Accuracy of transdermal alcohol monitoring devices in a laboratory setting

Authors: Eileen Brobbin[1] MSc, Paolo Deluca[1] PhD, Simon Coulton[2] PhD, and Colin Drummond[1] MD

Institute: [1]Institute of Psychiatry, Psychology & Neuroscience, Addiction Department, 4 Windsor Walk, KCL, London, UK
[2] Centre for Health Service Studies, University of Kent, Canterbury, UK

Corresponding author: Eileen Brobbin, Addiction Science Building, IoPPN, 4 Windsor Walk, Denmark Hill, SE5 8BB, eileen.brobbin@kcl.ac.uk

Ethics reference: HR/DP-20/21-24414

Abstract

Aim: The development of transdermal alcohol sensors (TAS) presents a new method to monitor alcohol consumption with the ability to objectively measure data 24/7. We aimed to evaluate the accuracy of two TAS (BACtrack Skyn and Smart Start BARE) in a laboratory setting.

Method: Thirty-two adults received a dose of ethanol 0.56g/kg body weight as a 20% solution whilst wearing the two TAS and provided Breath Alcohol Concentration (BrAC) measurements, for 3.5 hours post alcohol consumption. Pearson correlations and repeated measures ANOVA tests were conducted on the peak, time to peak and AUC data. Bland Altman plots were derived. A time series analysis and cross-correlations were conducted to adjust for time lag.

Results: Both TAS were able to detect alcohol and increase within 20 minutes. BrAC peaked significantly quicker than Skyn and BARE. BrAC and Skyn peaks were negatively significantly correlated (r=-.381, p=.035, N=31) while Skyn and BARE peaks were positively significantly correlated (r=.380, p=.038, N=30). Repeated measures ANOVA found a significant difference between BrAC, Skyn and BARE ($F(1.946, 852.301) = 459.873$, $p<.001$)). A time-series analysis found when BrAC-Skyn and BrAC-BARE were adjusted for the delay to peak, there was still a significant difference. Failure rates: 1.7% (Skyn) and 4.8% (BARE).

Conclusion: Some evidence for TAS validity as both consistently detected alcohol. Failure rates and time lag show improvements in older device generations. However, neither TAS presented strong equivalence to the breathalyser, even when the lag time was adjusted. With further testing and technology advancements, TAS could be a potential alcohol monitoring tool.

Keywords: alcohol; alcohol monitoring; alcohol treatment; transdermal alcohol sensor; wearable alcohol biosensor

50 word summary:
Two of the newest transdermal alcohol sensor (TAS) devices were worn in laboratory conditions for one afternoon to compare their accuracy of alcohol monitoring, to a breathalyser. Findings suggest that the two TAS (BACtrack Skyn and SmartStart BARE) recorded significantly similar data post-alcohol consumption, but not with the breathalyser.

Introduction

Transdermal alcohol sensor (TAS) devices can be worn to measure transdermal alcohol concentration (TAC) excreted through a sensor against the skin. This sensor analyses the concentration of alcohol in the skin vapours emitted and calculates the TAC (Dumett *et al.*, 2008). Current methods used to measure alcohol consumption within clinical settings include breathalysers, blood tests, or self-report measures such as the Alcohol Use Disorder Identification Test-Consumption (AUDIT-C) (Saunders *et al.*, 1993) or Timeline Follow Back (Sobell and Sobell, 1992; Kuntsche *et al.*, 2022). However, all these methods have limitations in the accuracy, amount of detail collected, or period over which they can collect data. They can also be costly and burdensome to use (Kuntsche *et al.*, 2022). TAS could address some of these limitations, as they can be worn over extended periods while providing continuous measurement (Litten, Bradley and Moss, 2010; Leffingwell *et al.*, 2013). This means TAS could capture data when self-report is not possible (for example, when heavy drinking leads to lack of capacity and blackouts) or when it is not possible for multiple daily breathalysers, blood, or urine tests (Kuntsche *et al.*, 2022). A breathalyser can only capture alcohol consumption which has occurred within a short window before breathalyser administration (Kuntsche *et al.*, 2022). However, as this technology is still relatively new, TAS brands validity and reliability still require investigation. There are studies published on TAS accuracy or acceptability (Caluzzi *et al.*, 2019; Fairbairn and Kang, 2019; Croff *et al.*, 2020; van Egmond *et al.*, 2020; Brobbin *et al.*, 2022b, 2022a; Courtney, Russell and Conroy, 2022), but most of these are limited by only testing one brand of TAS and each study uses a slightly different design, population, or analysis, making it difficult to draw direct comparisons.

There are various TAS brands available, and all are slightly different in their technology, appearance, wear, or output reporting. SCRAM is the most validated TAS (Sakai *et al.*, 2006; Dougherty *et al.*, 2012; Barnett, Meade and Glynn, 2014; Hill-Kapturczak *et al.*, 2014; Barnett *et al.*, 2017; Alessi, Barnett and Petry, 2019; Fairbairn and Kang, 2019; Roache *et al.*, 2019; van Egmond *et al.*, 2020; Wang *et al.*, 2021; Ash *et al.*, 2022) and has been available in the US criminal justice system for the past 20 years, and in the UK justice system since 2020. SCRAM has various TAS models but is usually worn on the ankle and is the largest device in size. SCRAM uses a home modem to send data to the server and monitors alcohol consumption every 30 minutes. The ION Wearable TAS is worn on the wrist and is available for purchase on their website (*ION Wearable*, no date). They use non-invasive enzymatic alcohol sensors with replaceable cartridges. BACtrack Skyn is the latest developed, available, TAS using fuel-cell technology and is available for research purposes and beta testing. It is worn on the wrist and can sync with Apple iOS devices to provide data on an app and a website server. If the Skyn is being worn close to the paired Apple device, real-time syncing and viewing of alcohol consumption is possible. The latest version of the Skyn App also has the option to estimate a range of BAC (Blood Alcohol Concentration) based on the TAC data. However, this data conversion is currently only usable in personal Skyn use and not in a research context. WrisTAS was another wrist-worn TAS that has been widely validated (Swift *et al.*, 1992; Webster and Gabler, 2008; Marques and McKnight, 2009; Bond *et al.*, 2014; Luczak, Rosen and Wall, 2015; Simons *et al.*, 2015; Croff *et al.*, 2020) but has since been bought by Smart Start. Other TAS are in various stages of development as evidenced in previous literature reviews (van Egmond *et al.*, 2020; Brobbin *et al.*, 2022a, 2022b). The BARE is a newer prototype developed by Smart Start and is not currently available to purchase. The BARE manufacturers state the target audience is clinical use, unlike the other

brands which are targeted for general consumer use. More details about the two TAS used in this study (BACtrack Skyn and Smart Start BARE) are included in Supplementary Information.

When considering implementing TAS within clinical settings, their use would not only be detecting alcohol presence. Recording when, the amount, peaks of use and the accuracy are important. They could be used to discern between low, moderate, or high levels of drinking (Dougherty *et al.*, 2012; Hill-Kapturczak *et al.*, 2014). TAS could become a valuable tool for alcohol treatment to allow for continuous alcohol monitoring TAS brands, with the option to view output in near-real time, could provide alcohol services with the opportunity to initiate timely and personalised interventions, for patients going through alcohol treatment, and afterwards, when there is a high chance of relapse (Moos and Moos, 2006). TAS could also benefit alcohol research, helping researchers continuously measure alcohol consumption and potentially implement Contingency Management (CM) for alcohol reduction in near-real time (Barnett *et al.*, 2011, 2017; Dougherty *et al.*, 2014, 2015; Averill *et al.*, 2018; Richards *et al.*, 2023). While only the SCRAM can be locked and considered tamper-proof, the other brands have a temperature sensor to monitor removals but are not as secure as SCRAM. The use of a removable device to deliver CM will need further exploration and potential penalties for dealing with device removals, data loss, and wearers possibly turning the device off to obscure data collection.

