

Field and Laboratory Alcohol Detection With 2 Types of Transdermal Devices

Paul R. Marques and A. Scott McKnight

Background: Two types of transdermal electrochemical sensors that detect alcohol at the skin surface were evaluated. One, the AMS SCRAMTM device, is locked onto the ankle and is based on a fuel cell sensor; the other, a Giner WrisTASTM device, worn on the wrist, is based on a proton exchange membrane. SCRAM is used by several court systems in the United States to monitor alcohol offenders, WrisTAS, a research prototype, is not commercially available.

Methods: The 2 devices were worn concurrently by 22 paid research subjects (15 men, 7 women), for a combined total of 96 weeks. Subjects participated in both laboratory-dosed drinking to a target of 0.08 g/dl blood alcohol concentration (BAC), and normal drinking on their own; all subjects were trained to use and carry a portable fuel-cell breath tester for BAC determinations. Overall 271 drinking episodes with BAC \geq 0.02 g/dl formed the signal for detection—60 from laboratory dosing, and 211 from self-dosed drinking, with BAC ranging from 0.02 to 0.230 g/dl (mean 0.077 g/dl).

Results: False negatives were defined as a transdermal alcohol concentration response equivalent $<$ 0.02 g/dl when BAC \geq 0.02 g/dl. The overall true-positive hit rate detected by WrisTAS was 24%. The low detection rate was due to erratic output and not recording during nearly 67% of all episodes; reportedly a chipset, not a sensor problem. SCRAM correctly detected 57% across all BAC events, with another 22% (total 79%) detected, but as $<$ 0.02 g/dl. When subjects dosed themselves to BAC \geq 0.08 g/dl, SCRAM correctly detected 88% of these events. SCRAM devices lost accuracy over time likely due to water accumulation in the sensor housing. **Neither unit had false-positive problems when true BAC was $<$ 0.02 g/dl.**

Conclusions: Each device had peculiarities that reduced performance, but both types are able to detect alcohol at the skin surface. With product improvements, transdermal sensing may become a valuable way to monitor the alcohol consumption of those who should be abstaining.

Key Words: Transdermal Alcohol, DWI, Blood Alcohol Concentration, Monitoring, Detection.

SOME INDIVIDUALS WITH alcohol problems are viewed as posing significant risks to the public, and as a result, courts often order them to remain completely abstinent from alcohol. Sometimes these sentences are enforced by jail or the threat of jail along with other forms of monitoring. Although jail can often successfully enforce abstinence, it is at best an expensive and short-term remedy because that environment does not offer much opportunity to practice self-control. It serves the interests of retribution more so than rehabilitation.

Alcohol ignition interlock devices very effectively prevent the use of a car after alcohol has been consumed (Marques et al., 2001; Voas et al., 1999; Willis et al., 2004), but because

these are specific to a car, they are not a deterrent to drinking when not driving. Sometimes the courts will confine a problem drinker to his home using electronic proximity monitors that are coupled with regular breath tests, either through a telephone-linked device with voice recognition or via house calls from probation services. Alternatively, direct alcohol biomarkers, such as ethanol itself (typically a 6 to 12 hour window of detection), or markers such as ethyl glucuronide or ethyl sulfate (with 36 or more hours of detection) can be measured in urine samples one or more times per week to document consumption (Borucki et al., 2005; Helander and Beck, 2005; Wurst et al., 2003). All of these approaches have strengths and weaknesses, and all require periodic specimen collection.

A recent addition to the monitoring options that has undergone developmental research is transdermal alcohol detection. Approximately 5% of ethanol consumed is lost through breath, and about 1% is lost through the skin (Swift, 2003). Two electrochemical devices are available that detect alcohol gas at the skin surface and have data storage and retrieval options as well as circumvention detection protocols. The 2 devices each approach the problem of measurement differently. This report is an evaluation of the accuracy and sensitivity of these devices under both laboratory-dosed and field

From the Pacific Institute for Research and Evaluation, Calverton, Maryland.

Received for publication February 11, 2008; accepted November 14, 2008.

Reprint requests: Paul R. Marques, PhD, Senior Research Scientist, Pacific Institute for Research and Evaluation, 11720 Beltsville Road, Suite 900, Calverton, MD 20705-3111; Fax: 301-755-2799; E-mail: marques@pire.org

Copyright © 2009 by the Research Society on Alcoholism.

DOI: 10.1111/j.1530-0277.2008.00887.x

self-initiated drinking while used for several weeks of continuous wear.

The Alcohol Monitoring Systems, Inc. (AMS, Littleton, CO) device, the SCRAMTM (Secure Continuous Remote Alcohol Monitoring), measures ethanol gas at the skin surface, often referred to as transdermal alcohol concentration (TAC) using a fuel-cell sensor. The system consists of 3 components: (1) a SCRAM "bracelet" that is locked onto the ankle; (2) a SCRAM modem for uploading data; and (3) a remote server, SCRAMNetwork, for aggregating data from offenders and for reporting these data to monitoring staff. Based on current practice, the unit is worn 24 hours a day, 7 days a week, for up to several months. Typically, it is set to sample air at 60-minute intervals in a confined area above the skin surface enclosed by a rubber muff; it switches to a 30-minute sampling protocol if alcohol is detected. It was designed for security and remote reporting to minimize circumvention and to render data usable by courts or corrections. In most applications, the SCRAM modem is scheduled to read the bracelet log during normal sleep hours and transfer data by autodial to the SCRAM server. During error-free operation with no alcohol detected 24 samples are uploaded daily. Within minutes after upload, the data are available for a monitoring authority to review. The ankle bracelets evaluated weighed about 8 ounces (225 g). In addition to the alcohol sensor, other sensors detect changes in proximity and skin surface reflectivity and temperature near the alcohol sensor. The 2 nonalcohol sensors are parts of the circumvention detection protocols. Battery life is approximately 30 to 45 days. As of January 2008, AMS reports 7,000 devices are in daily service; since their introduction in 2003, 53,000 offenders have been monitored.

