HHS Public Access

Author manuscript

*Exp Clin Psychopharmacol*. Author manuscript; available in PMC 2025 April 01.

Published in final edited form as:

*Exp Clin Psychopharmacol*. 2024 April ; 32(2): 245–254. doi:10.1037/pha0000683.

# Signal Processing and Machine Learning with Transdermal Alcohol Concentration to Predict Natural Environment Alcohol Consumption

**Nathan A. Didier, M.S.**[1], **Andrea C. King, Ph.D.**[1], **Eric C. Polley, Ph.D.**[2], **Daniel J. Fridberg, Ph.D.**[1]

[1]The University of Chicago, Department of Psychiatry and Behavioral Neuroscience

[2]The University of Chicago, Department of Public Health Sciences

## Abstract

Wrist-worn alcohol biosensors continuously and discreetly record transdermal alcohol concentration (TAC) and may allow alcohol researchers to monitor alcohol consumption in participants' natural environments. However, the field lacks established methods for signal processing and detecting alcohol events using these devices. We developed software that streamlines analysis of raw data (TAC, temperature, and motion) from a wrist-worn alcohol biosensor (BACtrack Skyn) through a signal processing and machine learning pipeline: biologically implausible skin surface temperature readings (<28°C ) were screened for potential device removal and TAC artifacts were corrected, features that describe TAC (e.g. rise duration) were calculated and used to train models (random forest and logistic regression) that predict self-reported alcohol consumption, and model performances were measured and summarized in auto-generated reports. The software was tested using 60 Skyn datasets recorded during 30 alcohol drinking episodes and 30 non-alcohol episodes. Participants (N = 36; 13 with alcohol use disorder) wore the Skyn during one alcohol drinking episode and one non-alcohol drinking episode in their natural environment. In terms of distinguishing alcohol from non-alcohol drinking, correcting artifacts in the data resulted in 10% improvement in model accuracy relative to using raw data. Random forest and logistic regression models were both accurate, correctly predicting 97% (58/60; AUC-ROCs = 0.98, 0.96) of episodes. Area under TAC curve, rise duration of TAC curve, and peak TAC were the most important features for predictive accuracy. With promising model performance, this protocol will enhance the efficiency and reliability of TAC sensors for future alcohol monitoring research.

Correspondence concerning this article should be addressed to Nathan A. Didier, M.S., Department of Psychiatry & Behavioral Neuroscience, The University of Chicago, 5841 S. Maryland Avenue (MC-3077), Chicago, IL 60637. Contact: ndidier@bsd.uchicago.edu.

Disclosures

All authors have read and approved the manuscript for submission and have made substantial contributions to the conception, design, gathering, analysis and interpretation of data, the writing of the article, and acknowledge that they have exercised due care in ensuring the integrity of the work. All authors have no declarations of competing interests to declare.

**Keywords**

transdermal alcohol concentration (TAC); machine learning; natural environment drinking; alcohol monitoring; biosensor

## Introduction

Recent advances in mobile technology, such as smartphones, fitness trackers, and wireless devices, have changed many aspects of daily life. In research and health care, mobile technologies have enhanced assessment of ambulatory biological states and health-related behaviors (Huhn et al., 2022). Regarding the monitoring of alcohol use specifically, wearable biosensors can detect alcohol consumption by assaying perspiration-derived transdermal alcohol concentration (TAC). These devices offer a continuous, nonintrusive, and passive alternative to detect alcohol drinking compared with measures obtained from samples of blood or exhaled breath (Norman et al., 2021; Fairbairn & Bosch, 2021).

The BACtrack Skyn (San Francisco, CA) was first developed in 2016 and is sold directly to consumers as a discreet, minimally intrusive wrist-worn alcohol biosensor with potential application to the criminal justice system, health-minded consumers, and field research. The use of transdermal alcohol biosensors like the Skyn is anticipated to expand (Fairbairn & Kang, 2021) as users report they are non-disruptive and comfortable to wear during sleep, exercise, and other daily activities (Ash et al., 2022; Wang et al., 2019, 2021) and allow ecologically relevant monitoring of real-time alcohol consumption. The Skyn samples TAC, temperature, and motion every 20 seconds to facilitate high resolution analysis of alcohol consumption. Prior work using the Skyn and other devices have focused on converting measured TAC into standard drinks and estimated blood alcohol concentration (eBAC) (Dougherty et al., 2012; Fairbairn et al., 2019; Fairbairn & Kang, 2020; Karns-Wright et al., 2018; Norman et al., 2021, Oszkinat, Shao et al., 2022). However, without reliable methods for signal processing and modeling, TAC cannot be directly translated into estimated alcohol use or eBAC because biological factors (e.g., metabolism and skin properties) and contextual factors (e.g., exercise, environmental temperature, alcohol spills) produce TAC readings that vary across individuals and settings (Barnett et al., 2014; Fairbairn & Kang, 2021; Luczak & Ramchandani, 2019; Swift, 2003; van Egmond et al., 2020). To our knowledge, there are no established best practices for cleaning and analyzing Skyn TAC data to determine and detect alcohol use, thus limiting broader deployment of these devices in scientific research, clinical interventions, or law enforcement.

There have been two published studies to date reporting on predictive models of self-reported alcohol consumption using Skyn-derived TAC. Both studies employed modest sample sizes (11–47 light and heavy drinkers) and participants wore the Skyn for one to four weeks. In one study (Courtney et al., 2023), TAC data were cleaned by removing recordings with skin temperature <29°C (indicating probable device removal) and smoothed using a 15-minute rolling average. TAC values >15μg of alcohol per liter of air (μg/L) were classified as alcohol events and subsequently screened for biological plausibility according to thresholds of duration, peak TAC, and TAC rise rate, resulting in 86.4% specificity for detecting self-

reported non-drinking days and 57.6% sensitivity for detecting self-reported drinking days (Courtney et al., 2023). A separate study used a receiver operating characteristic (ROC) to set a day-level threshold of area under TAC curve (AUC-TAC) that achieved 90% specificity, resulting in 69% sensitivity to detect self-reported alcohol drinking over 101 days (Ash et al., 2022). Of note, this study analyzed raw data (i.e., uncleaned and unsmoothed) but did examine temperature readings to identify potential device removal. While both studies showed initial promise regarding the feasibility of using the Skyn to monitor alcohol use in participants' natural environments, accuracy was hindered by insufficient data cleaning and predictive modeling.

There are three significant challenges when developing a reliable procedure for detecting alcohol drinking episodes from biosensor-derived TAC. First, the TAC signal is often noisy, which can be exacerbated by environmental contamination (e.g., beverage spills, hand sanitizer), device removal, or excessive movement (van Egmond et al., 2020; Wang et al., 2021). Spikes in the TAC signal caused by artifacts may lead models to incorrectly predict alcohol consumption, so it is important to develop reliable methods for removing noise from the data (van Egmond et al., 2020; Wang et al., 2021). Second, there are no standardized tools or protocols for feature engineering, i.e., calculating statistics that describe TAC datasets and curves, such as AUC-TAC, peak TAC, and TAC rise/fall rate (Russell et al., 2022). While the precise features of interest may vary according to the study design and hypotheses, clear definitions and procedures for deriving TAC features will be useful to aid replicability and streamline analyses. Third, classification models are needed to discriminate alcohol drinking from non-alcohol drinking episodes. Effective models will likely come from training multivariable versus single-variable models because additional variables (i.e., TAC features) will help models more accurately predict alcohol consumption (Fairbairn & Bosch, 2021). Classification models, such as random forest and logistic regression, are designed to manage the complexity of multiple features by algorithmically assessing their interactions and calibrating parameters to achieve optimal predictive utility. Machine learning, which involves training models and using them to make predictions on novel data, is a promising technique for evaluating the predictive utility of TAC features and TAC-based models (Fairbairn & Bosch, 2021; Piasecki, 2019). To date, however, no studies using the Skyn have employed machine learning to develop models that predict self-reported alcohol consumption.