Earlier studies have shown a general agreement between TAC and BrAC (Breath Alcohol Concentration) or BAC values using SCRAM (Dougherty *et al.*, 2012; Barnett, Meade and Glynn, 2014; Hill-Kapturczak *et al.*, 2014; Fairbairn and Kang, 2019), and Skyn (Fairbairn and Kang, 2019; Wang *et al.*, 2019, 2021; Fairbairn, Kang and Bosch, 2020; Ash *et al.*, 2022). However, other studies have shown varying levels of disagreement (Sakai *et al.*, 2006; Marques and McKnight, 2009; Croff *et al.*, 2020). A large majority of these studies have been conducted using SCRAM or WrisTAS (Leffingwell *et al.*, 2013; Greenfield, Bond and Kerr, 2014; Barnett, 2015; Dougherty *et al.*, 2015; Fairbairn *et al.*, 2019; Wang *et al.*, 2019; van Egmond *et al.*, 2020). There has been a smaller number of accuracy studies using Skyn (Fairbairn and Kang, 2019; Wang *et al.*, 2019, 2021; Fairbairn, Kang and Bosch, 2020; Rosenberg *et al.*, 2021; Ariss, Fairbairn and Bosch, 2022; Ash *et al.*, 2022; Richards *et al.*, 2022). The few studies conducted using Skyn found an overall high correlation between self-report, BrAC and Skyn TAC, that Skyn could distinguish between low and high alcohol doses and was considered acceptable and feasible to wear (Fairbairn and Kang, 2019; Fairbairn, Kang and Bosch, 2020; Rosenberg *et al.*, 2021; Wang *et al.*, 2021; Ash *et al.*, 2022; Courtney, Russell and Conroy, 2022; Richards *et al.*, 2022).

This study adds to this research work with the latest version of Skyn compared to a newer wrist sensor (BARE). This study aims to investigate the accuracy of these two TAS to measure alcohol concentration within a laboratory setting for a population of alcohol users not meeting the criteria of alcohol dependence (AUDIT<16). It will also be the first to test the accuracy of BARE. As all three devices (breathalyser, Skyn, BARE) aim to measure alcohol consumption we expect values from all three devices to be correlated. We predict that the Skyn and BARE will be correlated in terms of both peak data and time-to-peak readings.

Material and Methods

Recruitment

Participants were recruited through a fortnightly research email at King's College London, which included the study advertisement. Those interested were asked to email the researcher. To be eligible, participants had to meet the inclusion criteria: 1) 18 years old and over, 2) regular drinker (five or more standard UK units for men, four or more standard UK units for women, consumed at least twice per month), 3) no desire to reduce their drinking, 4) AUDIT score <16, 5) willing to provide informed consent to participate in the study. Potential participants were excluded if they had: 1) drug or alcohol-related health problems, 2) a health diagnosis potentially complicated by alcohol ingestion, 3) illicit drug use (excluding cannabis) within the last 4 weeks, 4) a history of drug or alcohol treatment, 5) were prescribed contraindicated medication, 6) pregnant, 7) presence of Covid-19 symptoms.

These inclusion and exclusion criteria were used to target individuals who regularly consumed the amount of alcohol that was to be provided to them as part of the study. We did not want to provide alcohol to those who did not regularly consume this amount, had a higher risk of alcohol dependence by scoring >16 on the AUDIT, or had any health problems complicated by alcohol consumption.

Alcohol measures

AUDIT – Used to ensure participants scored below the indication of harmful alcohol use or dependence and moderate-severe alcohol use disorder (Saunders *et al.*, 1993).
Breathalyser – Lion Alcometer SD-400. Lion fuel cell sensor (standard version) Measure range: 0.02 – 2.00mg/l BrAC, operating range: -5 - +45 Celsius, with optimum operation between +10 - +40 Celsius. Calibrated one week before recruitment started.
Transdermal alcohol sensors – BACtrack Skyn (model T15/2021) and Smart Start BARE (prototype). Skyn provided data output every 20 seconds and included motion and temperature (Celsius) sensors. Skyn output can be displayed in 20-second, 1-minute, 5-minute or 30-minute intervals. We viewed this data at 1-minute intervals to best match the breathalyser and BARE readings. BARE only provided data output every two minutes (average of the data recorded every 20 seconds) and recorded temperature (Celsius) data. For the analysis, we used these two variables (alcohol signal mg/dl) and temperature but BARE also had columns for sample count, Prox status, Prox (light), and descriptive alcohol data (event peak, peak level, data slope, area under the curve (AUC)).

Procedure

Participants completed an AUDIT and the Nuffield Medical health questionnaire to ensure they met inclusion criteria and were asked not to consume alcohol for 24 hours prior and to eat a light lunch one hour before arrival. All procedures took place in the afternoon. On arrival, a breathalyser measurement was taken to confirm a reading of zero. Participants were weighed and their height was measured. The participant's weight determined the alcohol dose. All participants received 0.56g/kg of 37.5% ABV vodka as a 20% solution mixed with orange fruit juice, with this drink being consumed within 10 minutes. For example, a participant weighing 80kg would have a dose of 44.8g ethanol (5.5 UK units).

Automatic measurements were recorded by the TAS. Breathalyser measurements were taken every 15 minutes, starting after alcohol consumption for a total of 3.5 hours,

therefore there were 15 measurements taken. Participants were thanked for their time and provided with a snack and a £10 voucher. All participants were told they would have to remain until their BrAC was below a safe level of 0.25mg/l, below the UK drink drive limit (0.35mg/l), even if this was longer than the set 3.5 hours. Due to the amount of alcohol being consumed by each participant, it was not feasible for participants to stay until their BrAC values had reduced to zero. This study was approved by the King's College London ethics committee (reference: HR/DP-20/21-24414).

Data manipulation
Missing values were calculated using linear interpolation in Excel and any negative values were left-censored to zero, as conducted in previous research (Courtney, Russell and Conroy, 2022). We standardised the values by converting all device values to the same unit, mg/l. Breathalyser values were in mg/l units. The BARE values were converted from mg/dl to mg/l by multiplying the numerator by 10, and the Skyn values were converted from ug/L to mg/l by dividing by 1000. As the Skyn provided values for each minute we were able to use the minute that matched with the breathalyser. The BARE provided values every two minutes and so when the value did not match exactly with the same minute as the breathalyser the minute before was used for analysis.

Data analysis
A statistical power calculation was used to determine the sample size, which was 25 participants (alpha of 0.05, power of 0.90, estimated effect size of 0.60). Sociodemographic and descriptive device data were recorded and plotted. All device values were standardised. We reported on the failure rate for all devices and the peak, time to peak and AUC for each participant.

To test if the two TAS peak and time-to-peak data are correlated. Data was analysed at each time point (15-minute intervals) from alcohol consumption, for a total of 3.5 hours (15 timepoints). Pearson correlations were used for peak data and device timepoints, and paired t-tests were used for time to peak. We define the peak data as the highest value recorded for each participant by each device. That means, for the same participant the three devices (BrAC, Skyn and BARE) could have time of peak data at different times.

After a Shapiro-Wilks test of normality, repeated measures ANOVA tests were conducted to compare the Breathalyser and Skyn, Breathalyser and BARE, and Skyn and BARE. Mauchly's Test of Sphericity was conducted.

The AUC was calculated for each device from the proportion of cells within a matrix below the curve. The AUC was calculated to investigate if there was a difference in the area under the curve for each device. Pearson correlations for AUC and alcohol administered were conducted and correlation AUC between BrAC, Skyn and BARE. Bland-Altman plots (Bland and Altman, 1986) were used to compare devices for time to peak and peak alcohol concentration if possible.

A time series analysis and cross-correlation were conducted to adjust timepoints for any time lag in measurement between the devices. The timepoints were then adjusted for the found lag. The Shapiro-Wilks test of normality, repeated measures ANOVA, and Mauchly's

test of Sphericity were conducted with these adjusted data. The AUC was recalculated for this adjusted data. Pearson correlations with AUC were conducted with this time-adjusted dataset. To note, the AUC is incomplete as participant measurements were not taken until BrAC and TAS values returned to zero. All the analysis was conducted with IBM SPSS Statistics version 28.

Results

Baseline and descriptive device statistics

There was a total of 32 healthy adult participants (15 males, 17 females). Sample characteristics and descriptive standardised device data are reported in Table 1. Data recorded from each participant included their Breath Alcohol Concentration (BrAC) breathalyser, Skyn, and BARE data (mg/l). The average peak BrAC was 0.40 mg/l (SD: 0.15 mg/l) (0.083g/210L), the average peak Skyn was 0.10 mg/l (SD: 0.04 mg/l and the average peak BARE was 8.01 mg/l (SD: 6.08 mg/l).

Below, we have provided a visualization of the average data from the 3.5 hours (Figure 1).

Table 1. Sample characteristics.

