Objective Assessment in Clinical Psychological Science: Progress in Wearable Alcohol Biosensors

Catharine E. Fairbairn, Ph.D.*, Dahyeon Kang, Ph.D., Jiaxu Han, M.A., and Nigel Bosch, Ph.D.

Catharine E. Fairbairn, Ph.D., Department of Psychology, University of Illinois Urbana-Champaign, cfairbai@illinois.edu; Dahyeon Kang, Ph.D., Department of Psychiatry and Behavioral Sciences, University of Washington, danikang@uw.edu; Jiaxu Han, M.A., Department of Psychology, University of Illinois Urbana-Champaign, jiaxuh2@illinois.edu; Nigel Bosch, Ph.D., School of Information Sciences and Department of Educational Psychology, University of Illinois Urbana-Champaign; Champaign, IL 61820, pnb@illinois.edu

This research was supported by National Institutes of Health Grants R01AA025969 to Catharine E. Fairbairn and R01AA028488 to Catharine E. Fairbairn and Nigel Bosch. Our thanks to Alexa Boland, Scott Jung, Camille Lansang, Jay Park, Yujung Son, and the students and staff of the Alcohol Research Laboratory for help in the conduct of this research.

* Correspondence concerning this article should be addressed to Catharine Fairbairn, Ph.D., Department of Psychology, 603 East Daniel St., Champaign, IL 61820. Electronic mail: cfairbai@illinois.edu.

Abstract

Clinical psychology is a discipline reliant on self-reports, but uniquely susceptible to specific biases associated therewith. Here we review progress in objective behavioral assessment in the domain of alcohol research, introducing an emerging transdermal class of wearable alcohol biosensor. We note challenges of transdermal assessment, together with recent performance gains from updated devices and analytic tools, including machine learning. We indicate unanswered questions for transdermal technology, including whether devices might ultimately produce fine-grained drinking quantity estimates and device longevity. We further identify factors that can impede development of new objective measures, including the tendency to judge new tools against an implicit ideal and consider scientific findings divorced from methodological details. Finally, in evaluating transdermal and other objective measurement tools, we argue for consideration of the specific error type (random vs systematic) generally linked with novel vs existing tools, identifying measurement diversification as a priority for clinical psychology moving forward.

Keywords: Measurement, Behavior, Alcohol, Wearables, Machine Learning, Common Methods Bias

Measurement of behaviors is central to the understanding of the etiology, treatment, and diagnosis of psychopathology (Haynes et al., 2019; Kendler, 2005; Sanislow et al., 2010). Clinical psychologists have historically relied heavily on self-reports for behavioral assessment, employing questionnaire and interview-based measures for assessing facets of behavior ranging from sleep to eating to spending to alcohol and drug use (American Psychiatric Association, 2013; Hunsley and Mash, 2007; Sher and Trull, 1996; Watkins et al., 1995). Yet participants' ability to accurately report on their behaviors is impacted by a range of factors, including participants' memory/attention (Schwarz, 2007), behavioral automaticity (Nisbett and Wilson, 1977), as well as concern surrounding stigma and self-presentational constraints (Tourangeau and Yan, 2007). While various influences drive researchers' reliance on subjective self-reports, methodological barriers are likely to count among these, as the range of tools available for objective behavioral assessment has historically been limited (Fairbairn and Bosch, in press).

In more recent decades, technological advances, including in wireless communication, miniaturization, and artificial intelligence, have expanded the range of measurement options available for behavioral assessment in clinical psychological science (Fairbairn and Bosch, in press; LeCun et al., 2015; Pew Research Center, 2018). In the current article, we take the case example of alcohol use—an assessment domain notable for the challenges that can accompany subjective self-reports (Del Boca and Darkes, 2003). Specifically, we review progress in a class of objective wearable alcohol biosensor known as transdermal monitors. We further offer a brief overview and introduction to machine learning analytic methods as they relate to transdermal technology, exploring how machine learning has been applied to the dense time series yielded by transdermal monitors to improve device performance. Finally, we review challenges and common misconceptions linked with transdermal sensor validation research, examining also how

these might interface with challenges for assessment in clinical psychological science more broadly.

Self-Reports in Clinical Psychological Science

Behavior is key to the understanding of a range of psychopathologies, with many if not most clinical disorders being defined by idiosyncratic, maladaptive, excessive, or deficient patterns of behavior (American Psychiatric Association, 2013; Amianto et al., 2015; Krystal and State, 2014; Maisto and Saitz, 2003; Robins, 1981; Zwaigenbaum and Penner, 2018). While a subset of disorders are characterized by a primarily behavior-based profile of symptoms (e.g., eating disorders; antisocial personality disorder; autism spectrum disorder), other disorders defined primarily by non-behavioral manifestations nonetheless feature key symptoms in the behavioral domain (e.g., compulsions in OCD, sleep disturbance for depression, avoidance for anxiety). Beyond symptoms, behavior is also key to the broader understanding of individual-level functioning and, as such, even when investigating (primarily) non-behavioral disorders, researchers are often interested in precursors or consequences of disorders in the behavioral domain, including school performance, job functioning, relationship behaviors, and health behaviors (Dooley et al., 2023; Keller et al., 2009; Romano et al., 2001; Tucker et al., 2006; Wang et al., 2014).