A 2-part evaluation study of SCRAM was published by the University of Colorado investigators (Sakai et al., 2006). Part 1 included a 1-day laboratory analysis in which subjects arrived, were hooked up, drank, and had the bracelet removed. The devices were found to adequately discriminate lower and higher dosed subjects. Part 2 was a 7-day wear study in which subjects (both alcohol dependent and nondependent) logged drinking while wearing their SCRAM bracelets. The investigators reported no episodes of false-positive TAC results, and while blood alcohol concentration (BAC) and TAC could not be considered quantitatively equivalent, there was qualitative parity between reported drinking and SCRAM results. Also, Sakai et al. found that devices readily discriminated the consumption patterns of alcohol-dependent and social drinkers. Although the primary application of this product is positioned for the offender market, the technology also has potential uses for patient monitoring (e.g., alcohol abuse/dependence or liver transplant patients) by detecting and reporting lapses before they become relapses, long a goal of treatment compliance monitoring programs.

The Giner WrisTASTM (Wrist Transdermal Alcohol Sensor; Giner, Inc., Newton, MA) is a research prototype of a sensor that is not commercially available. Nonetheless, due to

support from the National Institute on Alcohol Abuse and Alcoholism, it has a documented research history and was developed with treatment applications in mind. This device affixes to the wrist with a Velcro strap and is about the size of a wristwatch; it is based on Giner's patented proton exchange membrane technology. In the WrisTAS, an electrode oxidizes the ethanol to form acetic acid that diffuses into a reservoir. Alcohol concentration is reflected by the level of oxidation current and is continuously monitored. The device writes a file entry, typically every 5 minutes, by averaging a near continuous signal that reflects TAC over that time. Data logged in the device can be periodically downloaded to a computer via a serial port interface. The data storage capacity of WrisTAS version 5 is approximately 21 days. Giner devices, selected for good performance characteristics, have been reported to be linear within normal pharmacologic ranges of ethanol dosing. Swift and colleagues (1992) and Swift (2000) reported that the WrisTAS linearity extends from 5 to 500 mg/dl (0.005 to 0.50 g/dl). This transdermal device was shown to output a TAC that parallels the more familiar BAC curves but shifted to the right with a 1- to 2-hour delay. The alcohol sensor in the Giner device can respond to changes in alcohol more promptly than can a fuel-cell sensor. This device also has non-alcohol sensors (temperature and skin resistance) that can be used for monitoring circumvention.

Contrary to the high degree of accuracy in selected WrisTAS units reported by Swift and collaborators, Greenfield and colleagues (2005) reported considerable variability in the ability of unselected WrisTAS version 5 devices to detect alcohol at the skin surface. They reported that the device performed with low reliability on 43% of the trials and that less than 25% of the trials were deemed high quality. The difficulties were believed to be unrelated to the alcohol sensor itself but rather a secondary consequence of an undependable input/output (I/O) chipset that controls file operations.

For any transdermal device, error can derive from the measuring device and from the alcohol signal itself. In an effort to characterize the nature of the transdermal alcohol signal from kinetics of a model system, Anderson and Hlastala (2006) reported that ethanol transport through the skin is substantially affected by the stratum corneum, the external most layer of the skin surface. They determined that detectable transdermal ethanol gas concentration is particularly affected by the thickness, temperature, and hydration state of the stratum corneum. Because of these variables, they concluded that TAC cannot be considered a quantitative estimate of BAC as TAC can vary by as much as 2:1, depending on local skin factors.

Neither of these problems renders the technology seriously deficient in serving its intended purpose for alcohol use monitoring. The TAC is not BAC but is usefully related to BAC, and sensor capability is distinct from file operations. This paper reports on an evaluation of 2 classes of devices during extended concurrent wear by subjects who were paid to participate in both laboratory and field evaluation following alcohol consumption.

MATERIALS AND METHODS

Subjects and Criteria

Subjects between 21 and 35 years gave signed consent and then participated in a screening process for subject selection. Inclusion criteria were regular drinking but not to unusual excess and no drug or alcohol-related health or criminal problems. Further, they should not be using contraindicated medications or be pregnant. Subjects were asked to provide a urine sample upon entry into the subject pool to confirm that they showed no positives for benzodiazepines or the standard panel of 5 illicit drugs. The urine toxicology screening devices were Biosites Triage Tox ScreensTM, and the results were machine read from the Biosites Triage Meter PlusTM. Pregnancy tests were performed with over-the-counter EPTTM home pregnancy tests. Procedures were approved by our Institutional Review Board under FWA#00007038.

A prospective subject's normal level of alcohol consumption was assessed by the AUDIT (Alcohol Use Disorders Identification Test), and a subset of questions from the AUDADIS (Alcohol Use Disorders and Associated Disabilities Schedule). From 55 telephone or e-mail inquiries, 32 people were eventually screened for consideration. Of those screened, 10 were excluded primarily due to self-reported risk levels of alcohol consumption. Three subjects screened positive for benzodiazepines or marijuana.

The final subject pool participating in the research included 15 men and 7 women. Across all subjects, 30 trials were conducted (mean age 26.7, 68% men). These include 18 four-week trials and 12 two-week trials. The supplemental 2-week trials served to add more drinking episodes. Five of the males and 3 of the females participated for an initial 4-week and the subsequent 2-week trial. Data were evaluated at the event level not the subject level.