Participant characteristics Mean (Standard Deviation)	All (N=32)	Males (n=15)	Females (n=17)	Sex difference
Age in years	25.84 (6.14)	26.53 (7.53) Range: 18-47	25.24 (4.76) Range: 18-39	p=.304
BMI	22.23 (3.31)	24.14 (3.28) Range: 20-31.5	20.55 (2.33) Range: 14.9-24.3	p=.330
Height in centimetres	172 (7.92)	177.32 (6.66) Range: 167.5-187.0	167.62 (6.03) Range: 155-175	p=.478
Weight in kilograms	66.28 (12.53)	75.56 (8.96) Range: 59.3-94.4	58.09 (9.05) Range: 40.7-78.0	p=.823
AUDIT scores	6.53 (3.12)	7.27 (3.33) Range: 2-12	5.88 (2.87) Range: 3-15	p=.681
Race/ethnicity n (%)				$X^2(3)=3.599$, p=.308
Caucasian	23 (72%)	12 (79%)	11 (65%)	
Asian	6 (19%)	1 (7%)	5 (29%)	
Black	2 (6%)	1 (7%)	1 (6%)	
Hispanic	1 (3%)	1 (7%)	0	
Alcohol data Mean (SD)	**All (N=32)**	**Males (n=15)**	**Females (n=17)**	**Sex difference**
Standardised data				
Peak BrAC mg/l	0.40 (0.15)	0.37 (.11)	0.43 (.18)	p=.162
Peak Skyn mg/l	0.10 (0.04)	0.10 (0.04)	0.10 (0.03)	p=.489
Peak BARE alcohol signal mg/l	799.35 (608.02)	570.00 (386.42)	908.24 (748.18)	p=.925
BrAC time to peak (minutes)	53.44 (32.98)	51.00 (29.89)	55.59 (36.27)	p=.417
Skyn time to peak (minutes)	120.32 (39.93)	125.67 (38.25)	115.31 (42.05)	p=.618
BARE time to peak (minutes)	140.97 (45.65)	139.14 (40.19)	142.47 (50.88)	p=.638

Significant p<0.05
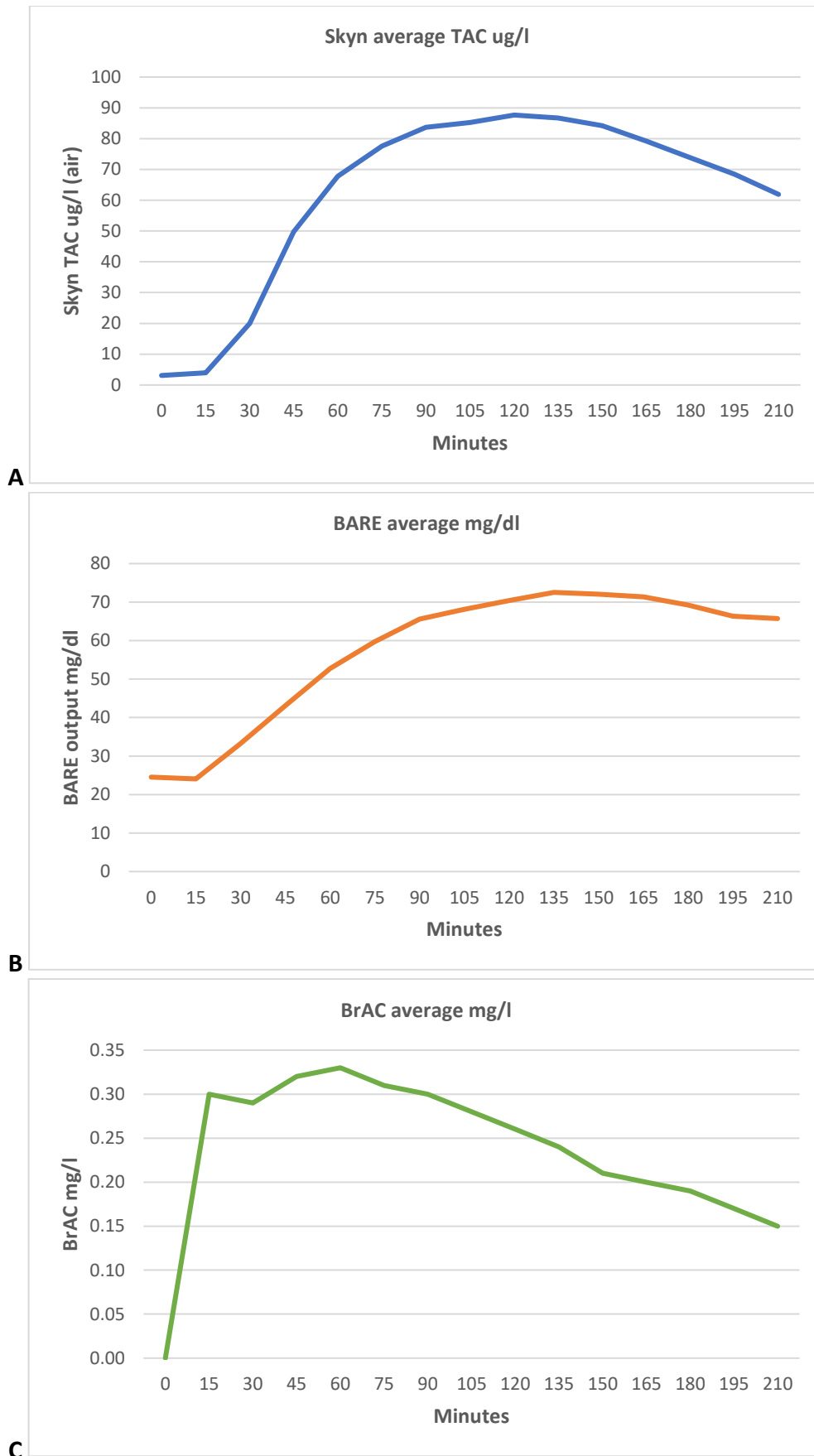AUDIT = Alcohol Use Disorder Identification Test
BrAC = Breath Alcohol Concentration
BMI = Body Mass Index

Figure 1. Average Skyn, BARE and BrAC data for all participants.

Figure 1 A, B and C: A) Average Skyn output every 15 minutes for 3.5 hours (ug/l air), **B)** BARE output every 15 minutes for 3.5 hours (mg/dl). **C)** Average BrAC measured every 15 minutes for 3.5 hours (mg/l).

BrAC = Breath Alcohol Concentration
TAC = Transdermal Alcohol Concentration

Failure rate

In this context, failure rate means missing data, not reported in device output. The Skyn output was reported in 1-minute intervals, so the failure rate is the number of 1-minute intervals that were not reported. Skyn failure rate was 1.7% for all timepoint measures (for 31 participants, 108 out of 6466 timepoints were missing and not reported in Skyn output). One participant (P27) failed to record usable data due to human error, we excluded this participant from this calculation. The BARE output was reported in 2-minute intervals, so the failure rate is the number of 2-minute intervals that were not reported. The BARE failed to record one participant (P30) for an unknown reason, possibly a device error. Including this participant, BARE had a failure rate of 4.8% (for 32 participants, 161 out of 3357 timepoints were missing and not reported in BARE output). Both TAS recorded a relatively stable baseline TAC at the start of the experiment, with a clear rise following alcohol consumption, a peak, and the start of a fall. All breathalyser measurements were successfully recorded.

BrAC and TAC data

The amount of ethanol (grams) each participant consumed, calculated by their weight, the peak, time to peak and the AUC from each device are included in Table 2.

Table 2. Participant information: ethanol consumed, weight, device peak, time to peak data and AUC.

Participant ID	Alcohol consumed	Weight	BrAC g/210L	Peak			Time to peak (in minutes)			AUC		
				BrAC	Skyn	BARE	BrAC	Skyn	BARE	BrAC	Skyn	BARE
1	39.70	70.90	0.081	0.39	0.16	830.00	15	112	150	3.56	1643.59	8110.00
2	34.27	61.20	0.056	0.27	0.12	500.00	45	146	152	2.55	1271.70	5285.00
3	49.62	88.60	0.077	0.37	0.12	3500.00	45	138	86	3.10	1353.01	35595.00
4	38.70	69.10	0.075	0.36	0.07	260.00	15	110	84	2.61	719.15	2790.00
5	33.21	59.30	0.113	0.54	0.05	870.00	15	190	132	3.03	338.12	10295.00
6	36.23	64.70	0.100	0.48	0.10	770.00	75	89	62	4.62	1030.39	8080.00
7	29.29	52.30	0.113	0.54	0.11	630.00	15	151	110	4.42	1112.55	6760.00
8	44.74	79.90	0.083	0.40	0.16	1870.00	60	185	74	3.72	1527.90	17665.00
9	43.68	78.00	0.079	0.38	0.12	730.00	60	132	140	2.56	630.58	3660.00
10	37.46	66.90	0.067	0.32	0.10	450.00	75	100	142	2.82	763.35	4005.00
11	34.66	61.90	0.081	0.39	0.10	290.00	90	147	126	3.75	912.21	3055.00
12	42.06	75.10	0.067	0.32	0.07	1140.00	15	135	130	2.47	546.38	10360.00
13	42.73	76.30	0.060	0.29	0.17	1690.00	45	97	126	2.91	1784.28	18285.00
14	33.32	59.50	0.060	0.29	0.12	840.00	135	147	198	2.88	1209.64	7480.00
15	28.00	50.00	0.092	0.44	0.09	550.00	30	134	126	3.44	863.14	5950.00
16	30.97	55.30	0.054	0.26	0.12	700.00	75	72	146	2.64	1029.61	6865.00
17	43.68	78.00	0.067	0.32	0.14	780.00	45	59	178	3.17	1326.60	7040.00
18	29.74	53.10	0.077	0.37	0.13	690.00	15	90	166	3.29	1194.46	7210.00