Behavior also represents a domain of functioning well suited to objective assessment (Baumeister et al., 2007). In contrast to some constructs of interest to clinical psychologists, including those considered intrinsically internal and experiential in nature (e.g., wellbeing; Garcia and Gustavson, 1997), behavior exists as an external and observable aspect of human experience (Vazire, 2010). Historically, however, researchers were faced with limited options for the assessment of behavior beyond self-reports, with available alternatives often requiring labor-

intensive human-based coding systems (e.g., Bakeman and Gottman, 1997; Ekman et al., 2002) and/or bulky equipment that constrained assessment to a lab environment (Ancoli-Israel et al., 2003; e.g., sleep actigraphy, direct food measurement; Stubbs et al., 2014). As such, studies employing (direct) behavioral measures of behavior remain comparatively rare. In a recent systematic review of papers published in three of the highest ranking journals in clinical psychological science, we identified self-reports as the single measure type most commonly employed for assessing behavior, with 41% of measures employing questionnaire or interview-based tools—a figure that outstripped direct behavioral measures by a sizable margin (31%, See Figure 1; Fairbairn and Bosch, in press). Of note, reliance on self-reports in clinical psychology extends beyond the assessment of behavior. Approximately 90% of non-experimental articles in our review tested primary study aims using at least one closed-ended questionnaire or interview-based measure, and over half (54%) employed designs wherein both predictor and outcome were measured via self-report (see Figure 1; Fairbairn and Bosch, in press).

While clinical psychology is a discipline that relies heavily on self-reports, it is one also arguably particularly susceptible to the challenges associated with subjective assessment. Researchers across subfields of psychology have long raised concerns regarding biases linked with self-reports, noting limitations in human capacity to provide accurate information surrounding their experiences (Allport, 1927; Nisbett and Wilson, 1977; Schwarz, 1999). Clinical psychological research integrates elements likely to exacerbate these sources of bias, including elements relevant to context, population, and the assessments themselves. Regarding clinical *assessments*, these often touch on topics that are sensitive in nature (e.g., illegal activities, thoughts of self-harm, sexual experiences)—domains notorious for eliciting high levels of misreporting, including in therapeutic and otherwise anonymized/"safe" reporting

settings (Tourangeau and Yan, 2007). Further, many of the core cognitive and affective processes underlying psychopathology are theorized to be primarily automatic in nature and so not consistently accessible to participants for the purposes of reporting (e.g., negative automatic thoughts; Beck, 1976; Nisbett and Wilson, 1977). Clinical *contexts* can involve the perception for secondary consequences linked with specific outcomes (e.g., legal ramifications), so potentially incentivizing misreporting (Berry et al., 1991; Rees et al., 1998). Finally, regarding clinical *populations*, numerous clinical disorders are defined in whole or in part by deficits in skills that would directly impact the validity of self-reports, including impaired verbal expression, impaired attention, lack of insight, and a propensity towards deceit (Ditzer et al., 2023; Lilienfeld and Fowler, 2006; Williams et al., 2008).

Depending on the specific nature of the research being conducted and the profile of measures contained therein, these sources of error can present not only as "noise" (diminishing researchers' ability to detect true effects of interest) but also as “confounds” (creating the perception of an association between variables where none exists). Importantly, where both predictor and outcome share a method of measurement, significant findings yielded from statistical tests might not reflect shared variance linked with true underlying constructs of interest, but rather simply shared variance linked with the method of measurement (“common methods bias”, see Sidebar; Campbell and Fiske, 1959; Podsakoff et al., 2003). As such, a significant positive association between self-reported marital stress and self-reported negative alcohol-related consequences could reflect a range of factors (see Figure 1), from a true underlying association to participant-level differences in memory (are they able to recall relevant instances of each?), participants’ negative affect at the time of assessment (do negative events—past, present, and future--seem broadly salient?), individual differences in acquiescent response

style, or even participants' own lay theories surrounding this association (do they themselves speculate their spouse drives them to drink?; Podsakoff et al., 2024, 2003). In the context of a field wherein over half of studies assess associations using designs wherein both predictor and outcome are measured via self-report (Fairbairn and Bosch, in press), and concern surrounding research practices that might increase false-positive effects is burgeoning (Tackett et al., 2019), the potential for non-replicability linked with self-reports should raise widespread concern. Given its capacity for capture via non-subjective means, the behavioral domain presents researchers with important opportunities for measurement diversification in clinical psychology.

Within the domain of behavioral measurement, few constructs have presented with as widely-documented a profile of challenges as has alcohol use (Del Boca and Darkes, 2003). Of note, alcohol represents a unique case example wherein engagement with the target behavior itself can interfere with an individual's ability to accurately report on the behavior being enacted. The consumption of alcohol is known to impair cognitive capacity required for self-reports, including memory, attention, and self-awareness (Fairbairn et al., 2021; Weissenborn and Duka, 2003; White, 2003). Beyond cognition, motivation to actively engage with self-monitoring is likely to wane over the course of a drinking episode, as (drinking) cue-inspired craving kicks in, cognitive resources diminish, and demands of the context may increase (i.e., the party really gets started; DiClemente and Prochaska, 1998; Marlatt, 1996). Further, the size and strength of drink "pours" can vary widely in magnitude, as can the concentration of alcohol in each individual drink, meaning even cognitively alert and motivated drinkers can struggle to accurately report (Barnett et al., 2009; Kerr et al., 2008). In addition, even for individuals who have all necessary resources to self-report, external and self-presentational factors may interfere. Overconsumption of alcohol is widely stigmatized (Krendl and Perry, 2023), and societal factors may leave some

individuals reluctant to report even moderate levels of consumption (e.g., under-aged drinkers). Further, results of alcohol use assessment can be linked with negative consequences in some cases, including serious legal and professional ramifications, so potentially incentivizing misreporting (Jones, 1996). Finally, a propensity towards deceit surrounding alcohol use has historically been considered a defining symptom of disordered drinking, reflecting individuals' attempts to protect an established pattern of behaviors from internal and external pressure to change (Farber, 2020; Schuckit et al., 2020). In sum, self-reports of alcohol have the potential to fail researchers just at the times they are most critical—i.e., in assessing heavy drinkers and active episodes of consumption.