Subjects were paid at a rate of \$100/week for participation, plus an added bonus was paid if they stayed with the study for the full duration of 4 weeks (\$400) or 2 weeks (\$200). The bonus was deemed warranted due to the disruptive effects should subjects drop out; no one dropped out.

In addition to drinking, subjects were instructed to keep a log of all food and alcohol consumed on any day with a drinking episode of 2 or more drinks. Subjects were instructed to e-mail drink logs every day including nondrinking days. Not all logs arrived the following day, but daily reminders were issued if the log was not received.

Basic Data Elements

As noted above, the study required subjects to log all alcoholic drinks or drink equivalents during the study period and, when not drinking in the laboratory situation, subjects had been trained to self-test BAC with a portable fuel-cell personal breath tester (PBT). Whenever a subject consumed more than 1 drink, he or she was instructed to follow standard procedure of either performing a mouth rinse with water and/or waiting 15 minutes after the last beverage sip before testing with the PBT. This was accomplished outside the office by entrusting all subjects with a SD-400TM breath tester (CMI Inc., Owensboro, KY) to bring with them whenever they might be drinking away from home. Accordingly, with 2 transdermal sensors (the WrisTAS and the SCRAM), a fuel-cell PBT, and a log of drinking, there are 4 types of data elements that represent ethanol consumption during the study.

Alcohol Consumption Procedures

There were 2 types of alcohol consumption in this study: (1) laboratory dosing in which subjects came to the research site were weighed to calculate dose and drank in the morning before any significant food consumption, and (2) self-dosing or free-form alcohol consumption of their own choosing. Each episode of alcohol con-

sumption is coded and logged for each subject. Laboratory-dosed and self-dosed drinking events were tracked separately because the method of dosing was very different.

Laboratory Dosing. In the laboratory at 9 to 10 AM, subjects were given an amount of distilled spirits expected to bring their BACs to 80 mg/dl (0.08 g/dl) when consumed over a 30-minute period. Subjects could add soft drinks or juice to their drinks and were told to space their consumption over the full 30-minute period. Fifteen minutes after drinks were finished, mouths were rinsed and breath testing began. BAC readings in the laboratory phase represented the average of 2 PBT fuel-cell testers used successively within the same 1-minute period. In accordance with human subject requirements, when subjects were dosed in the laboratory condition, a medical technician was present in the event of any adverse reaction, such as aspiration of vomit, severe flushing reaction, or other (there were none). Subjects were instructed to arrive for laboratory dosing with no measurable BAC and to limit drinking the night before so that TAC would have a chance to return to 0 prior to laboratory dosing. They were also asked to avoid drinking for several hours after laboratory dosing so that TAC could return to 0 prior to any further self-dosing. There were some occasions when laboratory and self-dosing TAC levels overlapped. There were no cases where laboratory dosing began with measurable BACs.

Self-Dosing. When self-dosing, subjects kept a log of all food and alcohol consumed, and recorded BAC during drinking episodes with 2 or more alcoholic drinks. Subjects were encouraged to log anything they felt noteworthy, but at a minimum, they were told the logs should contain:

- The time and size (small, medium, large) of a meal.
- The time and number of standard drinks or drink equivalents consumed. Unusual drink size or alcohol content was also noted.
- The time and BAC result. Subjects took the first BAC reading 30 minutes after the first drink, and every hour afterward until the BAC reached 0.00 g/dl or until the subject went to bed for the night.

Subjects sent drink logs daily via e-mail. On nondrinking days, subjects sent an e-mail stating that they had no alcohol. Drinking logs were submitted for all but 5 of 249 self-dosed drinking events. The information about meal size was collected as reserve information for resolving discrepancies between transdermal and blood alcohol concentrations.

Coding TAC Results

Drinking episodes were defined as having a peak BAC ≥ 0.02 g/dl, and the categorization task was to determine whether and how well transdermal devices detected the alcohol signal. This criterion was selected for 4 reasons: (1) the literature describes a delay of 60 to 120 minutes between peak BAC and peak TAC while the average biotransformation rate of ethanol is 0.017 g/dl, (2) BAC less than 0.02 g/dl produces little practical impairment in most drinkers, (3) other forensic devices like alcohol ignition interlocks do not attempt to lock out below 0.02 g/dl, (4) 0.02 g/dl TAC is the criterion used for the AMS (SCRAM) in practice to minimize false-positive accusations of probationers. Data from all sources were aggregated into a data file and composite graphical display. Two coders independently reviewed 41 episodes and achieved good (>96%) agreement on the definition of episodes, maximum BAC, maximum SCRAM TAC, maximum WrisTAS TAC, and time of these TAC maxima. Because there were no real disagreements between the 2 coders, the subsequent judgment of a single coder was used to maintain consistency. Using the spreadsheet, the following data were recorded for each drinking episode:

- The time and level of the maximum BAC recorded.
- The time and level of maximum TAC after the maximal BAC occurrence.
- Subcategories of TAC responses that were distinctive types of false negatives. The definitions of false negative subtypes are described below.

For each drinking episode and for each transdermal device, the coder categorized the TAC readings as one of the following:

True Positive

- *A hit*—the drinking episode was clearly visible with a TAC ≥ 0.02 g/dl.

Subtypes of False Negatives

- *<0.02 g/dl*—the alcohol was evident but the TAC response < 0.02 g/dl.
- *Low-confidence*—a change in TAC was apparent, though difficult to see without knowing that BAC was elevated, usually due to high variance in TAC before and after drinking episodes.
- *Too noisy*—the TAC readings were patternless, too noisy, or too variable to clearly distinguish a drinking episode.
- *Missing data*—apparent device failure: no TAC data were available for retrieval during or after a drinking episode.
- *Complete false negative*—the transdermal device was on and recording but based on TAC reading it seems to have completely missed a positive BAC ≥ 0.02 g/dl.