9

19	29.12	52.00	0.054	0.26	0.13	430.00	45	91	134	2.43	1181.52	4110.00
20	37.24	67.10	0.133	0.64	0.05	520.00	45	94	156	4.96	321.44	6285.00
21	31.92	54.70	0.196	0.94	0.03	500.00	15	11	208	5.12	247.77	4830.00
22	46.87	83.70	0.079	0.38	0.09	800.00	60	144	124	3.84	684.44	9510.00
23	37.24	66.50	0.092	0.44	0.08	530.00	60	172	182	4.55	558.56	4650.00
24	31.92	57.00	0.094	0.45	0.07	790.00	45	103	206	4.11	576.67	6605.00
25	26.43	47.20	0.138	0.66	0.11	520.00	30	72	6	5.49	985.37	5690.00
26	22.79	40.70	0.054	0.26	0.08	590.00	75	137	140	2.97	781.59	6290.00
27	39.20	70.00	0.111	0.53	-	750.00	15	-	146	4.68	-	8445.00
28	35.50	63.40	0.065	0.31	0.09	660.00	120	151	174	3.54	814.25	5260.00
29	42.28	75.50	0.046	0.22	0.07	820.00	75	106	164	2.02	602.49	10175.00
30	52.86	94.40	0.060	0.29	0.07	-	105	88	-	3.03	370.75	-
31	39.45	70.50	0.090	0.43	0.07	530.00	45	174	192	3.74	691.57	4295.00
32	43.74	78.10	0.063	0.30	0.04	310.00	105	153	210	3.15	379.80	2725.00

Table 2. Alcohol consumed (grams), Weight (kg), peak BrAC (g/210L), peak BrAC (mg/l), peak Skyn (mg/l), peak BARE (mg/l), time to peak (minutes) and AUC. A dash (-) marks missing data for that participant and device output.
AUC = Area Under the Curve
BrAC = Breath Alcohol Concentration

The data were analysed at 15-minute intervals to match breathalyser measurements, referred to as timepoints. There were a total of 15 timepoints over the 3.5-hour study length: 0 minutes (post alcohol consumption) = timepoint 1, 15 minutes = timepoint 2, 30 minutes = timepoint 3, 45 minutes = timepoint 4, 60 minutes = timepoint 5, 75 minutes = timepoint 6, 90 minutes = timepoint 7, 105 minutes = timepoint 8, 120 minutes = timepoint 9, 135 minutes = timepoint 10, 150 minutes = timepoint 11, 165 minutes = timepoint 12, 180 minutes = timepoint 13, 195 minutes = timepoint 14, 210 (3.5 hours) = timepoint 15.

Pearson correlations were conducted between BrAC and Skyn, BrAC and BARE and Skyn and BARE, as reported in Table 3. Only one significant positive correlation was found between BrAC and Skyn timepoints, at timepoint 2, 15 minutes post alcohol consumption. No significant Pearson correlations were found between BrAC and BARE. Pearson correlations between Skyn and BARE were conducted with significant positive correlations found at timepoints 5, 6, 10, 11, 12 and 13 (minutes: 60, 75, 135, 150, 165, 180). Pearson correlations found peak BrAC and peak Skyn were negatively correlated, while peak Skyn and peak BARE were positively correlated.

The TAS appeared to detect each alcohol event for each participant. TAC output data increased within 20 minutes of alcohol consumption. The average time to peak value for Skyn was 120.32 minutes (SD=39.93) and for BARE it was 140.97 minutes (SD=45.64), compared to the breathalyser at 53.44 minutes (SD=32.98) (Skyn lagged by 66.88 minutes and BARE by 87.53 minutes). The difference between Skyn and BARE time to peak was non-significant (t(29) = -1.723, p = .096). The time to peak between BrAC and Skyn and BrAC and BARE were significantly different (t(30) = -7.801, $p < .001$, t(30) = -10.202, $p < .001$, respectively). BARE peak was significantly correlated to Body Mass Index (BMI) (r = .615, $p <$ .001, N = 31).

Shapiro Wilks test found that data output was normally distributed for BrAC, Skyn and BARE. A repeated measures ANOVA was conducted. Mauchly's Test of Sphericity indicated that the

assumption of sphericity had been violated, and therefore a Greenhouse Geisser correction was used. The difference between the devices was statistically significant (Table 4).

Table 3. Pearson correlations.

Device correlations	Timepoint 2 (15 minutes)	Timepoint 3 (30 minutes)	Timepoint 4 (45 minutes)	Timepoint 5 (60 minutes)	Timepoint 6 (75 minutes)	Timepoint 7 (90 minutes)	Timepoint 8 (105 minutes)	Timepoint 9 (120 minutes)	Timepoint 10 (135 minutes)	Timepoint 11 (150 minutes)	Timepoint 12 (165 minutes)	Timepoint 13 (180 minutes)	Timepoint 14 (195 minutes)	Timepoint 15 (210 minutes)	Peak TAC
BrAC - Skyn	*r = .378, p = .033, N = 32**	r = .0.94, p = .609, N = 32	r = .082, p = .654, N = 32	r = .036, p = .845, N = 32	r = -.110, p = .549, N = 32	r = -.196, p = .283, N = 32	r = -.123, p = .502, N = 32	r = -.308, p = .086, N = 32	r = -.231, p = .203, N = 32	r = -.103, p = .582, N = 31	r = -.066, p = .724, N = 31	r = -.044, p = .814, N = 31	r = .073, p = .698, N = 31	r = -.153, p = .411, N = 31	*r = -.381, p = .035, N = 31**
BrAC - BARE	r = .196, p = .290, N = 31	r = .002, p = .992, N = 31	R = .074, p = .692, N = 31	r = .058, p = .758, N = 31	r = -.078, p = .676, N = 31	r = .004, p = .984, N = 31	r = -.174, p = .350, N = 31	r = -.234, p = .205, N = 31	r = -.244, p = .186, N = 31	r = -.182, p = .336, N = 30	r = -.269, p = .150, N = 30	r = -.208, p = .270, N = 30	r = -.173, p = .361, N = 30	r = -.097, p = .609, N = 30	r = -.113, p = .544, N = 31
Skyn - BARE	r = .038, p = .841, N = 31	r = .137, p = .461, N = 31	r = .300, p = .101, N = 31	*r = .437, p = .014, N = 31**	*r = .385, p = .032, N = 31**	r = .286, p = .119, N = 31	r = .273, p = .137, N = 31	r = .347, p = .056, N = 31	*r = .384, p = .033, N = 31**	*r = .381, p = .038, N = 30**	*r = .384, p = .036, N = 30**	*r = .389, p = .034, N = 30**	r = .372, p = .078, N = 30	r = .294, p = .115, N = 30	*r = .380, p = .038, N = 30**

Table 4. Repeated measures ANOVA.

	Shapiro Wilks Test			Mauchly's Test of Sphericity	Greenhouse Geisser
	BrAC	**Skyn**	**BARE**		
Results	.953 df=442, p<.001*	.949, df=442, p<.001*	.690, df=442, <.001*	$X^2 (2) = 6878.795$, p = .000*	$F(1.000, 441.000) = 574.820$, p <.001*

*Significant p<0.05
BrAC = Breath Alcohol Concentration
TAC = Transdermal Alcohol Concentration

AUC

The AUC was calculated for each participant for BrAC, Skyn and BARE data. The AUC was calculated using the start of the drinking event to the endpoint of the last measurement 3.5 hours later. Pearson correlations of AUC and the amount of alcohol administered (in grams) were conducted. Skyn AUC and BARE AUC were correlated ($r = .439$, $p = .015$, $N = 30$), BrAC AUC was not significantly correlated with Skyn AUC ($r = -.189$, $p = .317$, $N = 30$) or BARE AUC ($r = -.139$, $p = .480$, $N = 30$).

Analysing AUC up to the peak, there were no significant correlations between any of the devices.