Alcohol Use Measurement

Alcohol has been documented as a leading cause of death and disability among younger adults worldwide (Griswold et al., 2018), with one in four individuals in the U.S. meeting criteria for Alcohol Use Disorder during their lifetime (Grant et al., 2015). Regarding the measurement of drinking, as with other forms of clinical assessment, alcohol use measurement has most often involved self-reports, including both interview- and questionnaire-based measures. In clinical research, a commonly employed measure involves questions of use quantity/frequency, in which participants are asked to indicate their average number of drinking episodes (frequency) and typical number of drinks per episode (quantity) during an average month (Del Boca and Darkes, 2003). To gauge more fine-grained patterns of consumption, researchers have sometimes employed calendar-based recall tools involving cued reporting of specific day-level drink counts (e.g., 90-day timeline follow-back; Sobell and Sobell 1992). In clinical settings, screening for alcohol-related problems has most frequently featured written questionnaires such as the AUDIT and CAGE, requiring that patients estimate use patterns and use consequences over the course of

the prior year (Kranzler and Soyka, 2018; Maisto and Saitz, 2003), while the formal diagnosis of alcohol use disorder typically involves interview-based assessment of alcohol-related experiences and consequences over the same time span (American Psychiatric Association, 2013; World Health Organization, 2022). Finally, more recently, ambulatory methods have been developed that permit the recording of individual drinks/drinking episodes in real time (Piasecki, 2019; Shiffman, 2009), so potentially addressing issues linked with retrospective recall and behavioral aggregation associated with traditional self-report measures.

Alcohol ingestion can also be measured objectively in relatively precise terms over the shorter term via direct analysis of alcohol levels in bodily fluids. The concentration of alcohol in blood (blood alcohol concentration; *BAC*) is considered the gold standard for alcohol assessment (Swift, 2003; Widmark, 1932). Breath alcohol levels are a widely-employed, non-invasive proxy for assessing *BAC* (Jones, 1996), reflecting the concentration of alcohol present in capillaries of the lungs and producing values that correlate strongly with true blood alcohol levels (Gibb et al., 1984). Beyond these direct measures of alcohol levels, which offer point-in-time estimates of intoxication, indicators of longer-term patterns of consumption can be found in the analysis of blood, urine, and hair, including through assessment of direct alcohol byproducts (e.g., PEth; FAEE; Harris et al., 2021) and indirect indicators of alcohol's impact on the liver and body (e.g., CDT; EtG; Crunelle et al., 2014; Fakhari and Waszkiewicz, 2023; Harris et al., 2021).

Extant objective indexes of drinking have advantages, and, considered alone or in combination, may serve as excellent measurement choices for specific applications. At the same time, each of these measures is linked with limitations. Although we have a highly accurate test of immediate intoxication level in the form of blood-based analysis, such blood tests are costly and invasive, and further assess intoxication level only at a single point in time (Swift, 2003).

Indirect biological markers aimed at recording patterns of consumption over the longer term also provide useful information, but lack specificity for describing day-level drinking and further have the potential for differential accuracy dependent on demographics and overall levels of health (Harris et al., 2021). Breathalyzers represent the tool with the potential to provide the most precise, temporally-specific record of intoxication over the longer term, with the advent of affordable, smartphone-integrated breathalyzers making this a measurement option accessible to many (Delgado et al., 2017). However, breathalyzer readings can be biased by mouth alcohol, require repeated/motivated action on the part of the user, and further involve a conspicuous action sequence for reading acquisition (e.g., 5-7 second exhale into a tube), so limiting the utility of breathalyzers during times of active consumption (Caddy et al., 1978; Jones, 1996). In sum, the field has lacked a noninvasive tool for passive, continuous alcohol assessment over the longer term. Wearables represent a measurement option well positioned to fill this gap (Barnett, 2015; NIAAA, 2015; Swift, 1993).

Wearable biosensors have the potential to address measurement challenges by filling needs in three key areas in the domain of alcohol use assessment: objectivity, longevity, and continuity (Barnett, 2015; Fairbairn and Kang, 2020; Luczak and Ramchandani, 2019; Swift, 1993). Specifically, wearables offer *objective* assessment in a measurement domain where both internal and externally driven incentives for misreporting can be high. Further, in a domain where patterns of behavior over the longer term can have critical implications for health and functioning, yet active reporting over *extended time* spans is often impractical, wearables have the capacity to passively trace alcohol use over the course of months and even years. Finally, wearables offer the potential for *continuous* monitoring in an domain where motivational and cognitive resources for self-assessment vary dependent on engagement with the very behavior

being assessed (i.e., alcohol ingestion). Given this profile of advantages, it is perhaps unsurprising that proposed applications for wearable sensors have been numerous, including across prevention, intervention, medical, motor-vehicle safety, and research domains (see Table 1). While some applications would require a sensor with a high level of temporal specificity (just-in-time relapse prevention) and time-locked accuracy (motor vehicle safety alerts), other applications would have widespread utility even given sensors capable of offering a lower level of precision (e.g., alcohol “fitbit” providing a long-term record of high-risk, low-risk and abstinent drinking days; see Table 1; Fairbairn and Bosch, 2021).