Further Considerations

Because the research subjects were paid to participate and real world offenders are under some threat by the courts or corrections to participate, the compliance motivation of these 2 groups is different. Also, research subjects were selected because they do drink and are expected to drink, whereas offenders are told not to drink and if they do, consequences follow, potentially including jail time.

Also, the 2 devices under evaluation differ on several dimensions. The AMS SCRAM is available for the criminal justice marketplace and is in service now, whereas WrisTAS is a research prototype developed by Giner Inc., a sensor company. These are not direct competitor products; both are electrochemical devices, but based on different underlying technology. SCRAM uses an alcohol fuel cell; WrisTAS uses a hydrated proton exchange membrane.

The AMS SCRAM device available for use by courts or corrections restricts the estimated BAC report to 0.08 g/dl, even when the actual TAC exceeds that value. AMS maintains this convention because any BAC ≥ 0.08 g/dl is above an “actionable” level for court-ordered abstinence. For the purposes of this evaluation, however, AMS made available the underlying raw TAC results beyond 0.08 g/dl providing useful detail. It is worth noting that in 2008, AMS introduced a newer version of their SCRAM device (SCRAM 2) and Giner has now introduced WrisTAS version 7.

RESULTS

Drinking Data

A total of 309 episodes of drinking were logged either in the laboratory ($n = 60$) or by the subjects’ self-dosing ($n = 249$). Of those 309 drinking episodes, 271 achieved a fuel-cell breath tester measured BAC value equal to or greater than 0.02 g/dl. The BACs for these 271 episodes ranged from 0.020 g/dl to 0.230 g/dl with a mean of 0.078 g/dl. The data in Table 1 characterize the drinking episodes available for analysis. In the laboratory setting, the BAC results were symmetrical around 0.08 g/dl, the target dosed BAC level. All

Table 1. BAC Characteristics Lab-Dosing and Self-Dosing Episodes

Type		<i>n</i>	Minimum	Maximum	Mean	SD
Lab	BAC max	60	.041	.123	.08287	.015771
Self	BAC max	211	.020	.230	.07675	.044295
Total	BAC max	271	.020	.230	.07811	.039835

were told to eat very lightly or not at all before coming to the laboratory to minimize differences in absorption time.

Self-dosing is just normal drinking, and although the mean BAC for laboratory- and self-dosed drinking were similar, there was considerable positive skew with self-dosed drinking BACs ranging up to 0.230 g/dl, providing some experience evaluating the transdermal sensors with heavy drinking.

For the 38 self-dosed drinking episodes that were < 0.02 g/dl (not shown), the mean BAC was 0.011 with a standard deviation of 0.005. These 38 low BAC drinking episodes were excluded from the analysis for several reasons, including the expected 2-hour delay for peak TAC after peak BAC that would have reduced alcohol to near 0 according to the classic Widmark equation (Widmark, 1932) metabolism rate of 0.017 g/dl/h. The 271 episodes ≥ 0.02 g/dl form the basic data elements for this analysis. This convention of ignoring values lower than 0.02 g/dl is a common practice: it is the protocol used by AMS with their offender-monitoring algorithms, it is the lower limit of alcohol-interlock lockout levels in any jurisdiction, and it is the lower level of measurable impairment among adult drinkers.

WrisTAS Coding From Judgment

Of the 271 episodes of drinking with known maximal BAC ≥ 0.02 g/dl, only 64, or 23.6%, were judged from WrisTAS data to be clearly ≥ 0.02 g/dl. False negatives occurred for different reasons. Sometimes data could not be retrieved from the device for display and evaluation. Using the protocol described in the Materials and Methods section, the problems with the WrisTAS fell into the categories shown in Table 2. The most common problem (37.6%) was missing data (no data captured or retrieved from the device), followed by data that were too erratic or noisy to be clearly associated with a BAC (16.2%), or data that could not be judged a positive hit due to the rater’s low confidence (13.3%) coding it as a hit. False negatives in which there was no response from a

Table 2. WrisTAS and SCRAM Frequencies Coded by Type

	WrisTAS		SCRAM	
	Frequency	Percent	Frequency	Percent
Hit	64	23.6	155	57.2
<0.02 g/dl	3	1.1	61	22.5
Low-confidence	36	13.3	7	2.6
Too noisy	44	16.2	8	3.0
Missing data	102	37.6	1	0.4
Nonresponse	22	8.1	39	14.4
Total	271	100	271	100

working unit occurred in 8.1% of the cases. BAC episodes that resulted in a low TAC response were rare (1.1%).

SCRAM Coding From Judgment

Table 2 also shows the SCRAM device detected BACs of 0.02 g/dl or higher in 155 of the 271 positive BAC events, for a valid hit rate of 57.2% overall. Here too, false negatives occurred for reasons unique to this device. Sixty-one (22.5%) of the positive BAC events were detected as a positive TAC but less than 0.02 g/dl. False negatives in which no response was found occurred in 14.8% of the events. There were relatively few missing data problems.

Automated Alerts

With SCRAM, there are 2 ways to assess the detection rate: via coded rater judgments described in Materials and Methods and shown in Table 2, and through the “Alcohol Alerts” that are found on the AMS server. The Alerts derive from an AMS algorithm intended to call alcohol positive results to the attention of a supervisor. The coded judgments and the AMS Alerts were independently scored; Table 3 shows their cross-tabulation. Agreement for true-positive detection by these 2 methods was 93.5%, and the agreement for true negatives (BAC < 0.02 g/dl) was 91.5% (kappa = 0.85, p = 0.000). The 155 coded hits (row value) in Table 3 are the same events as the 155 hits shown for SCRAM row 1, column 4 of Table 2. Each method agreed on 145 events and differed on 10 events. Upon examination, all disputed events were near the thresholds for detection. There is no comparable Giner alert system to use with the WrisTAS data; however, the cross-tabulation in Table 3 to a limited extent serves to validate the hit coding system devised for use with both types of devices.