Thus, the goal of this study was to develop a signal processing and machine learning protocol to reliably differentiate alcohol drinking from non-alcohol drinking episodes using Skyn-recorded TAC. We developed software to evaluate the quality of raw Skyn data, detect and correct artifacts, calculate TAC features (AUC-TAC, peak, etc.), and perform machine learning with multivariable models. We tested this software using TAC data collected in an independent sample of 36 participants as part of the second wave of the larger Mobile Alcohol Response Study (Fridberg et al., under review). Participants were regular drinkers across the alcohol use continuum, from light drinkers to those with alcohol use disorder (AUD), and wore the Skyn device for approximately 13 hours before, during, and after one alcohol drinking and one non-alcohol drinking episode in their natural environment. Here, we describe the functions of our software and report on the accuracy of our machine learning approach to detect alcohol consumption from field-derived TAC.

# Methods

## Participants and Screening

Participants were recruited from April 2021 to September 2021 and all study procedures were approved by the University of Chicago Institutional Review Board. This study was funded by NIH (R21-AA029746, R01-AA013746) and is not preregistered anywhere. Participants wore the Skyn device and used their smartphones to record each alcohol or non-alcoholic beverage consumed and subjective responses throughout the three-hour episode as part of a high-resolution ecological momentary assessment (HR-EMA) study protocol (Fridberg et al., in press; see also Fridberg et al., 2019, 2021 for information about the HR-EMA protocol).

Thirty-six participants were enrolled in the study. Study candidates were recruited via online advertisements, flyers, and word-of-mouth referrals. To be consistent with prior work (King et al., 2011), eligible participants were 21–35 years of age, consumed at least one alcohol drink weekly, did not meet criteria for any major medical disorders, had at least 10 years of education, were fluent in English, and owned an Android or iOS smartphone. Study candidates who met these initial criteria were invited to a one-hour screening conducted via secure video conference. The virtual screening visit consisted of informed consent, an online timeline follow-back (O-TLFB) calendar assessing alcohol use over the past month (Sobell & Sobell, 1992; Rueger et al., 2012), health and demographic surveys, the Alcohol Use Disorders Identification Test (Babor et al., 2001), and the AUD module from the Structured Clinical Interview for the DSM-5 (First, 2015). Participants were excluded if they met current or past criteria for a major psychiatric disorder (schizophrenia, bipolar disorder, OCD, panic disorder, eating disorder), endorsed suicidal ideation within the past 6 months, used cannabis >3 times per week or other illicit drugs >2 times in the past month, smoked more than 5 cigarettes per day, had received substance use treatment within the past 6 months, or reported that they were currently trying or planning to abstain from alcohol.

## HR-EMA and Skyn Protocol

Enrolled participants completed a 30-minute orientation session with a research assistant via secure videoconference which oriented them to procedures for using the HR-EMA app (Metricwire, Inc.) and Skyn biosensor (Model T10; firmware 2.0.8). During orientation, participants were instructed that they would complete HR-EMA of their behavior and experiences during two episodes (one alcohol drinking and one non-alcohol drinking) while wearing the Skyn biosensor. The research assistant and participant reviewed the participant's typical alcohol consumption pattern as reported on the TLFB to identify a day of the week when they were likely to consume alcohol, and a day of the week when they were likely to abstain from drinking (Fridberg et al., in press; 2019; 2021). Participants were encouraged to complete the alcohol drinking and non-alcohol drinking episodes on those days in the coming week, respectively. For ethical and safety reasons, participants were neither instructed nor required to consume a specific amount of alcohol during the alcohol drinking episode. For the non-alcohol episode, participants were instructed to consume one or more non-alcoholic beverages and to abstain from alcohol consumption.

At least 30 minutes before the start of each episode, participants attached the Skyn to the wrist of their non-dominant hand and powered on the device. Prior to consuming the first beverage in an episode, participants submitted a photo of their hand/wrist with their beverage and the Skyn clearly visible. Each 3-hour HR-EMA period consisted of 8 surveys (with the first survey self-initiated after finishing the first beverage in the period, and all others auto-initiated) with participants responding to queries about the type and size of each drink consumed since the last prompt interval. Participants were instructed to wear the Skyn continuously starting at least 30 minutes before the episode, throughout the episode, and for at least 8 hours after each episode; this guideline was intended to facilitate capture of descending TAC after the alcohol drinking episode.

### Skyn Acceptability

The day following each episode, participants were queried on the acceptability of using the Skyn: (1) “Overall, the Skyn was easy to use”; (2) “Using the Skyn was intrusive”; (3) “Using the Skyn took too long”; (4) “I would recommend the Skyn to other potential participants”; and (5) “Overall, I was satisfied with using the Skyn.” Each item was rated on a 1 to 5 scale, with 1=“strongly disagree” and 5=“strongly agree” (Fridberg et al., 2019; Wang et al., 2021).

### Compensation

After completing both episodes, participants returned the Skyn devices in pre-paid mailers and were compensated $140 for their participation.

### Open-Source Software

The signal processing and machine learning software, titled Skyn Data Manager, was programmed in Python 3.8 and used Pandas (McKinney, 2010) and SciKit-Learn (Pedregosa et al., 2011). To process the Skyn data for our entire cohort, we activated the software with a single terminal command. Software code, the trained predictive models from this study, and an instructional manual are publicly available for experimental support and testing; see online repository at https://github.com/ndidier3/skyn_data_manager.

### Skyn Datasets, Quality Control, and Dataset Cropping

Each of the 36 participants wore the Skyn during one alcohol drinking episode and one non-alcohol drinking episode, resulting in a total of 72 episodes. Twelve episodes (17%) were not captured by the biosensor due to participant error (forgetting to power on the Skyn [n=6], accidental damage [n=4], < 2 hours of recorded data due to mid-episode battery loss [n=2]), resulting in 60 Skyn datasets recorded from n = 32 participants during the alcohol drinking and non-alcohol drinking episodes. These datasets included TAC (μg/L), temperature (degrees Celsius), and motion (G). While the Skyn records data at 20-second intervals, we chose to download the data with 1-minute intervals to minimize processing burden and simplify interpretation of the time-series data.

For each TAC dataset, we examined temperature and duration of the dataset to verify that the device was worn properly at least 2 hours during the 3-hour HR-EMA episode (Figure 1.A). This 2-hour cutoff was chosen to ensure that TAC datasets were recorded during most of

the HR-EMA episode and that enough data was included to capture important TAC features. As normal skin surface temperature ranges from ~32–36°C (~90–97°F) (Freitas, 2003), we conservatively classified data points with temperature readings <28°C as invalid (2.7% of all data points). Data corresponding to potential device non-wear were removed from 2 datasets with >10% of datapoints outside of the valid temperature range; both datasets retained more than 2 hours of valid data recorded during the HR-EMA period. Using participants' beverage photo submission as a pre-episode timestamp, datasets were cropped so that data began at this timestamp and ended either when the device was removed or at 18 hours after the photo submission (Figure 1.B). Cropping is a customizable and optional software capability.

## Signal Cleaning

TAC data often contain artifacts characterized by large and sudden increases in TAC followed by a steep drop-off (Wang et al., 2021). We implemented a novel algorithm to detect and correct such artifacts by iteratively assessing each TAC value and taking the absolute difference between the current and previous TAC value (ΔTAC). If the natural log of ΔTAC was greater than 50% (minor threshold) or 75% (major threshold) of the natural log of the local peak TAC (i.e., the peak TAC occurring within 100 minutes of the current value), then the current TAC value was considered an artifact. When ΔTAC surpassed the minor threshold, it was considered a minor artifact and only the current TAC value was removed. When ΔTAC surpassed the major threshold, it was considered a major artifact and all subsequent TAC values above a return threshold were removed (For more detail, see Figure 1.C and Figure 2). After removing artifacts, a Gaussian processor imputed the removed values (Jafrasteh et al., 2022). Finally, we used a Savitzky–Golay filter to smooth the corrected datasets (Savitzky & Golay, 1964).