Transdermal Alcohol Monitors

A variety of wearable devices have been proposed for the passive and continuous measurement of alcohol consumption. These have included microneedle arrays (Vinu Mohan et al., 2017), gait-based sensors (e.g., alcohol sensing shoes; Lee et al., 2017), tattoos with sweat-inducing chemicals (Kim et al., 2016), as well as sensors based in infrared laser technology (Ridder et al., 2009). Several of these methods are currently under development and may ultimately yield viable tools, potentially changing the landscape of alcohol measurement in years to come (e.g., infrared sensors for assessing alcohol content within interstitial fluids; Ridder et al., 2009). To date, however, the most well validated wearable alcohol sensor is of a type commonly referred to as "transdermal" (Brobbin et al., 2022; van Egmond et al., 2020; Yu et al., 2022). Formulated on the basis of observations of alcohol pharmacokinetics made nearly a century ago (Nyman and Palmlöv, 1936), transdermal sensors measure drinking via a device that rests on the skin’s surface. Transdermal sensors are now in widescale use by individuals around the world, including by thousands of participants in research protocols and clinical trials (Davis-

Martin et al., 2021; Roberts and McKee, 2019) and over 1 million clients in the criminal justice system (Alcohol Monitoring Services, 2023).

Transdermal sensors rely on a physiological property of alcohol that many of us have borne witness to with our senses, albeit sometimes unknowingly. Alcohol fumes can at times seem to “waft” from the bodies of intoxicated individuals, encompassing not only the breath but also the skin (e.g., "alcohol seeped from our pores”; R. L. Jones 2021). Such expressions are not only a literary device but in fact reflect a scientific reality. Transdermal devices draw on alcohol’s property as a substance that distributes itself across all bodily fluids, including blood, urine, interstitial fluid, and sweat (Nyman and Palmlöv, 1936; Swift, 1993). Specifically, following ingestion, alcohol is absorbed through the lining of the stomach and small intestine into the bloodstream. From there, it circulates throughout the body, seeping through blood vessel walls to reach sweat glands and ultimately the cells of the skin itself (Swift, 2003). Approximately 1% of ingested alcohol is excreted through the skin (Nyman and Palmlöv, 1936; Swift, 2003). As such, alcohol can be detected via sensors placed on any portion of the human body, from palms to ankles to feet and foreheads.

Transdermal sensors measure the alcohol present in vapor form in the pocket of air immediately above the skin’s surface (Anderson and Hlastala, 2006; Giles et al., 1987). This air pocket contains alcohol excreted through the skin via two distinct processes: active excretion via sweat glands (sweat) as well as passive diffusion from the epithelial cells of the skin (insensible perspiration) (Anderson and Hlastala, 2006; Swift, 1993). Leveraging the same low-cost fuel cell technology integrated within drugstore breathalyzers sold around the world, transdermal devices measure alcohol via an electrochemical process involving the transformation of alcohol into an electrical current, the magnitude of which is directly measurable at the sensor (Swift, 2000). A

meta-analysis of laboratory-based research with these sensors indicates strong correlations between transdermal and blood/breath alcohol concentration in controlled dosing settings, (average $r = 0.87$; Yu et al., 2022).

Regarding engagement with transdermal technology, these sensors have been received with variable levels of enthusiasm within the alcohol research community. With the inception of the first viable sensors, enthusiasm for transdermal alcohol sensors among scientists ran high (see Fairbairn and Bosch, 2021; Giles et al., 1987; Swift, 1993). As results of tests with early prototype sensors began to accumulate, however, initial enthusiasm gave way to skepticism, with results suggesting transdermal sensors were not the “breathalyzer for the skin” that some had initially hoped they might be (Anderson and Hlastala, 2006; Marques and McKnight, 2009; Webster and Gabler, 2007). Specifically, a range of challenges were identified for transdermal sensors, including in the domain of *accuracy* (congruence between transdermal estimates of drinking and "true" drinking) as well as that of *timing* (temporal displacement between transdermal estimates of drinking and "true" drinking) (Luczak and Ramchandani, 2019).

Regarding the accuracy of transdermal sensor output, concerns were raised about both person-level and contextual factors that might confound the relationship between *BAC* and transdermal alcohol concentration (*TAC*) (Anderson and Hlastala, 2006; Luczak and Rosen, 2014). While creative calibration protocols were developed to address person-level variability (Luczak and Rosen, 2014), context-level confounds (inherently transient and sometimes difficult to measure) were considered as presenting a particular challenge (Anderson and Hlastala, 2006). Transdermal sensors assess alcohol levels present in both sweat and also insensible perspiration. While alcohol levels in insensible perspiration are attenuated relative to blood, alcohol levels in sweat are not (Swift, 2000). As such, any context-level factors with the potential to influence the

relative proportion of sweat and insensible perspiration at the skin's surface (e.g., aerobic exercise, ambient temperature, humidity, skin hydration) might also lead to variability in the relationship between *TAC* and *BAC* (Anderson and Hlastala, 2006). Further, beyond the sweat/insensible perspiration ratio, additional contextually dependent factors with the potential to confound *TAC*-*BAC* links include airborne alcohol external to the sensor (e.g., perfume; hand sanitizer) and bodily movements leading to variability in the distance between the sensor and skin (Anderson and Hlastala, 2006; Fairbairn et al., 2025).

Regarding timing, as noted above, alcohol arrives at the skin's surface by way of the blood and, as such, alcohol is detectable in the bloodstream prior to the time at which it is detectable at the skin's surface (Swift, 2003). The average duration of this lag time is not precisely quantified and is likely to vary (Anderson and Hlastala, 2006). However, some early studies indicated that the average duration of this lag might stretch as long as 4-5 hours (Marques and McKnight, 2009). Such findings augured an era of pessimism surrounding transdermal sensors, including the widespread belief that sensor applications requiring anything approaching time-sensitive assessment (e.g., real-time relapse prevention; research applications involving a time-locked drinking record) might ultimately emerge as infeasible (Anderson and Hlastala, 2006). Taken together, early research with transdermal technology led to concerns surrounding its viability, with some going so far as to speculate that the relationship between *BAC* and *TAC* is too variable to permit accurate translation of data from transdermal sensors, and further that lag times are too lengthy to permit temporally specific transdermal estimation of *BAC* (Anderson and Hlastala, 2006; Marques and McKnight, 2009; Webster and Gabler, 2007).