No Real False Positives Found

Although there were occasionally elevated readings from the alcohol sensors that were unrelated to alcohol consumption, virtually all of these were explainable as either some external source of interference or transient blips that did not have the characteristics of consumed alcohol. Dozens of automotive and home products, including personal care products,

Table 3. Two Methods for Estimating SCRAM™ Hits ≥0.02 g/dl, and Nonevents <0.02 g/dl

	AMS alcohol alert		Total
	No alert	Alert	
SCRAM hit			
No	106	10	116
Yes	10	145	155
Total	116	155	271

AMS, Alcohol Monitoring Systems; SCRAM, Secure Continuous Remote Alcohol Monitoring.

contain alcohols, and these can cause transient elevations in the readings that rise and fall with timeframes that are not physiologically tenable as a result of ingested alcohol. **Presumably, if people work in an environment where there is a sustained level of solvent or perfumes in the air, this can cause extended elevation; however, none of the subjects showed evidence of this during the study.**

Differences in Self-Dosing and Laboratory-Dosing

True-Positive Hit Rate. Table 2 shows overall detection rates across all episodes, regardless of maximal BAC and without separately breaking out laboratory-dosed from self-dosed drinking. The rapid rise and fall of BAC in the laboratory-dosing situation resulted in lower transdermal device hit rates than did the more protracted, normal alcohol use that came with self-dosing. Across both types of dosing and using judgment as the criterion, at higher peak BAC levels, the rate of true-positive transdermal hits increases for both devices. Figure 1 portrays the true positive hit rates for SCRAM and WrisTAS for 4 BAC ranges: ≥0.02, ≥0.04, ≥0.06, and ≥0.08. Data for this figure exclude those instances where the device did not respond at all (missing data) but includes all other coded categories. The performance difference between the 2 devices primarily reflects the difficulty of discerning signal from the noisy patterns in some of the WrisTAS results. The important feature of Fig. 1 is the similarity of patterns. Detection of the rapid rise and fall associated with laboratory drinking was poorer than self-dosed drinking, a finding that may reflect the rapid clearance of ethanol from circulation after a brief BAC peak that is not sustained. Area under the BAC curve across several hours would probably be a better predictor of TAC response level than is a peak BAC. Areas cannot be calculated for self-dosed drinking.

Magnitude of Peak BAC to TAC Differences for True-Positive Hits. When subjects dosed themselves with ethanol (normal drinking), the BAC episodes that were judged to be “hits” had a mean peak SCRAM TAC (0.081 g/dl), which was lower by 0.014 g/dl relative to the mean peak BAC

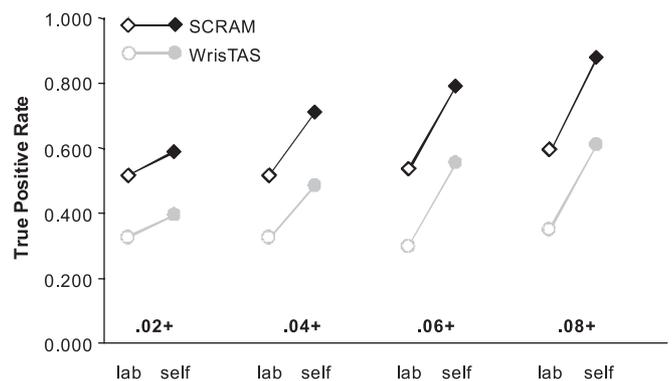


Fig. 1. True-positive rates for SCRAM and WrisTAS with 2 types of dosing (lab = open markers; self = closed markers) and 4 BAC ranges.

(0.095 g/dl) for all pairs. This is a difference of 12% peak to peak. By contrast, in the laboratory-dosing studies where the TAC peak response was muted, paired comparisons within a drinking episode had a mean SCRAM TAC peak (0.047 g/dl), 44% lower than mean peak BAC (0.085 g/dl). The laboratory mean peak difference was 3 times lower than the self-dose peak difference for episodes of drinking that were detected as true positives by SCRAM, even when the BAC peak was nearly the same.

TAC output from the WrisTAS overestimates BAC levels in the self-dose situation, if no correction is made for varying WrisTAS baselines. The paired comparisons of BAC and TAC for WrisTAS hits with self-dosing differ by 0.073 g/dl (0.166 TAC and 0.093 BAC) and represent a WrisTAS TAC higher than peak BACs by 86% (assuming a 0 TAC reading is truly 0, which for WrisTAS it often is not). By contrast in the laboratory situation, for true-positive hits, the WrisTAS devices logged TAC peaks just 0.019 g/dl lower than mean peak BAC (e.g., TAC was 21% lower than BAC). The SCRAM and WrisTAS devices appear to have different optimal BACs when accuracy is highest. SCRAM more accurately estimated self-dosing levels, and WrisTAS more accurately estimated laboratory-dosing levels. Neither device accurately estimated both types of drinking.