## Feature Engineering

To standardize curve demarcation across datasets and to account for variation in baseline (i.e., pre-drinking episode) TAC values, we set curve thresholds for each of the 60 datasets, calculated as one standard deviation above the mean of the first 15 TAC values. A curve was defined as all consecutive TAC values before and after the peak (i.e., the largest TAC value in the dataset) that exceeded the curve threshold. After demarcating each curve accordingly, we calculated TAC features used in previous studies (Russell et al, 2022, Courtney et al., 2023, Ash et al., 2022), including AUC-TAC, peak, rise duration, fall duration, rise rate, and fall rate. We also calculated several other features that describe a dataset, including average TAC difference, TAC alteration rate, TAC count, and minor/major artifact counts. Additional details on the feature engineering procedure are presented in Figure 3.

## Machine Learning

The final step was machine learning. To train and evaluate models that predict whether participants reported consuming alcohol, we tested two models – an elastic net logistic regression (Zou & Hastie, 2005) and a random forest model (Breiman, 2001) – using a procedure called group k-fold cross-validation (Arlot & Celisse, 2010). This procedure consisted of excluding one participant's feature data, training the model to optimally predict self-report using all other feature data, and then using the trained model to predict whether

the excluded feature data corresponded to self-reported alcohol consumption; this process was repeated for each of the 32 participants, resulting in predictions for all 60 datasets. Because the model is naïve of the study's case-control design, each prediction is made independently; for instance, if the model predicts alcohol was consumed in episode A for participant X, this does not influence the model's prediction regarding episode B for participant X. To evaluate model performance, we calculated true positives (correct predictions of alcohol drinking), true negatives (correct predictions of non-alcohol drinking), false positives (alcohol predictions when participant reported not consuming alcohol), false negatives (non-alcohol predictions when participant reported consuming alcohol), accuracy (correct predictions / all episodes), sensitivity (true positives / alcohol drinking episodes), specificity (true negatives / non-alcohol drinking episodes), and AUC-ROC. To interpret the importance of each feature used in the random forest, we computed the mean decrease in impurity, a measure from 0 to 1 for each feature with higher values indicating greater relative contribution towards predictive accuracy (Sandri & Zuccolotto, 2008). Last, to examine whether signal cleaning improved model performance, we created a set of comparator results by repeating group k-fold cross validation but with models trained using features from raw, unprocessed data instead of corrected data.

## Results

### Participant demographics, drinking behavior, and acceptability outcomes

Of the 36 participants accepted into the study, 32 (89%) participants provided at least one valid Skyn dataset; 28 (78%) participants provided valid datasets for both alcohol and non-alcohol episodes while 4 (11%) only provided a single valid dataset. Participants were mostly white (87%), in their mid-late 20s, and educated (mean 16.7 years of education; Table 1). Participants were diverse regarding drinking characteristics, ranging from 2 to 42 drinks/week (mean=15.5) and 13 (41%) met criteria for AUD (Table 1). Participants with complete Skyn datasets for both the alcohol drinking and non-alcohol drinking episodes (n=28) did not differ from those with only a single dataset (n=4) on demographic or baseline drinking variables.

Valid Skyn datasets corresponded to 30 alcohol drinking episodes and 30 non-alcohol drinking episodes. During alcohol drinking episodes, participants consumed 5.0±3.6 (M±SD) standard drinks (range = 1 to 20). Of these episodes, 16 (53%) were heavy drinking occasions (i.e., consuming 5+ drinks for men, 4+ for women) and 14 (47%) were light-moderate drinking occasions. No participant reported any alcohol consumption during the non-alcohol drinking episodes. Participants endorsed high overall satisfaction with study procedures and found the Skyn minimally intrusive and easy to use (Table 1).

### Artifact Cleaning

The cleaning algorithm detected/imputed major artifacts in 32 (53%) datasets and detected/imputed minor artifacts in 43 (72%) datasets. Out of all TAC values processed, 8.3% were labeled major artifacts and 3.4% were labeled minor artifacts.

### Machine Learning Outcomes

Results of the machine learning analyses are shown in Table 2. In terms of distinguishing alcohol drinking from non-alcohol drinking episodes, random forest and logistic regression models trained with features of corrected vs. raw data resulted in 10% improvement in accuracy. When trained with corrected data, the random forest and logistic regression both correctly predicted whether participants self-reported consuming alcohol in 58 of 60 episodes (96.7%; AUC-ROCs = 0.98, 0.96). The random forest model and logistic regression both had 100% specificity (30/30) to detect non-alcohol drinking episodes. Both models had two false negative predictions associated with the same pair of episodes, resulting in a sensitivity of 93.3% (28/30) to detect alcohol-drinking episodes. Mean decrease in impurity (Figure 4) indicated that AUC-TAC (0.31), rise duration (0.24), peak (0.17), and fall duration (0.09) made the largest contributions in the random forest model.

## Discussion

This study showed that a machine learning approach using TAC features derived from Skyn biosensors (Model T10 with firmware 2.0.8) can distinguish between real-world alcohol drinking and non-alcohol drinking episodes with high accuracy (~97%). The sensitivity rate for detecting alcohol use (93.3%) outperformed prior studies examining the Skyn (Range = 58% to 69%; Ash et al., 2022; Courtney et al., 2023), and other wrist-worn TAC sensors (Range = 40% to 86%; Bond et al., 2014; Brobbin et al., 2022; Croff et al., 2020; Simons et al., 2015). While TAC sensors can be insensitive to low levels of alcohol consumption (Karns-Wright et al., 2018; Roache et al., 2015; Piasecki, 2019; Wang et al., 2019), our best model detected alcohol in 13 of 14 light-moderate drinking episodes and similarly detected alcohol in 15 of 16 heavy drinking episodes. Increased sensitivity often comes at the cost of specificity (Ash et al., 2022, Roache et al., 2019), but this highly sensitive model performed with perfect specificity, i.e., never predicted alcohol consumption when the participant reported not drinking alcohol.

With promising model performance, the present study's signal processing and machine learning protocol may contribute to the alcohol biosensor field in several important ways. First, it highlights the value of developing a cleaning algorithm for removing noise and artifacts from field-recorded TAC data prior to feature engineering; without using our cleaning procedure, model prediction accuracy dropped by 10%. Rather than determining whether a TAC curve is an artifact by applying uniform thresholds to TAC features (Courtney et al, 2023), we assessed each TAC value to determine whether its ΔTAC accounts for an unwarranted proportion of the local peak. Comparing ΔTAC to the local peak is important, as a large ΔTAC may not indicate an artifact unless it contributes to most or all of the local peak. This cleaning approach pinpoints the characteristic feature of TAC artifacts - a sudden and large change in TAC (Wang et al., 2019) - while also accounting for individualized rates of metabolism and skin properties that produce variability across TAC datasets (Barnett et al., 2014; Fairbairn & Kang, 2021; Luczak & Ramchandani, 2019; Swift, 2003; van Egmond et al., 2020). We also designated a minor and major threshold such that more severe artifacts (i.e., greater ΔTAC in proportion to local peak) would undergo more extensive cleaning. After removing minor and major artifacts, we filled in the resultant gaps

of data using Gaussian imputation (Jafrasteh et al., 2022). Notably, this cleaning algorithm capitalizes on the computational efficiency of automation via programmatic scripts, as manually assessing each ΔTAC would be impractical. Parameters used for minor and major thresholds were tuned via trial-and-error, i.e., parameters were repeatedly modified by the software developer until visual comparison of TAC plots before and after cleaning showed that artifacts were removed. For future implementations, these cutoffs could be calibrated as a part of machine learning training.