Bland Altman Plot

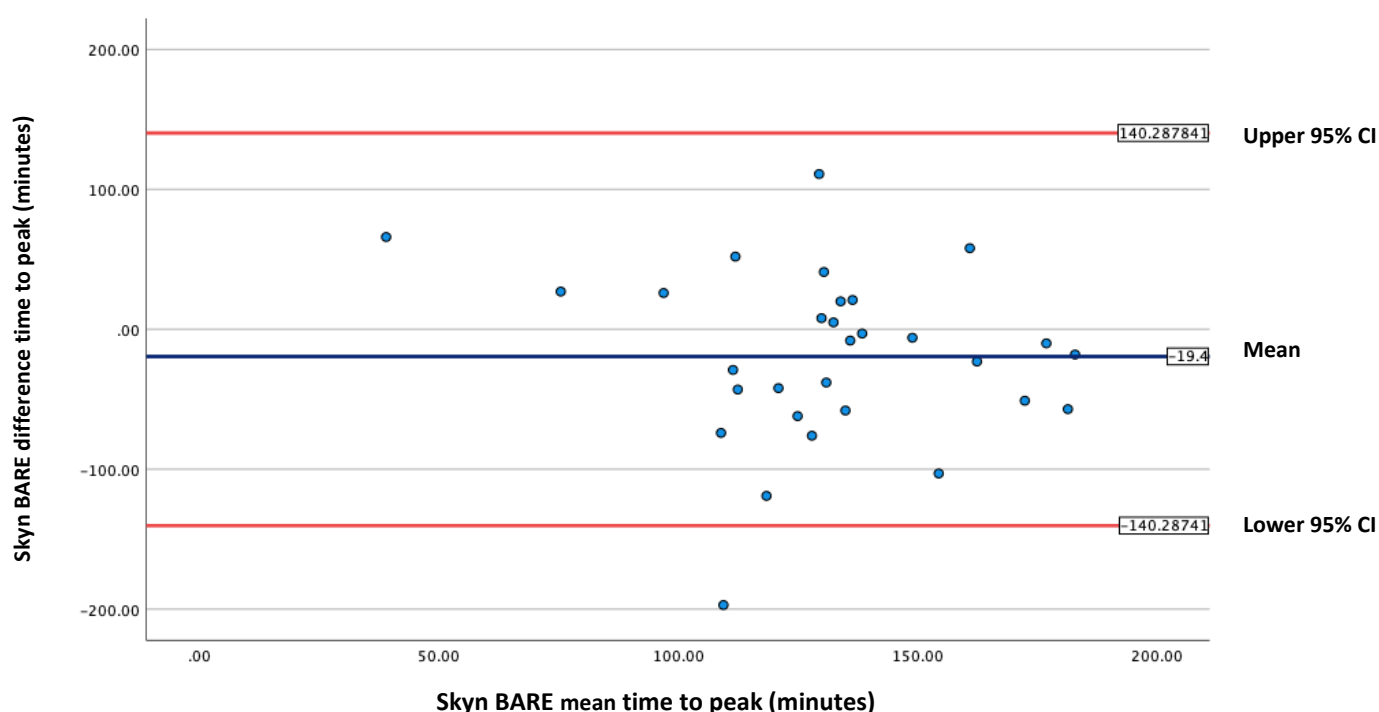**Figure 2. Bland Altman plot for time to peak Skyn and BARE.**

Figure 2. BA plot for time to peak (minutes) Skyn and BARE devices. Plot of differences between Skyn and BARE (y-axis) versus the mean of the Skyn and BARE (x-axis). The middle blue line represents the mean difference between the TAS. The top red line represents the upper 95% CI and the bottom red line represents the lower 95% CI.

We created a Bland Altman (BA) plot (Bland and Altman, 1986) for the time to peak for Skyn and BARE (Figure 2) to visually compare the TAS. For each participant, we averaged the time to peak for Skyn and BARE results (x-axis) and calculated the difference between the time to peak for Skyn and BARE results (y-axis). This BA plot provides a visual representation of the agreement between Skyn and BARE for the time-to-peak data. The mean is -19.400, the upper 95% CI is 101.488 and the lower 95% CI is -140.288.

If the two TAS were in perfect agreement, we would expect the blue circles to be close to the middle horizontal line with an intercept of zero. The BA plot below (Figure 2) shows varying disagreement between the time to peak for Skyn and the time to peak for BARE.

Time series analysis

We completed a time series analysis and cross-correlation to determine if there was a time lag for the TAS. We found that whilst Skyn and BARE data were in sync, there was a lag of two timepoints (30 minutes) for Skyn behind BrAC and a lag of four timepoints (60 minutes) for BARE behind BrAC. Therefore, we adjusted these data by these respective timepoints to account for the time lag. We then conducted the repeated measures ANOVA and AUC with these adjusted data.

Shapiro Wilks test found that these data were normally distributed. A repeated measures ANOVA was conducted. Mauchly's Test of Sphericity indicated that the assumption of sphericity had been violated X^2 (2) = 5070.296, p = .000, and therefore a Greenhouse Geisser correction was used. The difference between the devices was statistically significant: F(1.00, 330.000) = 518.675, p <.001. Correlations with adjusted data found that lagged data from Skyn and BARE were positively correlated (r = .113, *p* = .040, N = 331) but still not significantly correlated to BrAC. There was a significant correlation between AUC Skyn and AUC BARE (*r* = .450, *p* = .013, *N* = 30). There was no significant correlation between AUC BrAC and AUC Skyn (*r* = -.177, *p* = .341, *N* = 31) or AUC BARE (*r* = -.106, *p* = .570, *N* = 31).

Analysing AUC up to the peak, AUC Skyn and AUC BARE were significantly positively correlated (*r* = .364, *p* = .044, *N* = 31).

Discussion

This study aimed to test the accuracy of two TAS devices, compared to a breathalyser and each other. As all three devices (breathalyser, Skyn, BARE) aim to measure alcohol consumption we expected values from all three devices to be correlated. We predicted that the Skyn and BARE will be correlated in terms of both peak data and time-to-peak readings. Our results found that both TAS showed reactivity to the moderate amount of alcohol consumed with the output increasing rapidly for each participant. Data measured by Skyn and BARE were significantly correlated. Peak data and the AUC for Skyn and BARE were also significantly correlated and there was no significant difference in their time to peak. These results suggest that these TAS are reliably measuring the same phenomena. However, neither remained positively correlated to BrAC in peak or time to peak. We expected to see more consistent and stronger correlations between the TAS and BrAC. This could be due to the different ways that they measure alcohol, one through the breath and the other through the skin after being metabolised. In addition, this study only measured 3.5 hours and so did not capture the entire drinking event with output returning to zero.

Our results show there were no significant correlations between BARE and BrAC. In practical terms, this may lead to caution about BARE use with researchers and clinicians if BARE is not shown to be accurate and reliable for measuring alcohol consumption. However, we found that the two TAS were well correlated and appeared to be measuring the same events, significantly correlated with peak data, time to peak, AUC and many of the timepoints across the afternoon. As stated, there will be a slight difference between TAS data vs. BrAC due to their differences and TAS requiring alcohol metabolism before detection. The Skyn has had far more time and testing (Fairbairn and Kang, 2019; Wang *et al.*, 2019, 2021; Fairbairn, Kang and Bosch, 2020; Rosenberg *et al.*, 2021; Ariss, Fairbairn and Bosch, 2022;

Ash *et al.*, 2022; Courtney, Russell and Conroy, 2022; Richards *et al.*, 2022), compared to BARE, which is still in an early prototype stage. Considering this, we would have expected BARE to have a much higher failure rate than we found and a longer time to peak. The BARE failure rate of 4.8% is far lower than earlier Skyn prototype studies (Fairbairn and Kang, 2019; Fairbairn, Kang and Bosch, 2020) and there was no significant difference between Skyn and BARE for peak and time-to-peak data. This is encouraging for BARE's technology at this early phase. However, comparing BARE prototype accuracy to Skyn prototype accuracy, the Skyn still showed significant positive correlations to self-report and BrAC (Fairbairn and Kang, 2019; Wang *et al.*, 2019; Fairbairn, Kang and Bosch, 2020) in its prototype phase.

Another factor for practical use is access to the TAS output. Skyn currently requires data download once every 72 hours and access to a suitable pairing device (an up-to-date Apple iOS). If the wearer does not own an iPhone to sync it themselves, regular meetings with service or hospital staff would be required for data download, increasing resources, effort, and burden for both groups. If data are not successfully downloaded in time, then it will be written over. If the wearer owns an Apple iOS device, they can use the Skyn app and see data in near-real time. However, access to real-time data for the patient may not be wanted (by the Skyn wearer or in a research setting). In addition, owning the latest Apple products in clinical addiction populations may not be as high as among the general public (Milward *et al.*, 2015). BARE does not require a smartphone, and near real-time viewing of BARE data is not currently possible. At present, BARE can hold data for approximately 7 days, so would require weekly downloads.