Challenges of transdermal alcohol assessment, including those linked with temporal lags and contextual confounds, are indeed real. However, the conclusion that such challenges

preclude the possibility for useful transdermal sensor application was ill founded. Both device and analytic tools available to scientists within early stages of transdermal sensor testing were poorly suited to more fine-grained alcohol use estimation tasks. Specifically, most early transdermal validation research examined relatively bulky ankle bracelets featuring a sparse 30-minute sampling interval (e.g., AMS SCRAM device; see Fairbairn and Bosch, 2021; Yu et al., 2022). The ankle positioning and large size of these sensors diminished user acceptability and exaggerated skin-sensor variability, while the sparse sampling interval constrained capabilities for time-sensitive drinking detection (Fairbairn et al., 2020). Further, this 30-minute sampling interval can diminish sensitivity for detecting more nuanced over-time patterns in *TAC* readings—nuances that can aid with the critical tasks of addressing transdermal sensor lag times and also in parsing alcohol consumption from contextual confounds (Fairbairn and Bosch, 2021). For example, a rapid increase in *TAC* levels that could only be observed by frequent sampling might indicate a qualitatively different event from a gradual rise—e.g., increase in drinking vs increase in sweating (Figure 2). Beyond the devices themselves, early analytic methods for converting *TAC* into *BAC* relied on relatively basic conversion algorithms (e.g., linear regression; see Fairbairn and Bosch, 2021; Yu et al., 2022). Such analytic methods are typically ill suited to modeling complex, temporally displaced associations that can vary dependent on a range of contextual and individual factors. In sum, given the nature of these device and analytic tools, lackluster performance of early transdermal sensors is as likely to reflect a failure of method for extant tools vs a failure of concept for transdermal assessment more broadly.

Progress for Transdermal Devices: Rapid Sampling Sensors

The past several decades have given rise to technological developments with key implications for behavioral assessment in not only alcohol science but more broadly within

clinical psychological research. Advances in miniaturization, wireless communication, and computational methods represent a game changer for the direct measurement of behavior in everyday life (Fairbairn and Bosch, in press; Harari et al., 2016; Mehl et al., 2024). This time period has not only borne witness to the development of new technological tools but also, in many cases, their rapid dissemination and uptake. Smartphone owners now represent the overwhelming majority of individuals in developed countries (85%; Pew Research Center, 2022). As such, we exist in a strange new reality in which billions of individuals around the world now move through everyday life with a powerful supercomputer at their fingertips.

Such developments have had key implications for transdermal alcohol sensing technology. Spurred on by widely publicized funding initiatives (e.g., NIH's Wearable Alcohol Biosensor Challenge; NIAAA, 2015), in the years since 2015, a new generation of transdermal sensor has emerged (Fairbairn and Kang, 2019; van Egmond et al., 2020; Wang et al., 2019). These new sensors are characterized by a compact, wrist-worn design, similar to a Fitbit (Figure 3). New generation sensors measure *TAC* on the basis of passive airflow (vs the earlier, bulkier active air pump designs) and further feature wireless smartphone integration and cloud-based data storage (Wang et al., 2019). These novel design and data-storage elements result in new-generation devices capable of sampling *TAC* at over 90 times the rate of older generation sensors (15-20 second sampling interval; Fairbairn and Bosch, 2021). New generation devices also integrate rapid sampling of additional data streams beyond *TAC*, including motion, temperature, and humidity. Devices currently available for research include the BACtrack Skyn, Smart Start ORBIS (previously BARE), Arborsense GRADE, and the SOBRSafe SOBRSure device. Acceptability for these sensors among users is high relative to older-generation monitors, with

users rating devices as relatively attractive for regular wear and reporting minimal discomfort (Courtney et al., 2023; Murgia et al., in press; Rosenberg et al., 2023; Wang et al., 2021).

Of particular note, lag times between *BAC* and transdermal estimates of *BAC* (*eBAC*) are significantly shorter for new- vs old- generation sensors (Fairbairn and Kang, 2019; see Yu et al., 2022). Sequences of readings collected over time contain information not only on where values currently "are" but also where they are "going," a circumstance that has led to the development of a sub-field of time-series analysis centered on future forecasting (Figure 2; Box et al., 1994). As such, research indicates new generation sensors are generally capable of detecting alcohol within 30-60 minutes of use initiation (Brobbin et al., 2022; Fairbairn et al., 2025; Fairbairn and Kang, 2019). Regarding comparisons across device generations, a laboratory study in which participants (N=30) consumed a fixed alcohol dose (0.08% target *BAC*) or a control beverage while wearing both new-generation wrist and old-generation ankle monitors indicated lag times for new-generation sensors were less than half the duration of lag times for old-generation sensors (54 vs 120 minutes; Fairbairn and Kang, 2019). Regarding findings within the broader literature, a meta-analysis of 30 studies of transdermal sensors found that lag times for ankle-worn sensors were approximately double the magnitude of lags for sensors worn on the arm, wrist, or hand, and further that *TAC*-*BAC* correlations tended to be lower for ankle-worn devices (Yu et al., 2022; see Figure 3).