Second, assessing temperature readings can help to corroborate that the device was worn and reduce the number of false negative TAC readings (Ash et al., 2022; Courtney et al., 2023). A biologically implausible skin surface temperature, such as <28°C, provided a threshold for us to detect potential device removal. However, it is possible that ambient temperature may influence temperature readings and thereby limit the reliability of such a threshold, either by disguising non-wear in warm climates or falsely indicating non-wear in cold climates. A more robust method may be achieved by training models to identify device removal through multivariable analysis of temperature and motion features, similar to how models were developed in this study to identify alcohol consumption.

Third, the current study implemented novel approaches to engineering features that describe TAC datasets. We used HR-EMA timestamps to crop data before drinking initiation and after the device was removed. After cropping and correcting datasets, TAC curves were demarcated relative to baseline values using a curve threshold, enabling us to engineer features that describe specific aspects of the curve (e.g., AUC-TAC, rise duration, etc.). Engineering features within shorter/cropped time frames may have improved model efficacy, as features calculated from larger time frames (e.g., day-level) are likely to capture more non-alcohol data and weaken sensitivity to detect alcohol consumption. Further, using a baseline-adjusted curve threshold to determine the beginning and end of each curve helped to normalize the features and correct for confounding effects caused by TAC variability across context and devices, i.e., the zero-value for TAC varies across datasets (Wang et al., 2019; Fairbairn & Bosch, 2021).

Fourth, this is the first study to our knowledge that trained multivariable models (versus single-variable models) with Skyn-derived TAC features to predict self-reported drinking. AUC-TAC, rise duration, peak, and fall duration all provided important information that differentiated alcohol drinking from non-alcohol drinking episodes. Though this study engineered several novel features (e.g., average TAC difference), these features contributed less to random forest predictions than more established features like AUC-TAC. For testing our models, instead of splitting once into training and testing datasets (Ash et al., 2022) which limits predictions to the testing dataset, we made predictions on all 60 valid episodes by implementing group k-fold cross-validation (Steyerberg, 2018). As in prior work (Ash et al., 2022, Courtney et al., 2023), we treated participant self-report of alcohol consumption as the ‘ground truth’ for model predictions. To reduce participant burden and increase reliability of self-report, participants wore the Skyn during one alcohol drinking episode and one non-alcohol drinking episode, rather than wearing the device for weeks or months. Nonetheless, participants could have provided incorrect reports of alcohol use or non-use, which may explain the 2 false negative predictions associated with self-reported alcohol

drinking episodes completed by participants with AUD. Their corresponding TAC curves were visually flat with low peak and low AUC-TAC, so it is reasonable that the models predicted these episodes to be non-alcohol drinking. Besides false self-report, such incorrect predictions could have resulted from insufficient training data or features, undetected device removal, or Skyn device hardware failure.

There are some limitations and next steps worth mentioning. This study's case-control design - each participant completing one alcohol and one non-alcohol drinking episode - optimized for the development of binary models that determine whether alcohol was consumed (or not). The present models were trained with equally represented conditions (30 alcohol, 30 non-alcohol), so it will be important to confirm model performance under different conditions where either alcohol or non-alcohol episodes predominate. Expanding the software to make multi-class predictions (e.g., heavy drinking versus light drinking) is an important next step, as heavy drinking episodes place individuals at high risk of alcohol-related harms (Davis-Martin et al., 2021). Larger samples recorded across diverse contexts and drinking intensities will help facilitate the development of multi-class predictive models and confirm the present models' performance across environmental, physical, and biological variation. Similarly, predictions were made retroactively rather than in real-time, and were based upon several consecutive hours of TAC data. Thus, at this time, the current approach may be more useful for health tracking applications (e.g., identifying whether someone consumed alcohol on a given day or not) rather than for "in the moment" interventions. Related, the software described here was not designed to provide an index of participants' level of intoxication (i.e., eBAC). Recent work has focused on developing and refining models to generate eBAC from TAC recorded in the laboratory or field (Fairbairn et al., 2019; 2020; Luczak et al., 2018; Oszkinat, Luczak, et al., 2022), a process that is complicated by the potential for individual differences (e.g., skin thickness, rate of alcohol metabolism, sex) or environmental factors (e.g., ambient temperature and humidity) that confound the association between TAC and eBAC (Fridberg et al., 2022; Wang et al., 2019). Last, 12 episodes were not recorded by the Skyn due to participant error, indicating that participants may need additional training and guidance to reliably provide valid TAC data.

In conclusion, our signal processing and machine learning protocol demonstrated promising results for cleaning Skyn-derived TAC and distinguishing real-world alcohol drinking from non-alcohol drinking events. Importantly, the software described in this paper automates data quality checks, cropping, artifact detection and cleaning, smoothing, feature engineering, and machine learning for TAC data recorded by the Skyn biosensor. This software is capable of processing whole cohorts of data, advancing the ability for researchers to efficiently identify alcohol consumption in participants' natural environment. Future directions include training models to predict device removal, drinking intensity, or BrAC, examining whether demographics or other biomarkers contribute to model accuracy (Richards et al., 2023), and continuing to validate the present models with TAC data from diverse environments and samples.

## Acknowledgements

This research was supported by NIH grant R21-AA029746 (DJF, ACK) and R01-AA013746 (ACK). The funding source had no involvement in the study, other than providing financial support.

## References

Arlot S, & Celisse A (2010). A survey of cross-validation procedures for model selection. Statistics Surveys, 4(none), 40–79. 10.1214/09-SS054

Ash GI, Gueorguieva R, Barnett NP, Wang W, Robledo DS, DeMartini KS, Pittman B, Redeker NS, O'Malley SS, & Fucito LM (2022). Sensitivity, specificity, and tolerability of the BACTrack Skyn compared to other alcohol monitoring approaches among young adults in a field-based setting. Alcoholism: Clinical and Experimental Research, 46(5), 783–796. 10.1111/acer.14804 [PubMed: 35567595]

Babor T, Higgins-Biddle J, Saunders J, & Monteiro M (2001). AUDIT: The Alcohol Use Disorders Identification Test: Guidelines for use in primary care (2nd ed.). Geneva, Switzerland: World Health Organization.

Barnett NP, Meade EB, & Glynn TR (2014). Predictors of detection of alcohol use episodes using a transdermal alcohol sensor. Experimental and Clinical Psychopharmacology, 22(1), 86–96. 10.1037/a0034821 [PubMed: 24490713]

Bond JC, Greenfield TK, Patterson D, & Kerr WC (2014). Adjustments for Drink Size and Ethanol Content: New Results from a Self-Report Diary and Transdermal Sensor Validation Study. Alcohol: Clinical and Experimental Research, 38(12), 3060–3067. 10.1111/acer.12589

Breiman L (2001). Random Forests. Machine Learning, 45(1), 5–32. 10.1023/A:1010933404324

Brobbin E, Deluca P, Hemrage S, & Drummond C (2022). Accuracy of Wearable Transdermal Alcohol Sensors: Systematic Review. Journal of Medical Internet Research, 24(4), e35178. 10.2196/35178

Courtney JB, Russell MA, & Conroy DE (2023). 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, 108, 30–43. 10.1016/j.alcohol.2022.11.004 [PubMed: 36473634]

Croff JM, Hartwell ML, Chiaf AL, Crockett EK, & Washburn IJ (2021). Feasibility and reliability of continuously monitoring alcohol use among female adolescents and young adults. Drug and Alcohol Review, 40(7), 1143–1154. 10.1111/dar.13045 [PubMed: 32185847]

Davis-Martin RE, Alessi SM, & Boudreaux ED (2021). Alcohol Use Disorder in the Age of Technology: A Review of Wearable Biosensors in Alcohol Use Disorder Treatment. Frontiers in Psychiatry, 12. https://www.frontiersin.org/articles/10.3389/fpsyt.2021.642813