This is not the first study to report some disagreement or smaller correlations between TAS and BrAC, BAC or self-report. Sakai et al. (2006) found that while each laboratory participant had alcohol-positive TAC and that SCRAM could distinguish between low and high alcohol dosing, individual TAC readings were not reliably equivalent to participant BrAC measurements. Then Marques and McKnight (2009) reported poor performance in the WrisTAS, with a small true positive hit rate of detecting alcohol events at 24% and another study, using WrisTAS, found the TAS to be less reliable among adolescent females ($\leq$20 years) (Croff *et al.*, 2020). In the agreement of self-reported and TAC for a drinking event, TAC and eBAC were moderately correlated (r = 0.49–0.57). Croff et al. (2020) suggest that from their results, while TAS can provide more detailed data and accurate eBAC than self-report, it may be better for reporting the frequency of drinking events. In addition, there are details that self-report can record that TAS cannot, for example, the type of alcohol being consumed. Also, self-reporting can report on days of data without being overwritten, in theory, there is no storage limit to self-report. Our study had a range of ages between 18-47 years and a mean age of 26 years. Our results could reflect this suggestion of TAS data being less reliable in adolescents and younger adults. However, these studies used WrisTAS or SCRAM, yet the BACtrack has previously been found to perform better than SCRAM when using a model for estimating real-time BrAC measurements (Fairbairn, Kang and Bosch, 2020) and high correlations (Fairbairn and Kang, 2019; Rosenberg *et al.*, 2021; Wang *et al.*, 2021).

If TAS did not provide reliable output, then this would have consequences for its use. If used in clinical treatment as a motivational tool for alcohol reduction and the TAC was inaccurate, this could diminish accurate reinforcement (Stitzer and Petry, 2006), ultimately reducing

motivation to reduce and achieve abstinence. Or, if TAS reported an alcohol event when they were abstinent, then the individual would not receive the positive reinforcement earned. Reinforcement interventions are effective in substance use treatment but completing them incorrectly could reduce the efficiency of behaviour change (Davis *et al.*, 2016). Within hospitals, before specific alcohol-related operations, proof of abstinence of up to 3 months may be required. TAS could be used in combination with hospital and addiction service care as proof of abstinence. If the TAS was incorrect, then this could seriously impact their treatment. Also, if TAS underreported alcohol consumption when used with the public for general health information, they could end up consuming more than they think they are, negatively impacting their health.

A current challenge of TAS is output interpretation. These two TAS provide values as TAC ug/l (air) (Skyn) or as mg/dl (BARE) and, at the time of this study, these TAS companies provided no rules to determine a drinking event, high, moderate, or low drinking levels or direct comparison to know the number of drinks consumed or BrAC comparison from the output (Wang *et al.*, 2019). A formula could be created, similar to SCRAM rules to calculate drinking events. SCRAM provides a determination of whether alcohol was consumed, and the criteria (peak TAC, absorption, and elimination rates) which have been successfully used in research (Barnett *et al.*, 2011, 2017; Barnett, Meade and Glynn, 2014; Fairbairn and Kang, 2019; Fairbairn, Kang and Bosch, 2020) and is currently used in the criminal justice system. Gunn et al., recent publication addresses some of the same challenges we found with Skyn and details useful recommendations, including the use of software (TASMAC version 2.0) to process Skyn output (Gunn *et al.*, 2023). However, this publication appeared after our data analysis was completed. In addition, since this study was conducted, BACtrack reported changes to the capability of providing a BAC equivalent to TAC. In time, it is likely other TAS will also provide this comparison. Skyn output being formulated by software rather than manually by a researcher would improve Skyn data collection and analysis. However, due to how TAS work (alcohol consumption, metabolism, to the appearance of alcohol in this sweat vapour), there will always be a slight delay for TAC compared to BrAC, with TAC having a longer elimination time (Karns-Wright *et al.*, 2017; Norman *et al.*, 2020). It also could be argued that outside of research settings currently investigating TAC vs. BrAC or BAC, that TAC comparison to BrAC is not necessary. If TAS can use its output and determine its own rules for alcohol detection of events, tampering, and removal, such as SCRAM has, then it may not be necessary for there to be a comparable scale to other methods, for example, a breathalyser. Each of these tools (TAS, breathalysers, blood/urine tests, self-report) have their own scales and within use for treatment and personal aid, it could be deemed that a scale to compare all against each other is not as important. This study, and previous (Swift *et al.*, 1992; Sakai *et al.*, 2006; Marques and McKnight, 2009; Barnett, Meade and Glynn, 2014; Bond *et al.*, 2014; Hill-Kapturczak *et al.*, 2014; Barnett *et al.*, 2017; Karns-Wright *et al.*, 2018; Fairbairn and Kang, 2019; Rash *et al.*, 2019; Croff *et al.*, 2020; Fairbairn, Kang and Bosch, 2020; Rosenberg *et al.*, 2021; Wang *et al.*, 2021; Ash *et al.*, 2022), show how TAS can detect and react to alcohol consumption and show the curve of a drinking event which is the proposed purpose of their use in clinical settings. This curve of data TAS can show includes information such as the speed of intoxication and duration of intoxication and presence of alcohol through the TAC, features that are not possible with self-reported data. Therefore, even without TAC conversion to BAC or BrAC, TAS has already been shown to be predictive of alcohol consumption (Russell, Turrisi and Smyth, 2022).

Theorised TAS settings of use include alcohol monitoring in the criminal justice system, clinical monitoring in alcohol treatment, monitoring consumption when driving, and monitoring alcohol consumption for healthcare reasons, including general health information. SCRAM is already being used in the US, England, and Wales for use in the criminal justice system. No country has yet implemented them as a tool for clinical practice, to our knowledge. Our findings suggest that there is still further investigation required to do for TAS to report highly consistent and accurate data over extended periods. Yet, even if they are considered sufficiently accurate and reliable to monitor alcohol consumption, using them within our laboratory study for one afternoon we found we had other considerations, such as missing data, device removals and ease of use, which could impact their efficiency as a clinical tool (Gunn et al., 2023). There is also limited work on their acceptability, feasibility and adherence within clinical populations and identifying facilitators and barriers to TAS clinical use which need further exploration (Brobbin et al., 2022a; Richards et al., 2023).

While this study adds to the broader literature on testing TAS in controlled settings with healthy participants, (Dougherty et al., 2012; Hill-Kapturczak et al., 2014; Roache et al., 2015, 2019; Karns-Wright et al., 2017; Fairbairn and Kang, 2019; Fairbairn, Kang and Bosch, 2020), limited work has subsequently occurred within specialist community alcohol treatment clinical settings (Brobbin et al., 2022a). Considering clinical treatment is a posited use of TAS, it must be considered why more studies have not proceeded to test accuracy and use in clinical settings. This could be due to a factor about the TAS itself or research-related challenges. Further research should explore TAS accuracy 1) within the intended settings, 2) over longer periods, and 3) capture whole drinking events with typical drinking amounts that are larger amounts and longer time frames than those able to be ethically provided in laboratories. For example, those who consume alcohol for 24 hours.

It will also be important to continue to examine the technological advancements of BARE. As mentioned, its specific clinical targeted use and features discussed make it appealing for clinical treatment, overcoming some of the burdens of other TAS and reduced size compared to SCRAM. As seen with Skyn and other TAS, technology advancements have led to improved output, failure rate and time to peak (Sakai et al., 2006; Fairbairn and Kang, 2019; Fairbairn et al., 2019; Wang et al., 2019; Norman et al., 2020; Rosenberg et al., 2021). Alongside use in clinical treatment within alcohol services, TAS could be used in other medical contexts. Here further research is needed to consider if knowing if there has been alcohol consumed is enough or if specifically, how much has been consumed is also important and necessary. This study is the first to publish on BARE but we are aware of its limitations and that further assessment is required on all aspects of BARE use in clinical settings.

Limitations

With ethical considerations, we gave participants a restricted amount of alcohol that they reported regularly consuming, and we did not exceed this amount. Outside of the laboratory individuals might be drinking larger quantities of alcohol. So, accuracy at higher quantities will need to be investigated ethically, in the future. It would also be useful to investigate the accuracy of small doses of alcohol. Due to study restrictions and the amount of alcohol participants consumed the study ended 3.5 hours after alcohol consumption

rather than requiring participants to stay until their BrAC reached zero. Participants were able to leave after the 3.5 hours and if their BrAC was under 0.25mg/l (a safe limit). For most when leaving, their BrAC was below 0.20mg/l. So, although this study captured most of the TAC curve, it was not complete. This should be considered regarding the analysis, in particular when calculating the AUC.