Beyond lag times, the dense time series produced by new-generation sensors can have utility for parsing contextual confounds and thus for the broader accuracy of transdermal sensors (Fairbairn and Bosch, 2021). Thirty minutes of transdermal alcohol readings from a sensor that samples every 20 seconds can carry information beyond what any one of these readings might convey when considered in isolation (Fairbairn et al., 2025). Regarding the utility of these

metrics for addressing issues in transdermal assessment, divergence along features (i.e., predictors) such as trends, absolute successive difference values, total variability, and minimum/maximum might be used to differentiate internal/ingested (e.g., a beer) from external/airborne (e.g., hand sanitizer) alcohol (Figure 2). In line with these suppositions, data from the same multi-sensor laboratory alcohol-administration study described above, albeit this time with an expanded participant pool (N=110), provided evidence that dense time series yielded by new generation devices have the potential to increase device accuracy (Fairbairn et al., 2020). Results indicated that models for estimating *BAC* from transdermal data were 44% more accurate when built on rapid sampling Skyn (new-generation) vs slower sampling SCRAM (old-generation) data. Further, and important, the relevance of time-series features was reinforced even in comparisons that held constant specific device employed—models predicting *BAC* on the basis of time series features from 30 minutes of Skyn *TAC* data were 38% more accurate than models built only from the single *TAC* reading most proximal to *BAC* assessment. Taken together, results appear to suggest what common sense would also indicate—device design matters.

Progress in Analytic Methods: Machine Learning Algorithms

In addition to hardware and device development, the past decade has seen marked development in approaches for the analysis and interpretation of data derived from wearable devices (Didier et al., 2023; Fairbairn et al., 2020; Fairbairn and Bosch, 2021). Specifically, the years since 2010 marks an era sometimes referred to as the “AI Spring,” reflecting rapid progress in a family of analytic methods known as machine learning (LeCun et al., 2015; Norvig and Russell, 2021). Machine learning methods differ from conventional statistical approaches in that,

rather than imposing a pre-determined form for relationships between variables, these models instead directly "learn" the shape of the relationships from the data themselves.

To provide a specific example, a researcher seeking to understand the relationship between *TAC* and *BAC* using standard statistical tools might construct a linear regression model in which *TAC* is entered as a predictor of *BAC*. To allow for more complexity, the researcher might consider adding quadratic, cubic, or even segmented (e.g., spline) shapes for the association. All of these would be first theorized by the researcher and subsequently imposed on the data, with goodness of fit being gauged via tests of significance. A machine learning approach effectively takes this process and flips it on its head. Rather than the researcher deciding on probable shapes and forms for the relationship between *TAC* and *BAC*, machine learning models undergo an explicit "training" stage, during which a portion of a dataset is examined for patterns in relationships among predictors and outcomes. Following training, the model would theoretically be able to account for a relationship between variables of virtually any shape, capable of conforming to any pattern observed in collected data (Figure 4). To offer an analogy, if models were to adhere to data in the manner in which clothing does the human frame, linear regression models would hang like a muumuu, whereas machine learning models can cling like a leotard (Fairbairn and Bosch, 2021).

The tendency for machine learning to adhere closely to training dataset characteristics has important implications for understanding not only its strengths but also its limitations. The relative flexibility of machine learning models means these algorithms shine in cases where the relationship between variables is complex and the number of parameters available for understanding this relationship (i.e., number of "features" or predictors) is large. However, and importantly, machine learning models can "learn" not only generalizable relationships (i.e., the

"good stuff") but also measurement error and idiosyncratic dataset characteristics. Like the mind of a toddler, machine learning models are absorbing it all—not only the careful moral lessons we inject into bedtime stories but also (inconveniently) the occasional profanity we hurl at drivers in traffic. To offer an example relevant to the translation of transdermal sensor data, if a machine learning model were to be trained on a dataset in which participant self-reports were used as a proxy for *BAC*, and participants tended to lose track of drinks when engaging in high intensity binge episodes, the resulting machine learning algorithm would similarly fail to differentiate *BAC*s at more extreme levels. Similarly, if trained using a dataset featuring primarily women examined in laboratory contexts, the resulting machine learning algorithm might fail to generalize to individuals of other gender identities and data collected in the field. In contrast, a regression model, whose shape is determined by the researcher, would be more robust to idiosyncratic dataset characteristics. Since large datasets increase the likelihood of sample representativeness and generalizability, and can also reduce effects of measurement error and other noise by permitting exposure to a large number of instances, machine learning models are known for performing best when given access to large datasets for model training (Adjerid and Kelley, 2018; Bahri et al., 2024). Finally, the complexity of machine learning output has implications for transparency, as for many of the more complex model types it is not possible to clarify the processes by which the model reached its prediction. As such, machine learning may be less well suited to applications wherein such transparency is key (e.g., forensic settings; Vollmer, 2020).

Important, machine learning models have utility for the analysis of sequences of readings taken over time, including for the analysis of dense time-series data of the type produced by new-generation sensors (Figure 5). In particular, newer machine learning model types, including

convolutional neural networks and transformers, have permitted a form of "levelling up" for time series analysis, permitting the automatic recognition of patterns in time series data of nearly any shape (see Figure 5; Lee et al., 2009; Xu et al., 2018). Other recent computational developments include those permitting the fast extraction of multiple time series features (Christ et al., 2018) as well as methods for avoiding model "overfitting" (i.e., "regularization"; Chen and Guestrin 2016). Taken together, these computational developments have yielded advances in future prediction problems across fields, including crop yield forecasting (Chlingaryan et al., 2018), climate analysis (Jones, 2017), and speech recognition (Ma et al., 2025). Regarding transdermal application, in fixed-dose laboratory research featuring direct device/analytic comparisons (Fairbairn et al., 2020), we found the accuracy of *BAC* estimates based on (time-series based) machine learning models was more than twice that of linear regression, suggesting machine learning has the potential to substantially improve transdermal sensor performance.