Sakai and colleagues (2006) used the Bland-Altman method (1986) to examine SCRAM results. This method of arranging data allows for a visual way to compare 2 types of measurement in order to examine possible underlying scale differences. Slopes that deviate from the horizontal suggest underlying differences in measurement. For this evaluation, data were restricted to results where TAC was judged a "hit" so that scale differences could be evaluated. The average of paired TAC and BAC maximum values are plotted on the x -axis and the difference between them on the y -axis. If TAC

results were reliably related to BAC results across all values, the scatter plot would line up with a slope of 0. The charts in Fig. 2 show a fitted line along with 95% confidence intervals. The overall measured mean difference between BAC and TAC is represented by a shaded, dashed line from the y -axis. The left panel (Fig. 2A) is for WrisTAS and the right panel for SCRAM (Fig. 2B). The evidence suggests that for WrisTAS, the TAC values are positively accelerated relative to BAC as BAC increases. The differences are only a scale error, and a linear adjustment in reported values would likely correct it. It would also help to have a reliable way to connote 0 TAC for WrisTAS. Most of the SCRAM results capture 0 on the y -axis within the 95% confidence intervals. The difference in estimation may represent the different types of events each device is calibrated to detect. SCRAM is calibrated to detect self-dosed illicit drinking in an offender population, and WrisTAS has had extensive trials in laboratory-dosing situations as a treatment research device, a circumstance under which most human subject protocols would preclude study of high-level sustained dosing.

Delay of Peak TAC within True-Positive Hits. There is a commonly reported 2-hour lag between peak BAC and TAC under normal circumstances. This lag has been described in the peer-reviewed research literature and in the materials from the private companies. The delay represents transit time of ethanol from liquid phase in the body core out to the skin surface in the gas phase. On the descending limb of a BAC curve, the TAC curve usually lags behind the BAC curve somewhat, such that the absorption and elimination curves are slightly asymmetrical.

True-positive hits from WrisTAS had a timeframe of results that were more statistically normative than SCRAM true-positive hits. Estimating performance of either device in

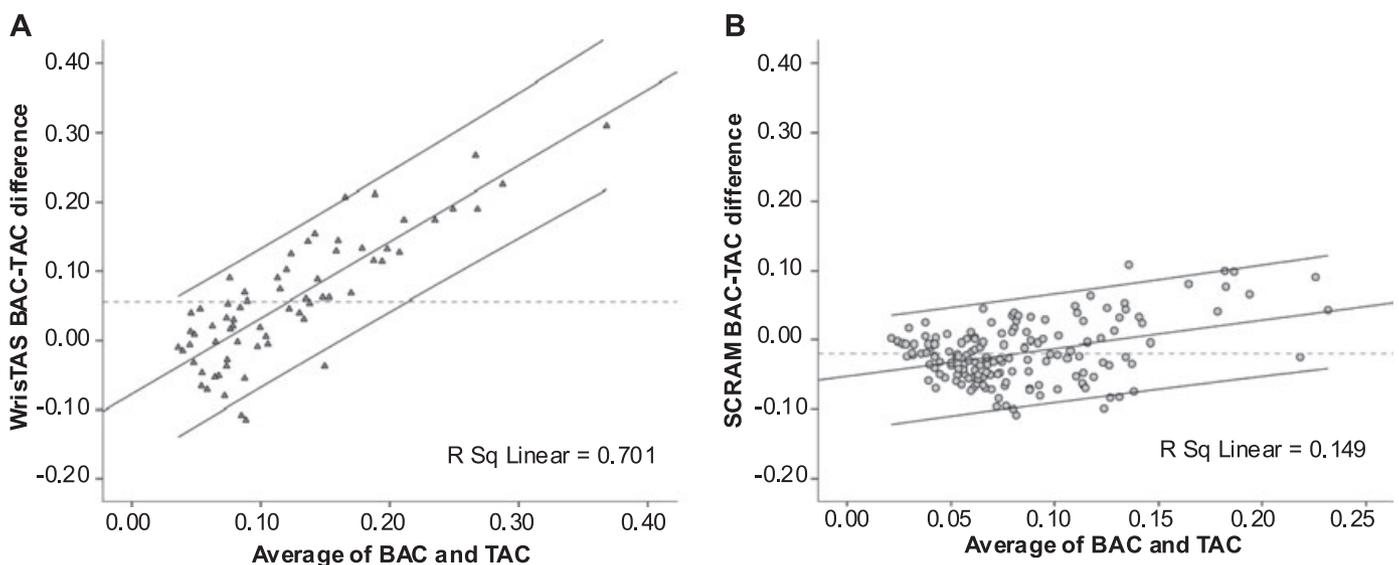


Fig. 2. Bland-Altman type data layout to compare BAC and TAC for WrisTAS (A), and SCRAM (B) over a range of measured values for true-positive hits (WrisTAS, Wrist Transdermal Alcohol Sensor; SCRAM, Secure Continuous Remote Alcohol Monitoring).

regard to time delay between BAC and TAC maxima can only be done with the laboratory-dosed studies where we controlled the alcohol dose. For the WrisTAS, the mean time delay for peak TAC was 2.28 ± 1.5 hour past the maximal BAC. This is very close to the expected range of delay based on the published literature. The SCRAM devices had mean TAC peak delays of 4.5 ± 2.9 hour relative to the BAC peaks. This lengthier delay exceeds by about 2 hours the expected delay in TAC maxima. A likely explanation is water accumulation inside the SCRAM unit that dilutes the ethanol causing peak reduction and delay in recording the TAC. This is a problem known to AMS, and its newer version of SCRAM reportedly solves this problem. However, there has been no independent confirmation reported to date.

An illustration of this problem is portrayed in Fig. 3. This study had individuals hooked to the same transdermal sensor for as long as 4 weeks but sometimes devices had to be swapped out earlier. Figure 3 shows a difference in true-positive detection (mean \pm 1 standard error) rates as a function of duration of wear across 4 intervals defined by maximal BACs attained per episode. The wear duration was dichotomized to place an equal number of drinking episodes in each side of the split ($n = 135$ short wear, $n = 136$ long wear). The short duration represents an average of 3.3 days of wear (dark lines with circles) and long duration an average of 12.5 days of wear (light line with squares).