Previous accuracy studies on Skyn have found significant correlations between Skyn and self-report (Rosenberg *et al.*, 2021) or BrAC (Fairbairn and Kang, 2019). Fairbairn and Kang (Fairbairn and Kang, 2019) used a similar design as this study, yet they found strong correlations between Skyn and BrAC peak and AUC correlations, while we did not. Differences between this study and theirs include: they used a larger sample size, had a control sample, participants were required to abstain from food for longer beforehand, were provided with a set meal based on weight and wore the devices for longer pre-alcohol consumption. These differences could have had an impact on TAS output due to variability in individual differences affecting alcohol metabolism. If factors such as how recently someone has eaten, or the quantity eaten significantly affect TAC then this could have implications on the TAS accuracy and practical issues of using TAS data in clinical treatment (van Egmond *et al.*, 2020). Fairbairn and Kang (2019) also had participants stay between 3-4 hours post-alcohol, as we did, meaning their participants did not stay until BrAC returned to zero and therefore an incomplete AUC.

To note, we used the latest generation Skyn device and the BARE prototype.

Abbreviations
ANOVA: Analysis Of Variance
AUC: Area Under the Curve
AUDIT: Alcohol Use Disorders Identification Test
BAC: Blood Alcohol Concentration
BMI: Body Mass Index
BrAC: Breath Alcohol Concentration
SD: Standard Deviation
TAC: Transdermal Alcohol Concentration
TAS: Transdermal Alcohol Sensor(s)
US: United States

References:
Alessi, S.M., Barnett, N.P. and Petry, N.M. (2019) 'Objective continuous monitoring of alcohol consumption for three months among alcohol use disorder treatment outpatients', *Alcohol*, 81, pp. 131–138. Available at: https://doi.org/10.1016/j.alcohol.2019.01.008.
Ariss, T., Fairbairn, C.E. and Bosch, N. (2022) 'Examining new-generation transdermal alcohol biosensor performance across laboratory and field contexts', *Alcoholism: Clinical and Experimental Research* [Preprint]. Available at: https://doi.org/10.1111/acer.14977.
Ash, G.I. *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), pp. 783–796. Available at: https://doi.org/10.1111/acer.14804.

Averill, F. *et al.* (2018) 'Transdermal alcohol monitoring combined with contingency management for driving while impaired offenders: A pilot randomized controlled study', *Traffic Injury Prevention*, 19(5), pp. 455–461. Available at: https://doi.org/10.1080/15389588.2018.1448079.

Barnett, N.P. *et al.* (2011) 'Contingency management for alcohol use reduction: A pilot study using a transdermal alcohol sensor', *Drug and Alcohol Dependence*, 118(2–3), pp. 391–399. Available at: https://doi.org/10.1016/j.drugalcdep.2011.04.023.

Barnett, N.P. (2015) 'Alcohol sensors and their potential for improving clinical care', *Addiction*, 110(1), pp. 1–3. Available at: https://doi.org/10.1111/add.12764.

Barnett, N.P. *et al.* (2017) 'A preliminary randomized controlled trial of contingency management for alcohol use reduction using a transdermal alcohol sensor', *Addiction*, 112(6), pp. 1025–1035. Available at: https://doi.org/10.1111/add.13767.

Barnett, N.P., Meade, E.B. and Glynn, T.R. (2014) 'Predictors of detection of alcohol use episodes using a transdermal alcohol sensor', *Experimental and Clinical Psychopharmacology*, 22(1), pp. 86–96. Available at: https://doi.org/10.1037/a0034821.

Bland, J.M. and Altman, D.G. (1986) 'Statistical methods for assessing agreement between two mehtods of clinical measurement', *The Lancet*, 327(8476), pp. 307–310.

Bond, J. *et al.* (2014) 'Adjustments for drink size and ethanol content: new results from a self-report diary and transdermal sensor validation study', *Alcohol Clinical Exp Res.*, 38(12), pp. 3060–3067. Available at: https://doi.org/10.1111/acer.12589.

Brobbin, E. *et al.* (2022a) 'Acceptability and feasibility of wearable transdermal alcohol sensors: A systematic review (Preprint)', *JMIR Human Factors* [Preprint]. Available at: https://doi.org/10.2196/40210.

Brobbin, E. *et al.* (2022b) 'Accuracy of wearable transdermal alcohol sensors: A systematic review', *J Med Internet Res* [Preprint]. Available at: https://doi.org/10.2196/35178.

Caluzzi, G. *et al.* (2019) 'Transdermal monitors to assess alcohol consumption in real-time and real-life–a qualitative study on user-experience', *Addiction Research and Theory*, 27(4), pp. 354–361. Available at: https://doi.org/10.1080/16066359.2018.1530765.

Courtney, J.B., Russell, M.A. and Conroy, D.E. (2022) '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', *Alcohol* [Preprint]. Available at: https://doi.org/10.1016/j.alcohol.2022.11.004.

Croff, J.M. *et al.* (2020) 'Feasibility and reliability of continuously monitoring alcohol use among female adolescents and young adults', *Drug and Alcohol Review* [Preprint], (January). Available at: https://doi.org/10.1111/dar.13045.

Davis, D.R. *et al.* (2016) 'A review of the literature on contingency management in the treatment of substance use disorders, 2009–2014', *Preventive Medicine*, 92, pp. 36–46. Available at: https://doi.org/10.1016/j.ypmed.2016.08.008.

Dougherty, D.M. *et al.* (2012) 'Comparing the detection of transdermal and breath alcohol concentrations during periods of alcohol consumption ranging from moderate drinking to binge drinking', *Exp Clin Psychopharmacology*, 20(5), pp. 373–381. Available at: https://doi.org/10.1037/a0029021.

Dougherty, D.M. *et al.* (2014) 'Use of continuous transdermal alcohol monitoring during a contingency management procedure to reduce excessive alcohol use', *Drug Alcohol Depend.*, 142, pp. 301–306. Available at: https://doi.org/10.1016/j.drugalcdep.2014.06.039.

Dougherty, D.M. *et al.* (2015) 'Transdermal Alcohol Concentration Data Collected During a Contingency Management Program to Reduce At-Risk Drinking', *Drug Alcohol Depend.*,

148(3), pp. 77–84. Available at: https://doi.org/10.1016/j.drugalcdep.2014.12.021.Transdermal.

Dumett, M.A. *et al.* (2008) 'Deconvolving an estimate of breath measured blood alcohol concentration from biosensor collected transdermal ethanol data', *Applied Mathematics and Computation*, 196(2), pp. 724–743. Available at: https://doi.org/10.1016/j.amc.2007.07.026.

van Egmond, K. *et al.* (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), pp. 1918–1932. Available at: https://doi.org/10.1111/acer.14432.

Fairbairn, C.E. *et al.* (2019) 'Estimating the quantity and time course of alcohol consumption from transdermal alcohol sensor data: A combined laboratory-ambulatory study', *Alcohol*, 81, pp. 111–116. Available at: https://doi.org/10.1016/j.alcohol.2018.08.015.

Fairbairn, C.E. and Kang, D. (2019) 'Temporal Dynamics of Transdermal Alcohol Concentration Measured via New-Generation Wrist-Worn Biosensor', *Alcoholism: Clinical and Experimental Research*, 43(10), pp. 2060–2069. Available at: https://doi.org/10.1111/acer.14172.

Fairbairn, C.E., Kang, D. and 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(April). Available at: https://doi.org/10.1016/j.drugalcdep.2020.108205.

Greenfield, T.K., Bond, J. and Kerr, W.C. (2014) 'Biomonitoring for improving alcohol consumption surveys the new gold standard?', *Alcohol Research: Current Reviews*, 36(1), pp. 39–46.

Gunn, R.L. *et al.* (2023) 'Use of the BACtrack Skyn alcohol biosensor: Practical applications for data collection and analysis', *Addiction* [Preprint]. Available at: https://doi.org/10.1002/add.16207.

Hill-Kapturczak, N. *et al.* (2014) 'Do variable rates of alcohol drinking alter the ability to use transdermal alcohol monitors to estimate peak breath alcohol and total number of drinks?', *Alcohol Clinical Exp Res.*, 38(10), pp. 2517–2522. Available at: https://doi.org/10.1111/acer.12528.

ION Wearable (no date). Available at: https://www.ionwearable.com (Accessed: 25 April 2023).