As noted previously, much of the challenge for transdermal assessment comes via contextual variability. Therefore, studies featuring a single fixed alcohol dose and a climate-controlled testing context offer an artificially "easy" testing scenario that is likely to substantially inflate the performance of transdermal alcohol sensors vs real-world (field) environments. To address this issue, our ongoing research introduces variability into laboratory dosing conditions, while also integrating large-scale objective ambulatory assessment of drinking outside the lab (NCT05692830; Figure 6). In this research, participants wear new-generation transdermal sensors while attending three laboratory visits involving variable alcohol doses (0.03%, 0.06%, and 0.09% target *BAC*) while being exposed to precisely timed environmental manipulations including aerobic exercise tasks, airborne alcohol exposure (e.g., hand sanitizer), and activities aimed at introducing variability in the distance between sensor and skin (arm/wrist movement).

Participants also supply breathalyzer readings at regular intervals throughout a 14-day period of ambulatory assessment. We recently reported initial findings from data from the first 100 participants to enroll in this ongoing trial (12,699 *BAC* readings; 5.39 million transdermal readings). For this research, machine learning models were built to operate in "real time" in that they incorporated only transdermal readings preceding, and not following, the moment of *BAC* assessment, so potentially mirroring information available for more temporally sensitive device applications (e.g., "just-in-time" relapse prevention). Initial results from this study indicated excellent accuracy for transdermal sensors in detecting episodes of drinking in real-time across a wide array of laboratory and field-based testing conditions, *AUROC*, 0.966; Sensitivity, 89.8%; Specificity, 90.6% (Fairbairn et al., 2025).

To determine the extent to which transdermal monitors have utility for predicting more fine-grained indicators of drinking (e.g., precise *BAC* levels) larger datasets will be required. Over the next several years, as big samples continue to accumulate, firmer answers to this question are likely to be forthcoming. However, prior research with pilot field trials might nonetheless offer relevant clues. A preliminary trial (*N*=27) featuring real-time machine learning predictions from participants observed in field contexts found the average absolute distance between true *BAC* and transdermally estimated *BAC* ranged between 0.019-0.041% (Ariss et al., 2023). Further, results of power-law-curve projections included in this research indicated that the accuracy of transdermal *BAC* estimates is likely to improve over 25% as larger datasets become available for model training (Ariss et al., 2023). Taken together, findings to date indicate that transdermal sensors have utility as alcohol use monitors, including for specific time-sensitive applications, and further present the intriguing possibility that they might ultimately serve as proximal indicators of *BAC*, with an average error near to that of a standard drink (0.02%).

In sum, results of studies employing up-to-date methods suggest that transdermal alcohol monitors can yield accurate estimates of drinking in a dimension close to real time. Results integrating direct methodological comparisons also suggest that new-generation sensors combined with data-driven methods yield results substantially more accurate than other techniques. Research over the next several years is likely to uncover the level of specificity with which is possible to estimate *BAC* levels from transdermal sensors, and thus whether such sensors can and should serve as *BAC* monitors in any form or rather simply a tool for the time-locked assessment of overall drinking status and broad drinking risk level. While information on the more fine-grained accuracy of transdermal sensors is still forthcoming, researchers have nonetheless devised creative applications for these sensors. Among the first applications identified was within the realm of intervention, where contingency management treatments based on transdermal sensor output were found to successfully reduce drinking among patients with alcohol use disorder (Alessi et al., 2017; Barnett et al., 2011; Dougherty et al., 2014). Transdermal sensors have also been used to better understand continuous patterns of use among individuals in various contexts, including among heavy drinking attendees of a music festival (Norman et al., 2021), patients receiving AUD intervention in an outpatient context (Alessi et al., 2019), and clientele at a local soup kitchen (Rash et al., 2019). In our own research, we applied transdermal sensors to identify factors driving drinking in everyday life, combining photographic indicators of drinking environment with continuous transdermal estimates of *BAC* to identify a range of immediate context-level predictors of heavy drinking, including unfamiliar contexts, bars, and crowds (Ariss et al., in press; Fairbairn et al., 2018). Finally, researchers have also made creative use of the dynamic data provided by transdermal sensors, with studies indicating that, over-and-above mean levels, dynamic *TAC* rise and fall rates significantly predict day-level

negative alcohol-related consequences (Russell et al., 2022) and even alcohol-induced blackouts (Richards et al., 2024).

Summary and Implications

At various points in the history of device development, challenges of transdermal sensor assessment were translated as indicating an impossibility for useful application. A critical mindset is essential when evaluating new technology, as the "hype" surrounding new gadgets can not infrequently outstrip their performance. At times, however, such skepticism can reflect a form of inertia and impede meaningful progress. We argue that misconceptions underlying early perceptions of transdermal technology are both common and also impactful not only within addiction science, but across areas where new technology is being assessed, and thus worthy of being named. These include: 1) Results of validation research can be judged independent of the specific analytic and device tools used to achieve the results. Although common across research domains, this myth can have particularly deleterious effects in fields where prototype technologies are being assessed and results are as yet sparse (the myth of the "**Methodless Findings**"); 2) An ideal method of measurement exists well-suited for all potential applications, and therefore early results from these new methods might be judged not against an imperfect (true) array of alternative measures but rather against an implicit ideal (the "**Perfect Solution**" fallacy).