Dougherty DM, Charles NE, Acheson A, John S, Furr RM, & Hill-Kapturczak N (2012). Comparing the detection of transdermal and breath alcohol concentrations during periods of alcohol consumption ranging from moderate drinking to binge drinking. Experimental and Clinical Psychopharmacology, 20, 373–381. 10.1037/a0029021 [PubMed: 22708608]

Fairbairn CE, & Bosch N (2021). A new generation of transdermal alcohol biosensing technology: Practical applications, machine -learning analytics and questions for future research. Addiction, 116(10), 2912–2920. 10.1111/add.15523 [PubMed: 33908674]

Fairbairn CE, & Kang D (2019). Temporal Dynamics of Transdermal Alcohol Concentration Measured via New-Generation Wrist-Worn Biosensor. Alcohol: Clinical and Experimental Research, 43(10), 2060–2069. 10.1111/acer.14172

Fairbairn CE, & Kang D (2021). Chapter 24 - Transdermal alcohol monitors: Research, applications, and future directions. In Frings D & Albery IP (Eds.), The Handbook of Alcohol Use (pp. 551–562). Academic Press. 10.1016/B978-0-12-816720-5.00014-1

Fairbairn CE, Kang D, & Bosch N (2020). Using machine learning for real-time BAC estimation from a new-generation transdermal biosensor in the laboratory. Drug and Alcohol Dependence, 216, 108205. 10.1016/j.drugalcdep.2020.108205

Fairbairn CE, Rosen IG, Luczak SE, & Venerable WJ (2019). Estimating the quantity and time course of alcohol consumption from transdermal alcohol sensor data: A combined laboratory-ambulatory study. Alcohol, 81, 111–116. 10.1016/j.alcohol.2018.08.015 [PubMed: 30179707]

First MB (2015). Structured Clinical Interview for the DSM (SCID). In The Encyclopedia of Clinical Psychology (pp. 1–6). John Wiley & Sons, Ltd. 10.1002/9781118625392.wbecp351

Freitas RA (2003). Nanomedicine, Volume I: Basic Capabilities. CRC Press. 10.1201/9781498712576

Fridberg DJ, Cao D, & King AC (2021). Alcohol subjective responses in heavy drinkers: Measuring acute effects in the natural environment versus the controlled laboratory setting. Alcoholism: Clinical and Experimental Research, 45(6), 1287–1297. [PubMed: 33864396]

Fridberg DJ, Faria J, Cao D, & King AC (2019). Real-Time Mobile Monitoring of Drinking Episodes in Young Adult Heavy Drinkers: Development and Comparative Survey Study. JMIR MHealth and UHealth, 7(11), e13765. 10.2196/13765

Fridberg DJ, Wang Y, & Porges E (2022). Examining features of transdermal alcohol biosensor readings: A promising approach with implications for research and intervention. Alcohol: Clinical and Experimental Research, 46(4), 514–516. 10.1111/acer.14794

Fridberg DJ, Lee Z, Fischer AM, Cursio JF, & King AC (in press). Assessing real-time positive subjective effects of alcohol using high-resolution ecological momentary assessment (HR-EMA) in risky versus light drinkers. Alcohol: Clinical and Experimental Research.

Huhn S, Axt M, Gunga H-C, Maggioni MA, Munga S, Obor D, Sié A, Boudo V, Bunker A, Sauerborn R, Bärnighausen T, & Barteit S (2022). The Impact of Wearable Technologies in Health Research: Scoping Review. JMIR MHealth and UHealth, 10(1), e34384. 10.2196/34384

Jafrasteh B, Hernández-Lobato D, Lubián-López SP, & Benavente-Fernández I (2022). Gaussian Processes for Missing Value Imputation (arXiv:2204.04648). arXiv. 10.48550/arXiv.2204.04648

Karns-Wright TE, Dougherty DM, Hill-Kapturczak N, Mathias CW, & Roache JD (2018). The correspondence between transdermal alcohol monitoring and daily self-reported alcohol consumption. Addictive Behaviors, 85, 147–152. 10.1016/j.addbeh.2018.06.006 [PubMed: 29910035]

King AC, de Wit H, McNamara PJ, & Cao D (2011). Rewarding, stimulant, and sedative alcohol responses and relationship to future binge drinking. Archives of General Psychiatry, 68(4), 389–399. 10.1001/archgenpsychiatry.2011.26 [PubMed: 21464363]

Luczak SE, Hawkins AL, Dai Z, Wichmann R, Wang C, & Rosen IG (2018). Obtaining continuous BrAC/BAC estimates in the field: A hybrid system integrating transdermal alcohol biosensor, Intellidrink smartphone app, and BrAC Estimator software tools. Addictive Behaviors, 83, 48–55. 10.1016/j.addbeh.2017.11.038 [PubMed: 29233567]

Luczak SE, & Ramchandani VA (2019). Special issue on alcohol biosensors: Development, use, and state of the field: Summary, conclusions, and future directions. Alcohol, 81, 161–165. 10.1016/j.alcohol.2019.07.001 [PubMed: 31299292]

McKinney W (2010). Data structures for statistical computing in Python. 56–61. Proceedings of the 9th Python in Science Conference. 10.25080/Majora-92bf1922-00a

Norman T, Peacock A, Ferguson SG, Kuntsche E, & Bruno R (2021). Combining transdermal and breath alcohol assessments, real-time drink logs and retrospective self-reports to measure alcohol consumption and intoxication across a multi-day music festival. Drug and Alcohol Review, 40(7), 1112–1121. 10.1111/dar.13215 [PubMed: 33174260]

Oszkinat C, Luczak SE, & Rosen IG (2022). Uncertainty Quantification in Estimating Blood Alcohol Concentration From Transdermal Alcohol Level With Physics-Informed Neural Networks. IEEE Transactions on Neural Networks and Learning Systems, PP. 10.1109/TNNLS.2022.3140726

Oszkinat C, Shao T, Wang C, Rosen IG, Rosen AD, Saldich EB, & Luczak SE (2022). Blood and breath alcohol concentration from transdermal alcohol biosensor data: Estimation and uncertainty quantification via forward and inverse filtering for a covariate-dependent, physics-informed, hidden Markov model. Inverse Problems, 38(5), 055002. 10.1088/1361-6420/ac5ac7

Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, & Duchesnay É (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 12(85), 2825–2830.

Piasecki TM (2019). Assessment of Alcohol Use in the Natural Environment. Alcohol: Clinical and Experimental Research, 43(4), 564–577. 10.1111/acer.13975

Richards VL, Wang Y, Porges EC, Gullett JM, Leeman RF, Zhou Z, Barnett NP, & Cook RL (20230116). Using alcohol biosensors and biomarkers to measure changes in drinking: Associations between transdermal alcohol concentration, phosphatidylethanol, and self-report in a contingency management study of persons with and without HIV. Experimental and Clinical Psychopharmacology. 10.1037/pha0000637

Roache JD, Karns-Wright TE, Hill-Kapturczak N, Mullen J, Liang Y, Lamb RJ, & Dougherty DM (2015). Using Transdermal Alcohol Monitoring to Detect Low-Level Drinking. Alcohol: Clinical and Experimental Research, 39(7), 1120–1127. 10.1111/acer.12750

Roache JD, Karns-Wright TE, Goros M, Hill-Kapturczak N, Mathias CW, & Dougherty DM (2019). Processing transdermal alcohol concentration (TAC) data to detect low-level drinking. Alcohol, 81, 101–110. 10.1016/j.alcohol.2018.08.014 [PubMed: 30179708]

Rueger SY, Trela CJ, Palmeri M, & King AC (2012). Self-administered web-based timeline followback procedure for drinking and smoking behaviors in young adults. Journal of Studies on Alcohol and Drugs, 73(5), 829–833. 10.15288/jsad.2012.73.829 [PubMed: 22846247]

Russell MA, Turrisi RJ, & Smyth JM (2022). Transdermal sensor features correlate with ecological momentary assessment drinking reports and predict alcohol-related consequences in young adults' natural settings. Alcoholism, Clinical and Experimental Research, 46(1), 100–113. 10.1111/acer.14739 [PubMed: 35066894]

Sandri M, & Zuccolotto P (2008). A Bias Correction Algorithm for the Gini Variable Importance Measure in Classification Trees. Journal of Computational and Graphical Statistics, 17(3), 611–628.