A logistic regression analysis of SCRAM hit rates using continuous “days of service” for SCRAM bracelets confirmed that days in service is a potent predictor of the likelihood of true positive hits ($p < 0.0001$, Wald = 24.5). The longer a bracelet is in use, the lower its ability to detect BAC. Also, as is evident from Fig. 3, the maximal level of BAC attained is an important determinant of SCRAM true-positive hit rate, and in the regression this too was an important predictor ($p < 0.0001$, Wald = 38.1). Less obvious, a marginally significant factor accounting for true positive SCRAM hits is

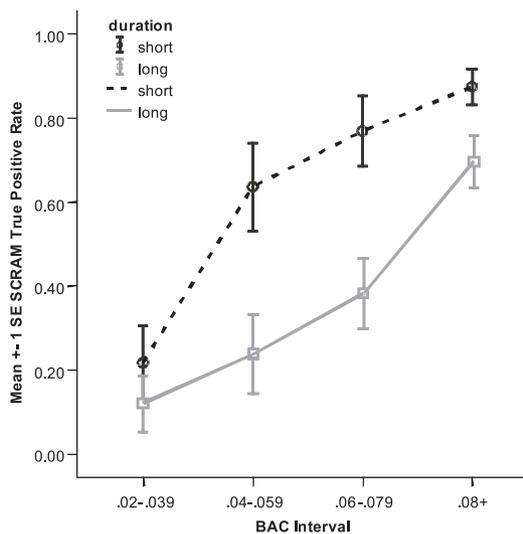


Fig. 3. SCRAM true-positive rate at 4 BAC intervals and 2 durations of device wear (SCRAM, Secure Continuous Remote Alcohol Monitoring).

the gender of the drinker during a drinking event. The mean maximal BAC attained by females (0.073 g/dl) and males (0.080 g/dl) were somewhat (not statistically) different, but even after controlling for maximal BAC attained, gender was still a marginally significant predictor that entered on step 3 of a forward conditional logistic regression equation ($p = 0.055$, Wald = 3.6). That is, controlling for BAC and service days of the devices, female drinking was less readily detected than male drinking.

For WrisTAS, no comparable “days of service” variable was created as described for SCRAM. Nonetheless, the likelihood of true positive WrisTAS hits is very strongly affected by the gender of the drinker, with females ($p < 0.0001$, Wald = 13.3) less likely to be detected even after controlling for maximal BAC attained. Gender of WrisTAS wearer appeared to be more predictive than maximal BAC attained ($p = 0.013$, Wald = 6.1) during a drinking event.

DISCUSSION

In general, the sensitivity and accuracy of these devices was poorer than expected. The WrisTAS had apparent problems with reliably recording or retrieving data, whereas the SCRAM device had apparent problems with water accumulation. Nonetheless, the devices can estimate consumed beverage alcohol as a gas at the skin surface some time after BAC has peaked and with product improvement are likely to better serve their intended function as abstinence monitoring aids. If Anderson and Hlastala (2006) are correct, it seems unlikely that the alcohol signal at the skin surface can be a precise estimate of BAC. Their biophysical model suggests that the stratum corneum, the outermost layer of skin, and other systemic factors, importantly affect the measurable ethanol gas concentration that leaves the skin. Individual differences, gender differences, or state differences within individuals, in hydration, temperature, and other factors theoretically have a large effect on the transdermal alcohol gas. A dermatologic study by Jacobi and colleagues (2005) found gender differences in some physiological characteristics of the stratum corneum. Both transdermal devices tested in this study detected less alcohol gas at the skin relative to BAC in females than in males. Additional findings are described in a National Highway Traffic Safety Administration (NHTSA) report (Marques and McKnight, 2007).

On the other hand, the monitoring of alcohol consumption does not need to depend on precise measurement of BAC; it depends on the ability of a technology to detect abstinence violations as measured by a signal in excess of some minimal amount, such as 0.02 g/dl. For these purposes, estimates suffice. As a monitoring device for offenders, the transdermal concept is valid. Despite the limitations of the actual equipment with false-negatives rates that are too high. These devices warrant further development and study.

The discovery of declining SCRAM accuracy over time is a very significant concern. It is likely that in the SCRAM version tested, the accuracy and the sensitivity problems are

due to liquid water accumulation, an assumption with which the manufacturer concurs. The new SCRAM, version 2, that supposedly solves this problem warrants evaluation to ensure that equipment in service for several months retains adequate sensitivity and accuracy.

The communication of the SCRAM bracelet with its remote server, along with data retrieval and reporting technology, is innovative. This type of automated data upload and transfer to a server is well suited to the alcohol offender monitoring market. **The system has temperature and skin reflectivity sensors to help detect circumvention efforts; our evaluation of these found them to be effective.** The system issues daily alerts to a program monitor when alcohol or tampering is detected; this likely prevents most offenders from beating this system. **The Scramnetwork server works well and proved to be a dependable authorization and data-tracking system.** Users are encouraged to shower and scrub skin under the SCRAM device, but the device cannot be immersed.

Although the performance of the WrisTAS version 5 devices in this study appears to have suffered from poor internal electronics, the sensor technology that is at the heart of the Giner device is inventive and seems capable of adequate alcohol detection. Swift and colleagues (Swift, 2003; Swift et al., 1992) published evidence over a 10-year span that demonstrated the accuracy of the Giner sensor. Earlier versions of the WrisTAS may have had superior electronics (R. Swift, personal communication). In the larger NHTSA report, we reported on an automated signal detection evaluation protocol we devised that did not rely on human coders, only a rule-based detection protocol. With it, the WrisTAS device detected TAC from a BAC of 0.02 g/dl as readily as from 0.08 g/dl, apparently by more expertly decoding some of the erratic readings that human coders could not determine to be true-positive hits. Unlike the SCRAM device, however, the WrisTAS cannot get very wet and has to be removed for showering. Particularly for WrisTAS, but marginally also for SCRAM, female drinking is less readily detected.