Karns-Wright, T.E. *et al.* (2017) 'Time Delays in Transdermal Alcohol Concentrations Relative to Breath Alcohol Concentrations', *Alcohol and Alcoholism*, 52(1), pp. 35–41. Available at: https://doi.org/10.1093/alcalc/agw058.

Karns-Wright, T.E. *et al.* (2018) 'The correspondence between transdermal alcohol monitoring and daily self-reported alcohol consumption', *Addictive Behaviors*, 85(May), pp. 147–152. Available at: https://doi.org/10.1016/j.addbeh.2018.06.006.

Kuntsche, E. *et al.* (2022) 'Comparing strengths and weaknesses of contemporary quantitative methods to collect data on alcohol consumption: an "at-a-glance" overview', *Addiction Research and Theory* [Preprint]. Available at: https://doi.org/10.1080/16066359.2022.2143495.

Leffingwell, T.R. *et al.* (2013) 'Continuous Objective Monitoring of Alcohol Use: Twenty-First Century Measurement Using Transdermal Sensors', *Alcoholism: Clinical and Experimental Research*, 37(1), pp. 16–22. Available at: https://doi.org/10.1111/j.1530-0277.2012.01869.x.

Litten, R.Z., Bradley, A.M. and Moss, H.B. (2010) 'Alcohol biomarkers in applied settings: Recent advances and future research opportunities', *Alcoholism: Clinical and Experimental Research*. Blackwell Publishing Ltd, pp. 955–967. Available at: https://doi.org/10.1111/j.1530-0277.2010.01170.x.

Luczak, S.E., Rosen, I.G. and Wall, T.L. (2015) 'Development of a real-time repeated-measures assessment protocol to capture change over the course of a drinking episode', *Alcohol and Alcoholism*, 50(2), pp. 180–187. Available at: https://doi.org/10.1093/alcalc/agu100.

Marques, P.R. and McKnight, A.S. (2009) 'Field and laboratory alcohol detection with 2 types of transdermal devices', *Alcoholism: Clinical and Experimental Research*, 33(4), pp. 703–711. Available at: https://doi.org/10.1111/j.1530-0277.2008.00887.x.

Milward, J. *et al.* (2015) 'Mobile phone ownership, usage and readiness to use by patients in drug treatment', *Drug and Alcohol Dependence*, 146(1), pp. 111–115. Available at: https://doi.org/10.1016/j.drugalcdep.2014.11.001.

Moos, R.H. and Moos, B.S. (2006) 'Rates and predictors of relapse after natural and treated remission from alcohol use disorders', *Addiction*, 101(2), pp. 212–222. Available at: https://doi.org/10.1111/j.1360-0443.2006.01310.x.

Norman, T. *et al.* (2020) 'Combining transdermal and breath alcohol assessments, real-time drink logs and retrospective self-reports to measure alcohol consumption and intoxication across a multi-day music festival', *Drug and Alcohol Review* [Preprint], (October). Available at: https://doi.org/10.1111/dar.13215.

Rash, C.J. *et al.* (2019) 'Monitoring alcohol use in heavy drinking soup kitchen attendees', *Alcohol*, 81, pp. 139–147. Available at: https://doi.org/10.1016/j.alcohol.2018.10.001.

Richards, V.L. *et al.* (2022) 'Correspondence between alcohol use measured by a wrist-worn alcohol biosensor and self-report via ecological momentary assessment over a 2-week period', *Alcoholism: Clinical and Experimental Research* [Preprint]. Available at: https://doi.org/10.1111/acer.14995.

Richards, V.L. *et al.* (2023) 'Identifying Desired Features That Would Be Acceptable and Helpful in a Wrist-Worn Biosensor–Based Alcohol Intervention: Interview Study among Adults Who Drink Heavily', *Journal of Medical Internet Research*, 25. Available at: https://doi.org/10.2196/38713.

Roache, J.D. *et al.* (2015) 'Using Transdermal Alcohol Monitoring to Detect Low-Level Drinking', *Alcoholism: Clinical and Experimental Research*, 39(7), pp. 1120–1127. Available at: https://doi.org/10.1111/acer.12750.

Roache, J.D. *et al.* (2019) 'Processing transdermal alcohol concentration (TAC) data to detect low-level drinking', *Alcohol*, 81, pp. 101–110. Available at: https://doi.org/10.1016/j.alcohol.2018.08.014.

Rosenberg, M. *et al.* (2021) 'Wearable alcohol monitors for alcohol use data collection among college students: feasibility and acceptability in a pilot study'. Available at: https://doi.org/10.1101/2021.02.17.21251959.

Russell, M.A., Turrisi, R.J. and 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), pp. 100–113. Available at: https://doi.org/10.1111/acer.14739.

Sakai, J.T. *et al.* (2006) 'Validity of transdermal alcohol monitoring: Fixed and self-regulated dosing', *Alcoholism: Clinical and Experimental Research*, 30(1), pp. 26–33. Available at: https://doi.org/10.1111/j.1530-0277.2006.00004.x.

Saunders, J.B. *et al.* (1993) 'Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO Collaborative Project on Early Detection of Persons with Harmful Alcohol Consumption-II', *Addiction*, 88(6), pp. 791–804. Available at: https://doi.org/10.1111/j.1360-0443.1993.tb02093.x.

Simons, J.S. *et al.* (2015) 'Quantifying alcohol consumption: Self-report, transdermal assessment, and prediction of dependence symptoms', *Addictive Behaviors*, 50, pp. 205–212. Available at: https://doi.org/10.1016/j.addbeh.2015.06.042.

Sobell, L. and Sobell, M. (1992) 'Timeline Follow-Back', in *Litten R.Z., Allen J.P. (eds) Measuring Alcohol Consumption*. Totowa, NJ: Humana Press, pp. 41–72. Available at: https://doi.org/10.1007/978-1-4612-0357-5_3.

Stitzer, M. and Petry, N. (2006) 'Contingency management for treatment of substance abuse', *Annual Review of Clinical Psychology*, 2, pp. 411–434. Available at: https://doi.org/10.1146/annurev.clinpsy.2.022305.095219.

Swift, R.M. *et al.* (1992) 'Studies on a Wearable, Electronic, Transdermal Alcohol Sensor', *Alcoholism: Clinical and Experimental Research*, 16(4), pp. 721–725. Available at: https://doi.org/10.1111/j.1530-0277.1992.tb00668.x.

Wang, Y. *et al.* (2019) 'Wrist-worn alcohol biosensors: Strengths, limitations, and future directions', *Alcohol*, 81, pp. 83–92. Available at: https://doi.org/10.1016/j.alcohol.2018.08.013.

Wang, Y. *et al.* (2021) 'Wrist-worn alcohol biosensors: Applications and usability in behavioral research', *Alcohol* [Preprint]. Available at: https://doi.org/10.1016/j.alcohol.2021.01.007.

Webster, G.D. and Gabler, H.C. (2008) 'Modeling of transdermal transport of alcohol: Effect of body mass and gender', *Biomed Sci Instrum*, 44, pp. 361–366.

Figure Legends

Figure 1. Average Skyn, BARE and BrAC data for all participants.

Figure 1 A, B, C: A) Average Skyn output every 15 minutes for 3.5 hours (ug/l air). B) BARE output every 15 minutes for 3.5 hours (mg/dl). C) Average BrAC measured every 15 minutes for 3.5 hours (mg/l).

BrAC = Breath Alcohol Concentration

TAC = Transdermal Alcohol Concentration

Figure 2. Bland Altman plot for time to peak Skyn and BARE.

Figure 2. BA plot for time to peak (minutes) Skyn and BARE devices. Plot of differences between Skyn and BARE (y-axis) versus the mean of the Skyn and BARE (x-axis). The middle blue line represents the mean difference between the TAS. The top red line represents the upper 95% CI and the bottom red line represents the lower 95% CI.

Table Legends

Table 1. Sample characteristics.

Significant p<0.05

AUDIT = Alcohol Use Disorder Identification Test

BrAC = Breath Alcohol Concentration

BMI = Body Mass Index

Table 2. Participant information: ethanol consumed, weight, device peak, time to peak data and AUC.

Table 2. Alcohol consumed (grams), Weight (kg), peak BrAC (g/210L), peak BrAC (mg/l), peak Skyn (mg/l), peak BARE (mg/l), time to peak (minutes) and AUC. A dash (-) marks missing data for that participant and device output.
AUC = Area Under the Curve
BrAC = Breath Alcohol Concentration

Table 3. Pearson correlations.
Table 4. Repeated Measures ANOVA.
*Significant p<0.05
BrAC = Breath Alcohol Concentration
TAC = Transdermal Alcohol Concentration