Savitzky Abraham., & Golay MJE (1964). Smoothing and Differentiation of Data by Simplified Least Squares Procedures. Analytical Chemistry, 36(8), 1627–1639. 10.1021/ac60214a047

Simons JS, Wills TA, Emery NN, & Marks RM (2015). Quantifying alcohol consumption: Self-report, transdermal assessment, and prediction of dependence symptoms. Addictive Behaviors, 50, 205–212. 10.1016/j.addbeh.2015.06.042 [PubMed: 26160523]

Sobell LC, & Sobell MB (1992). Timeline Follow-Back. In Litten RZ & Allen JP (Eds.), Measuring Alcohol Consumption: Psychosocial and Biochemical Methods (pp. 41–72). Humana Press. 10.1007/978-1-4612-0357-5_3

Steyerberg EW (2018). Validation in prediction research: The waste by data splitting. Journal of Clinical Epidemiology, 103, 131–133. 10.1016/j.jclinepi.2018.07.010 [PubMed: 30063954]

Swift R (2003). Direct measurement of alcohol and its metabolites. Addiction (Abingdon, England), 98 Suppl 2, 73–80. 10.1046/j.1359-6357.2003.00605.x [PubMed: 14984244]

van Egmond K, Wright CJC, Livingston M, & Kuntsche E (2020). Wearable Transdermal Alcohol Monitors: A Systematic Review of Detection Validity, and Relationship Between Transdermal and Breath Alcohol Concentration and Influencing Factors. Alcohol: Clinical and Experimental Research, 44(10), 1918–1932. 10.1111/acer.14432

Wang Y, Fridberg DJ, Leeman RF, Cook RL, & Porges EC (2019). Wrist-worn alcohol biosensors: Strengths, limitations, and future directions. Alcohol, 81, 83–92. 10.1016/j.alcohol.2018.08.013 [PubMed: 30179709]

Wang Y, Fridberg DJ, Shortell DD, Leeman RF, Barnett NP, Cook RL, & Porges EC (2021). Wrist-worn alcohol biosensors: Applications and usability in behavioral research. Alcohol, 92, 25–34. 10.1016/j.alcohol.2021.01.007 [PubMed: 33609635]

Zou H, & Hastie T (2005). Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 67(2), 301–320. 10.1111/j.1467-9868.2005.00503.x

**Public Significance:**

This study developed a method for processing and analyzing transdermal alcohol concentration from a wrist-worn biosensor. Using this procedure, we predicted whether participants consumed alcohol in their natural environment with 97% accuracy. The software and techniques described here will enhance future clinical and experimental research using alcohol biosensors to monitor drinking behavior.

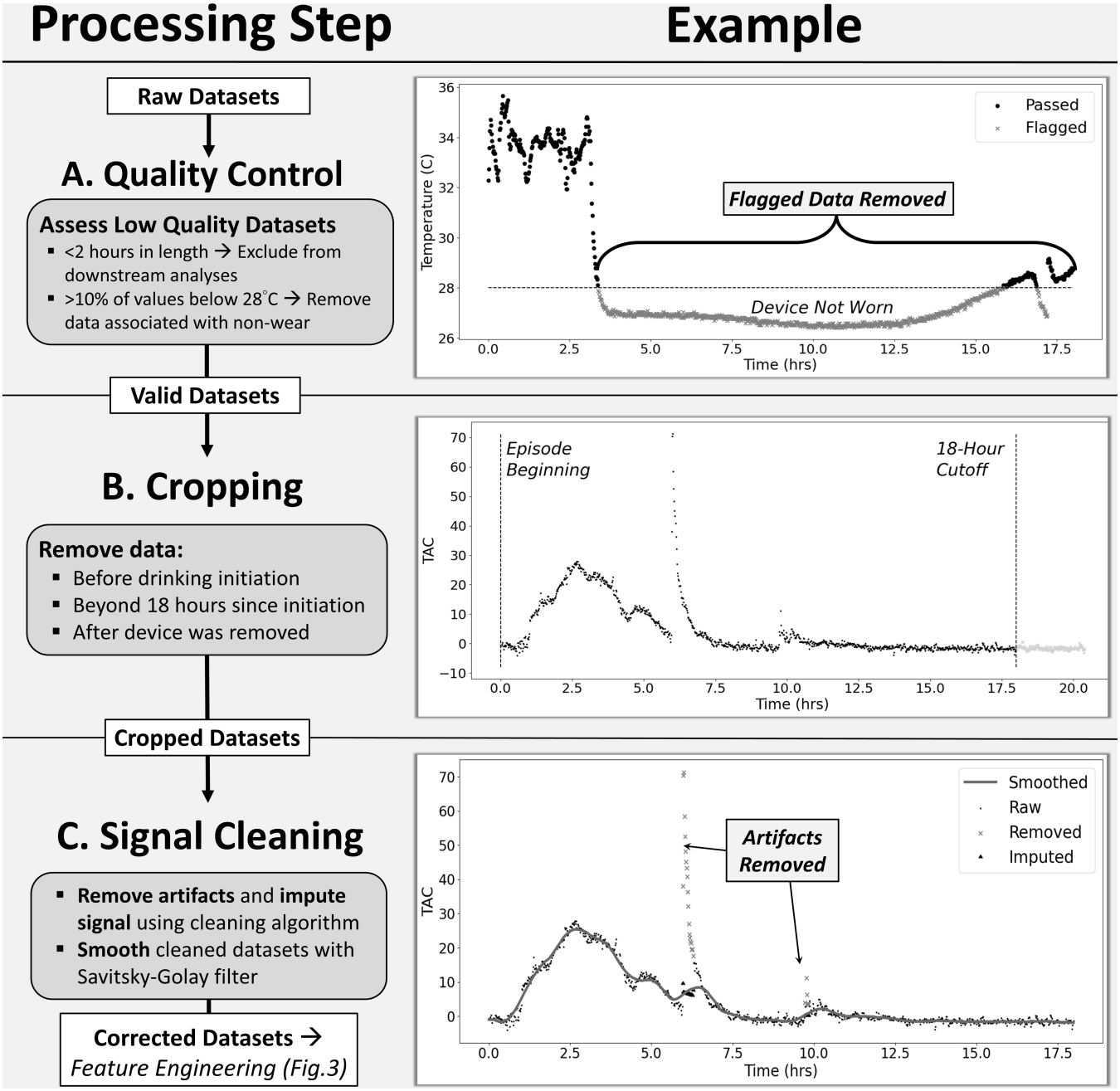



**Figure 1.**

HR-EMA = high-resolution ecological momentary assessment; TAC = transdermal alcohol concentration. The top example (A) shows how potential device non-wear was detected through visual inspection of temperature readings. The middle (B) and bottom (C) examples illustrate how TAC was cropped and cleaned; the data set represented in these examples are also represented in Figures 2 and 3.