Concluding that transdermal alcohol sensing has many more benefits than problems is unlikely to spark a rush to transdermal sensing as a monitoring remedy for alcohol abusive offenders; average daily cost is still a barrier, particularly with DWI offenders. The typical 2006 to 2007 cost of leasing SCRAM was around \$12 a day. This compares very well with regular electronic home confinement monitoring devices, which cost about \$12/day, but quite poorly with alcohol ignition interlock devices that usually cost about \$2.25/day. Accordingly, the use of transdermal sensing is probably going to be less used with routine DWI offenders than for multi-problem alcohol abusers. People who need to be entirely prevented from drinking, not just drinking-and-driving, are the natural market for SCRAM. Conceivably, transdermal technology could make an important contribution to the problem of low penetration of alcohol ignition interlocks among DWI offenders if it were required as an alternative to the interlock for those offenders who are judged by the court to be

unsuited for an interlock. It would provide better monitoring and security than simple license suspension, as is now most typical.

There is a parallel in these early findings about the accuracy of transdermal devices that is reminiscent of the early accuracy of alcohol ignition interlock devices. First generation interlock devices were often criticized for failing to match the performance characteristics of more conventional breath-test devices, despite interlocks having to operate in an often hostile automotive environment of heat, cold, dust, and vibration. Similarly, TAC is not BAC, and the expectation of parity is an impractical standard for a developmental technology. Transdermal sensing devices, like interlocks before them, need to be judged first on their potential contributions to public safety (or alcohol treatment) monitoring. Moreover, just as alcohol-interlock devices have improved dramatically in the 22 years since their first adoption, it is reasonable to expect that the transdermal-sensing equipment will also improve with further development and further study.

ACKNOWLEDGMENTS

The authors wish to acknowledge the support of the National Highway Traffic Safety Administration (Contract #DTNH22-02-D-95121). PIRE colleagues Scott Tippetts, Jim Fell, and Dr. Robert Voas provided helpful advice. In addition, this study was greatly facilitated by the cooperation of the senior managers and technical staff of the Alcohol Monitoring Systems Corporation in Littleton, Colorado, and similarly from Giner Incorporated in Newton, Massachusetts. Dr. Robert Swift of Brown University deserves special mention for providing helpful guidance during the study.

REFERENCES

- Anderson JC, Hlastala MP (2006) The kinetics of transdermal ethanol exchange. *J Appl Physiol* 100:649–655.
- Bland JM, Altman DG (1986) Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1:307–310.
- Borucki K, Schreiner R, Dierkes J, Jachau K, Krause D, Westphal S, Wurst FM, Luley C, Schmidt-Gayk H (2005) Detection of recent ethanol intake with new markers: Comparison of fatty acid ethyl esters in serum and of ethyl glucuronide and the ratio of 5-hydroxytryptophol to 5-hydroxyindole acetic acid in urine. *Alcohol Clin Exp Res* 29:781–787.
- Greenfield TK, Tujague J, Bond J, Kerr W (2005) A pilot study of self-report alcohol consumption: Diary versus transdermal alcohol sensor accuracy. Paper presented at the Research Society of Alcoholism Meeting, Santa Barbara, CA.
- Helander A, Beck O (2005) Ethyl sulfate: a metabolite of ethanol in humans and a potential biomarker of acute alcohol intake. *J Anal Toxicol* 29:270–274.
- Jacobi U, Gautier J, Sterry W, Lademann J (2005) Gender-related differences in the physiology of the stratum corneum. *Dermatology* 211:312–317.
- Marques P, Bjerre B, Dussault C, Voas RB, Beirness DJ, Marples IR, Rauch WR (2001) Alcohol Ignition Interlock Devices—I: Position Paper. International Council on Alcohol, Drugs and Traffic Safety (ICADTS), Oosterhout, the Netherlands. Available at: <http://www.icadts.org/reports/Alcoholinterlockreport.pdf>. Accessed January 1, 2009.

- Marques PR, McKnight AS (2007) Evaluating Transdermal Alcohol Measuring Devices. (DOT HS 810 875). National Highway Traffic Safety Administration, Washington, DC.
- Sakai JT, Mikulich-Gilbertson SK, Long RJ, Crowley TJ (2006) Validity of transdermal alcohol monitoring: Fixed and self-regulated dosing. *Alcohol Clin Exp Res* 30:26–33.
- Swift R (2000) Transdermal alcohol measurement for estimation of blood alcohol concentration. *Alcohol Clin Exp Res* 24:422–423.
- Swift RM (2003) Direct measurement of alcohol and its metabolites. *Addiction* 98:73.
- Swift RM, Martin CS, Swette L, LaConti A, Kackley N (1992) Studies on a wearable, electronic, transdermal alcohol sensor. *Alcohol Clin Exp Res* 16:721–725.
- Voas RB, Marques PR, Tippetts AS, Beirness DJ (1999) The Alberta Interlock Program: The evaluation of a province-wide program on DUI recidivism. *Addiction* 94:1849–1859.
- Widmark EMP (1932) Principles and Applications of Medicolegal Alcohol Determination. Biomedical Publications, Davis, CA.
- Willis C, Lybrand S, Bellamy N (2004) Alcohol ignition interlock programmes for reducing drink driving recidivism. *Cochrane Database of Systematic Reviews*. Art. No.: CD004168. DOI: 10.1002/14651858.CD004168.pub2.
- Wurst FM, Skipper GE, Weinmann W (2003) Ethyl glucuronide—the direct ethanol metabolite on the threshold from science to routine use. *Addiction* 98:51–61.