) Social anxiety disorder and alcohol use. *Alcohol Research & Health*, 26, 130–135.
- Choi, Y.K., Demiris, G., Lin, S.Y., Iribarren, S.J., Landis, C.A., Thompson, H.J. et al. (2018) Smartphone applications to support sleep self-management: Review and evaluation. *Journal of Clinical Sleep Medicine*, 14, 1783–1790.
- Corbin, W.R., Zalewski, S., Leeman, R.F., Toll, B.A., Fucito, L.M. & O'Malley, S.S. (2014) In with the old and out with the new? A comparison of the old and new binge drinking standards. *Alcoholism, Clinical and Experimental Research*, 38, 2657–2663.
- Davis-Martin, R.E., Alessi, S.M. & Boudreaux, E.D. (2021) Alcohol use disorder in the age of technology: A review of wearable biosensors in alcohol use disorder treatment. *Frontiers in Psychiatry*, 12, 642813.
- Demartini, K.S. & Carey, K.B. (2012) Optimizing the use of the AUDIT for alcohol screening in college students. *Psychological Assessment*, 24, 954–963.
- Dougherty, D.M., Hill-Kapturczak, N., Liang, Y., Karns, T.E., Cates, S.E., Lake, S.L. et al. (2014) Use of continuous transdermal alcohol monitoring during a contingency management procedure to reduce excessive alcohol use. *Drug and Alcohol Dependence*, 142, 301–306.
- Dougherty, D.M., Karns, T.E., Mullen, J., Liang, Y., Lake, S.L., Roache, J.D. et al. (2015) Transdermal alcohol concentration data collected during a contingency management program to reduce at-risk drinking. *Drug and Alcohol Dependence*, 148, 77–84.
- 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, 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.
- First, M.B., Williams, J.B.W., Karg, R.S. & Spitzer, R.L. (2014) *Structured clinical interview for DSM-V research version SCID-5-RV*. Washington, DC: American Psychiatric Association Publishing.
- Fucito, L.M., Ash, G.I., DeMartini, K.S., Pittman, B., Barnett, N.P., Li, C.R. et al. (2021) A multimodal mobile sleep intervention for young adults engaged in risky drinking: protocol for a randomized controlled trial. *JMIR Research Protocols*, 10, e26557.
- Fucito, L.M., DeMartini, K.S., Hanrahan, T.H., Whittemore, R., Yaggi, H.K. & Redeker, N.S. (2015) Perceptions of heavy-drinking college students about a sleep and alcohol health intervention. *Behavioral Sleep Medicine*, 13, 395–411.
- Fucito, L.M., DeMartini, K.S., Hanrahan, T.H., Yaggi, H.K., Heffern, C. & Redeker, N.S. (2017) Using sleep interventions to engage and treat heavy-drinking college students: A randomized pilot study. *Alcoholism, Clinical and Experimental Research*, 41, 798–809.
- Goldman, D. (2017) Investing in the growing sleep health economy. In M. Company (Ed.), *Private equity and principal investors practice (Anonymous)* (pp. 1–7). San Francisco, CA: McKinsey & Company.

- Hoepfner, B.B., Schick, M.R., Kelly, L.M., Hoepfner, S.S., Bergman, B. & Kelly, J.F. (2017) There is an app for that - Or is there? A content analysis of publicly available smartphone apps for managing alcohol use. *Journal of Substance Abuse Treatment*, 82, 67–73.
- Hultgren, B.A., Scaglione, N.M., Buben, A. & Turrissi, 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., Dougherty, D.M., Hill-Kapturczak, N., Mathias, C.W. & Roache, J.D. (2018) The correspondence between transdermal alcohol monitoring and daily self-reported alcohol consumption. *Addictive Behaviors*, 85, 147–152.
- 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, 16–22.
- 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, 212–219.
- O'Brien, E., Parati, G., Stergiou, G., Asmar, R., Beilin, L., Bilo, G. et al. (2013) European society of hypertension position paper on ambulatory blood pressure monitoring. *Journal of Hypertension*, 31, 1731–1768.
- Roache, J.D., Karns, T.E., Hill-Kapturczak, N., Mullen, J., Liang, Y., Lamb, R.J. et al. (2015) Using transdermal alcohol monitoring to detect low-level drinking. *Alcoholism, Clinical and Experimental Research*, 39, 1120–1127.
- 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.
- Sim, I. (2019) Mobile devices and health. *New England Journal of Medicine*, 381(10), 956–968. <https://doi.org/10.1056/NEJMr1806949>
- Sobell, L.C., Agrawal, S., Sobell, M.B., Leo, G.I., Young, L.J., Cunningham, J.A. et al. (2003) Comparison of a quick drinking screen with the timeline followback for individuals with alcohol problems. *Journal of Studies on Alcohol*, 64, 858–861.
- Sobell, M.B. & Sobell, L.C. (1993) *Problem drinkers: guided self-change treatment*. New York: Guilford Press.
- Suchting, R., Hébert, E.T., Ma, P., Kendzor, D.E. & Businelle, M.S. (2019) Using elastic net penalized cox proportional hazards regression to identify predictors of imminent smoking lapse. *Nicotine & Tobacco Research*, 21(2), 173–179.
- Troiano, R.P., Berrigan, D., Dodd, K.W., Masse, L.C., Tilert, T. & McDowell, M. (2008) Physical activity in the United States measured by accelerometer. *Medicine and Science in Sports and Exercise*, 40, 181–188.
- U.S. Department of Agriculture and U.S. Department of Health and Human Services. (2020). *Dietary Guidelines for Americans, 2020–2025*, 9th edn. U.S. Department of Agriculture and U.S. Department of Health and Human Services.
- Wang, L. & Miller, L.C. (2020) Just-in-the-moment adaptive interventions (JITAI): A meta-analytical review. *Health Communication*, 35, 1531–1544.
- 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.
- Weinstock, J., Petry, N.M., Pescatello, L.S. & Henderson, C.E. (2016) Sedentary college student drinkers can start exercising and reduce drinking after intervention. *Psychology of Addictive Behaviors*, 30(8), 791–801. <https://doi.org/10.1037/adb0000207>
- Worley, S.L. (2018) The extraordinary importance of sleep: the detrimental effects of inadequate sleep on health and public safety drive an explosion of sleep research. *P T*, 43, 758–763.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: 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, 783–796. Available from: <https://doi.org/10.1111/acer.14804>