## Cleaning Algorithm: Description and Examples

**For each TAC value in a dataset:**

1. Calculate Δ **TAC:** absolute difference between current TAC value ($TAC_i$) and previous TAC value ($TAC_{i-1}$)
   - $\Delta TAC = |TAC_i - TAC_{i-1}|$
2. Identify the greatest TAC value within a 100-minute period (**Local Peak**).
3. ΔTAC reaches a **Major Threshold** when:
   - ✓ Local Peak > 20 μg/L
   - ✓ Natural Log of ΔTAC accounts for 75% or more of the Natural Log of Local Peak
4. ΔTAC reaches a **Minor Threshold** when:
   - ✓ Local Peak > 5 μg/L
   - ✓ Natural Log of ΔTAC accounts for 50% or more of the Natural Log of Local Peak
   - ✓ Major Threshold was not reached
5. If Major Threshold reached:
   a. Remove all subsequent TAC values until TAC drops below **Return Threshold**, calculated as
      - $(Direction\ x\ Slope\ x\ N_{removed}) + (TAC_{i-1}) + (Direction\ x\ 3)$
      - ***Slope*** calculated using a linear regression fit of the previous 6 values
      - ***Direction*** = 1 if $TAC_i > TAC_{i-1}$, *else* ***Direction*** = -1
   b. Train a Gaussian process model using the TAC values before and after the removed value(s).
   c. Impute for removed value(s) using predictions generated by the Gaussian process model
6. If Minor Threshold reached: Remove only $TAC_i$ and apply steps 5b and 5c.

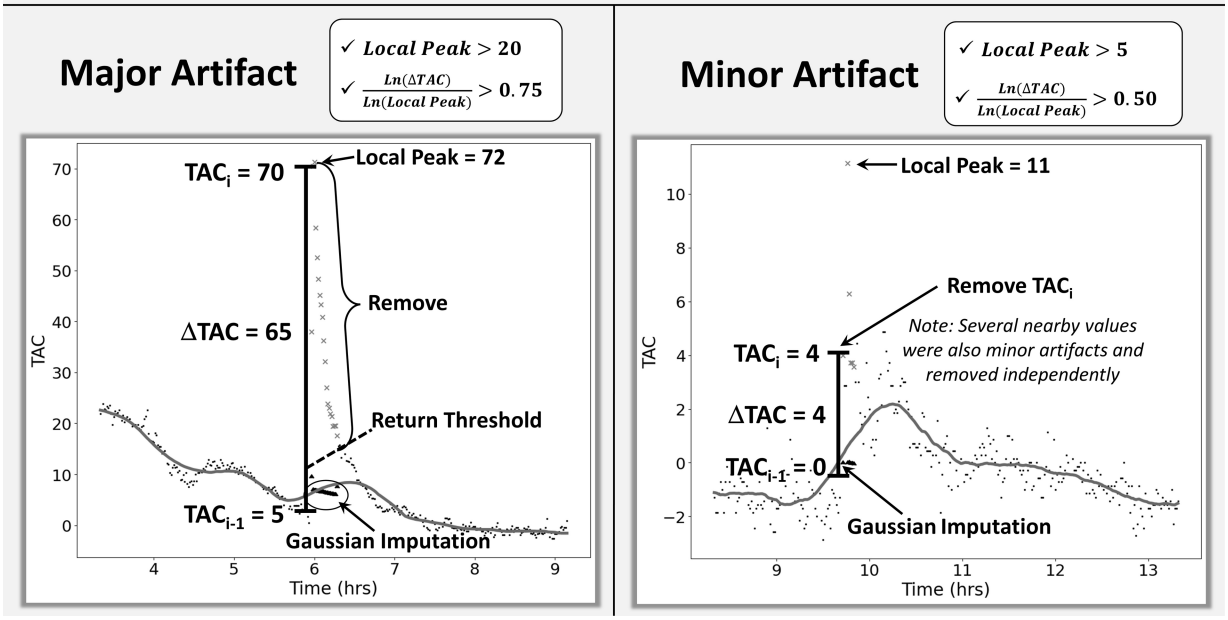


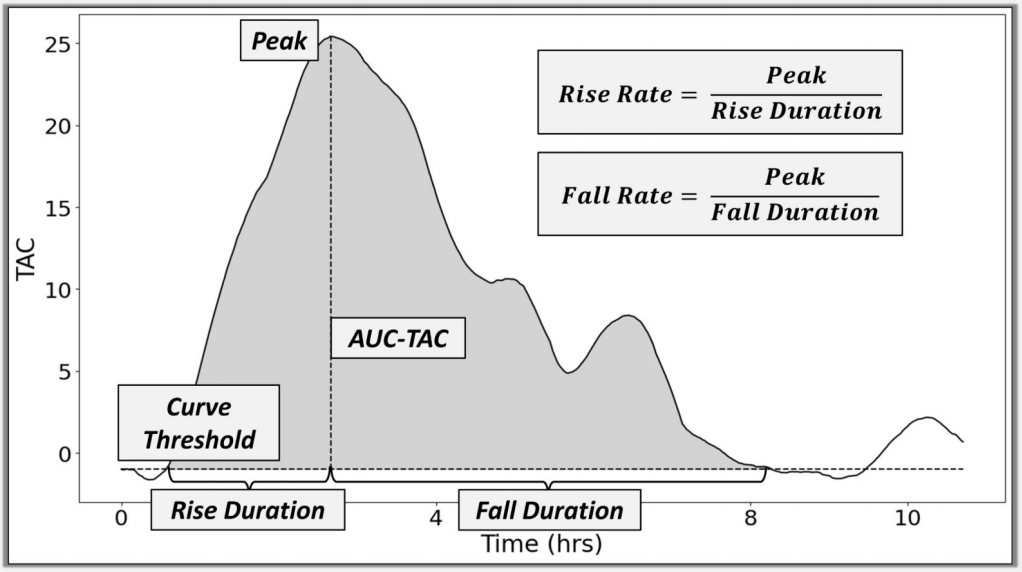

**Figure 2.**

TAC = transdermal alcohol concentration. When a major artifact is detected (left), a series of subsequent values may be removed for that cleaning iteration. Once TAC crosses the return threshold, major artifact cleaning is finished. In the return threshold, the slope is designed to project where TAC was heading if the signal was not obscured. The direction (±1) causes the return threshold to slope upward if the artifact is a spike and slope downward if the artifact is a drop. When a minor artifact is detected (right), only the data point assessed for that

iteration is removed. As shown in the minor artifact example, nearby data points may also be removed on separate iterations.

## Feature Engineering: Steps, Illustration, and Definitions

- **Demarcate curve** using Peak and Baseline Values
  1. Calculate the **Baseline Mean** and **Baseline SD** using first 15 values in dataset
  2. Calculate the **Curve Threshold:** Baseline Mean + Baseline SD
  3. Identify the **Peak** TAC: the largest value in the dataset
  4. Demarcate curve: all values before and after Peak that are above Curve Threshold
- **Calculate curve features**: Peak, Rise Duration, Rise Rate, Fall Duration, Fall Rate, Area Under TAC Curve (AUC-TAC)
- **Calculate dataset features**: Average TAC Difference, TAC Alteration Percent, TAC Count, Minor Outliers, Major Outliers




| Feature | Definition |
|---|---|
| Peak | The largest TAC value in a dataset |
| Rise Duration | Time from start of curve (first value above curve threshold) to Peak |
| Fall Duration | Time from Peak to end of curve (first value since peak that drops below curve threshold) |
| Rise Rate | Peak / Rise Duration |
| Fall Rate | Peak / Fall Duration |
| AUC-TAC | Area under TAC curve |
| Average TAC Difference | Average absolute difference between each pair of consecutive TAC values |
| TAC Alteration Percent | Percentage of consecutive TAC values that change direction from previous pair of TAC values |
| TAC Count | The number of TAC values in the dataset |
| Minor Outliers | The number of minor artifacts removed from the dataset |
| Major Outliers | The number of major artifacts removed from the dataset |

**Figure 3.**
AUC = area under curve; TAC = transdermal alcohol concentration.

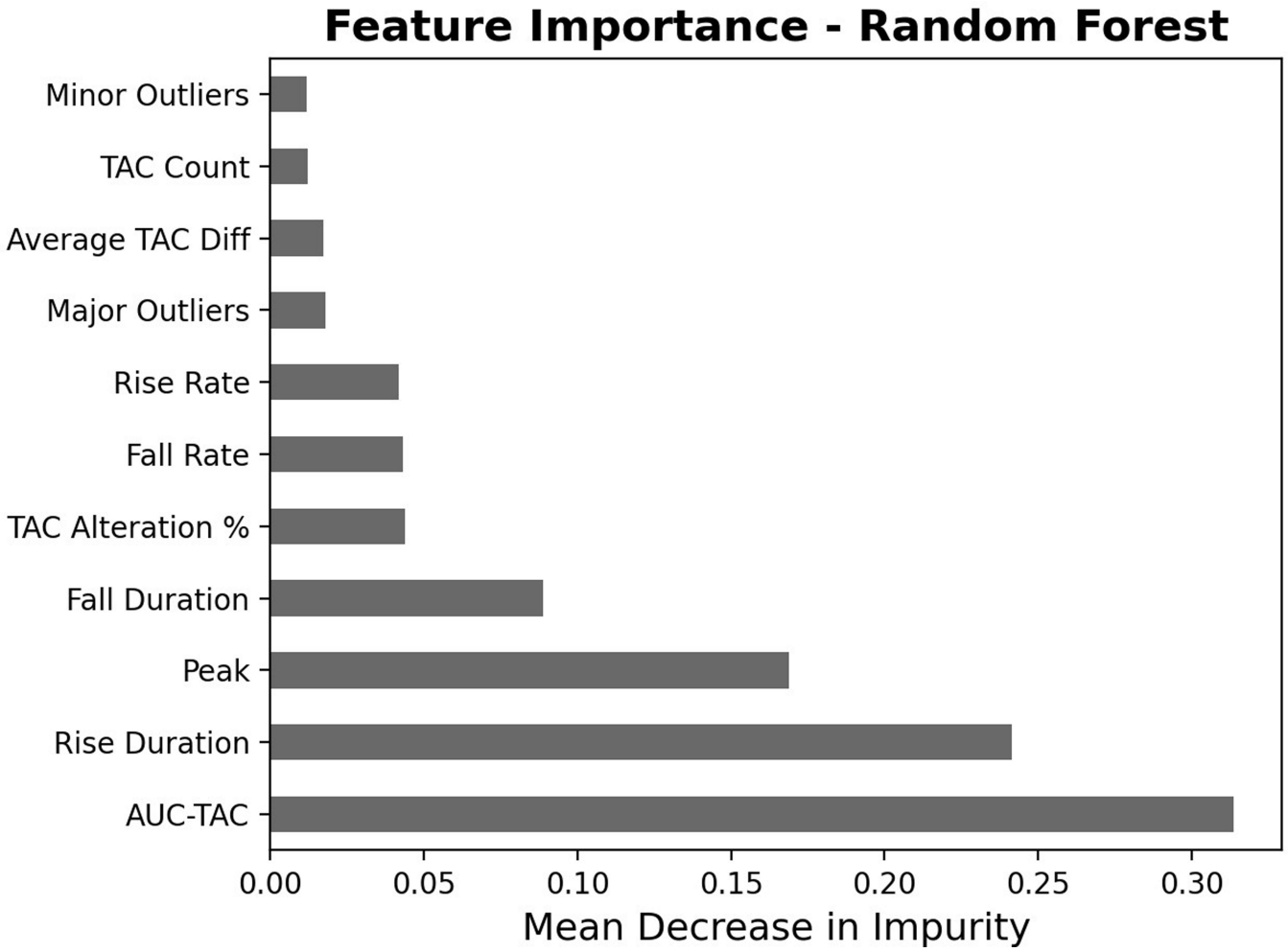



**Figure 4.**

Mean decrease in impurity is a measure of relative contribution toward predictive accuracy in the random forest, scored between 0 and 1. The four most important features were AUC-TAC (0.31), rise duration (0.24), peak (0.17), and fall duration (0.09). AUC = area under curve; TAC =transdermal alcohol concentration.

**Table 1.**

Sample Characteristics and Drinking Episodes

| | Mean (SD) or N (%) | Range |
|---|---|---|
| **Demographics** | | |
| Sex, female | 16 (50.0%) | -- |
| Age (years) | 27.3 (4.4) | 21 – 35 |
| Education (years) | 16.7 (2.2) | 10 – 21 |
| Race, white | 27 (87.1%) | -- |
| **Alcohol Drinking Characteristics** | | |
| Drinks per drinking day (in past month) | 4.5 (2.7) | 1.1 – 13.9 |
| Total drinks per week (in past month) | 15.5 (9.9) | 2.0 – 41.8 |
| Proportion[a] of heavy drinking days[b] (in past month) | 0.22 (0.19) | 0.00 – 0.71 |
| AUDIT[c] | 11.1 (5.7) | 3 – 23 |
| Participants meeting criteria for AUD[d] | 13 (40.6%) | -- |
| **Episode Outcomes** | | |
| Valid Skyn datasets | 60 | -- |
| Non-alcohol drinking episodes | 30 (50.0%) | -- |
| Alcohol drinking episodes | 30 (50.0%) | -- |
| Number of alcohol drinks consumed | 5.0 (3.6) | 1 – 20 |
| Heavy drinking episodes[b] | 16 (53.3%) | -- |
| Dataset duration (hours) | 12.8 (4.1) | 2.1 – 18.0 |
| **Skyn Acceptability** [e] | | |
| “Overall, the Skyn was easy to use” | 4.5 (0.8) | 2 – 5 |
| “I would recommend the Skyn to others” | 4.1 (1.2) | 1 – 5 |
| “Overall, I was satisfied with using the Skyn” | 4.2 (1.0) | 1 – 5 |
| “Using the Skyn was intrusive” | 2.0 (1.2) | 1 – 5 |
| “Using the Skyn took too long” | 1.7 (1.0) | 1 – 5 |

*Note*. Sample statistics are based on the 32 participants who provided at least one valid Skyn dataset.

[a] Proportion is calculated as heavy drinking days / 28 past-month days

[b] Heavy drinking defined as 5 or more alcohol drinks for men (4 or more for women);

[c] AUDIT = Alcohol Use Disorders Identification Test (Scoring Range: 0–40);

[d] AUD = alcohol use disorder; 2 or more symptoms based on Structured Clinical Interview for the DSM-5;

[e] Rated between “strongly disagree” (1) and “strongly agree” (5).

**Table 2:**

Machine Learning Performance

| | Random Forest | | Logistic Regression | |
|---|---|---|---|---|
| | **Corrected** | **Raw** | **Corrected** | **Raw** |
| **True Positive** | 28 | 26 | 28 | 24 |
| **True Negative** | 30 | 26 | 30 | 28 |
| **False Positive** | 0 | 4 | 0 | 2 |
| **False Negative** | 2 | 4 | 2 | 6 |
| **Sensitivity** [a] | 93.3% | 86.7% | 93.3% | 80.0% |
| **Specificity** [b] | 100% | 86.7% | 100% | 93.3% |
| **Accuracy** [c] | 96.7% | 86.7% | 96.7% | 86.7% |
| **AUC-ROC** | 0.984 | 0.898 | 0.962 | 0.907 |

*Note*. Group k-fold cross validation results by model type (random forest and logistic regression) and whether models were trained with features from Corrected or Raw datasets. True Positive = Correctly predicted alcohol episode; True Negative = Correctly predicted non-alcohol episode; False Positive = Incorrectly predicted alcohol episode; False Negative = Incorrectly predicted non-alcohol episode

[a] True Positives / (True Positives + False Negatives)

[b] True Negatives / (True Negatives + False Positives)

[c] (True Positives + True Negatives) / (True Positives + True Negatives + False Positives + False Negatives).