Regarding the former of these misconceptions, in the field of alcohol assessment, direct methodological comparisons suggest that new device and analytics substantially improve transdermal sensor performance, with new methods indicating the potential for real-time assessment of drinking based on transdermal sensor output (Yu et al., 2022). Regarding the latter, no extant methods for measuring drinking can yield results free of bias. Breathalyzer readings

are subject to mouth alcohol effects and require motivated and repeated action on the part of the user. Blood and urine-based methods are costly and invasive. Self-reports are biased by a range of factors, including ingestion of the very substance they seek to assess. Thus, in considering a new tool, the central question of concern is not “is this device as extraordinary as its most fervent proponents have sometimes claimed it to be?” But rather “how does this (imperfect) tool compare to the range of (also imperfect) alternatives that we currently have at our disposal, and might it fill a useful gap?”

Limitations and Future Directions

Recent developments in transdermal technology have opened a range of possibilities for transdermal sensors. At the same time, these new device and analytic tools have introduced unanswered questions. Below, we discuss limitations and future directions as they pertain to both transdermal device development, as well as the research that surrounds these devices.

Regarding the devices themselves, current iterations, while considerably improved in comparison to prior devices, yet still feature substantial limitations. First and foremost, an important open question is that of device longevity (Lansdorp et al., 2019). Specifically, transdermal devices have been proposed to degrade in accuracy levels over time with continued wear. Studies have found indications of diminution in new-generation sensor accuracy both across a single episode of wear (Fairbairn et al., 2025), as well as across longer time periods and multiple episodes of wear (Ash et al., 2022). The question of whether device degradation is a limitation truly intrinsic to transdermal assessment, or whether it might be avoided with future device improvement, is yet to be addressed.

Regarding other device limitations, the most widely used sensor (Skyn) is currently compatible only with iOS smartphones, making them inaccessible to the large population of Android users. Further, current wrist-worn devices are water resistant but not waterproof, meaning that participants are required to remove devices in order to bathe or shower. Lack of waterproofing can lead to a range of negative outcomes, including device damage (if users forget to remove the device) or interruptions in monitoring (if participants fail to reapply devices following removal). Data storage for new-generation sensors also represents an area of challenge, with currently available devices requiring regular smartphone app activation (Skyn) or close proximity to the smartphone (SOBRSure) for data to be effectively stored. Related to device acceptability, some participants experience itching and skin discomfort from using transdermal wrist sensors, with a small subset reporting developing a rash (Murgia et al., in press). Finally, and important, while the underlying fuel-cell technology is not one that is inherently costly, transdermal monitoring is currently expensive (e.g., $200/month per device). Prices would be expected to decrease as development costs diminish and a larger customer base established. Nonetheless, transdermal devices are currently outside the price range of most drinkers, and so addressing this question of access will represent a priority moving ahead.

Regarding future directions relevant to scientists, non-wear detection algorithms for wrist-worn devices are currently under-developed yet important for many applications, including medical and treatment use cases. Algorithms that integrate input from multiple data streams, including temperature and motion, will likely be required for generalizable non-wear detection. Second, effects of environmental factors on transdermal sensor accuracy, including variability in sweating, skin-sensor distance, and airborne alcohol, have yet to be precisely mapped. Characterization of effects of these confounds will be useful for both gauging their impact on

broader accuracy and in building translation algorithms capable of parsing their effects from the drinking-related effects of interest. Further, while past research has featured direct comparisons between machine learning and linear regression, the field is lacking research comparing machine learning and more sophisticated top-down analytic approaches, including first-principles models that account for alcohol pharmacokinetics (Dumett et al., 2008; Luczak and Rosen, 2014). In light of inevitable noise linked with transdermal assessment, some of which will be attributable to sources that are effectively unmeasurable, it is possible that analytic models informed by such top-down processes will have increments in accuracy beyond what might be derived from data-driven models applied in isolation (Oszkinat et al., 2022). Finally, while prior research has found no significant effects of race in transdermal sensor performance, patterns of means have indicated the potential for accuracy differences for African American participants (Fairbairn et al., 2025). In future, studies that over-sample racial minority groups would be indicated to ensure generalizability of transdermal algorithms to individuals of all races.

An additional unanswered question for transdermal algorithms is their sensitivity for detecting low levels of drinking (e.g., $BAC < 0.03\%$, or episodes consisting of 1-2 standard drinks) (Barnett et al., 2014; Roache et al., 2015). In fine tuning algorithms, researchers inevitably face tradeoffs between sensitivity and specificity. For many applications, including those that extend beyond the criminal justice system, false-positive readings may need to be minimized or eliminated. Dependent on the application and required level of specificity, it is possible that algorithms designed to minimize false positive readings will ultimately not simultaneously be capable of preserving sufficient sensitivity to detect low-level drinking.

Further, and important, research is needed to more precisely identify the temporal sensitivity of new-generation transdermal sensors across diverse contextual conditions. In our

recent combined laboratory-field study, we found that approximately 70% of alcohol use episodes were detected via transdermal sensors within 30 minutes of first consumption (Fairbairn et al., 2025). However, within this initial field examination, alcohol consumption was not assessed continuously but was rather determined according to positive readings provided at points of active *BAC* assessment. As such, more research is needed on the exact time course for transdermal drinking detection before the initiation of more time-sensitive sensor applications is advisable (e.g., sensor-derived indicators of driving advisability).

Importantly, looking ahead, research has taught us that methods matter when it comes to transdermal sensor validation research. Many of the more pressing questions are now relatively nuanced, including those related to identifying precise temporal-specificity and accuracy levels. In light of this, we argue that studies featuring small/homogenous samples, sparse-sampling sensors, and self-report measures of drinking quantity will offer minimal utility for addressing most research questions in transdermal sensor validation research. Further, when considered outside the bounds of their methods, such studies can at times represent more of a hindrance than a help. Moving ahead, studies that rely on precise and objective indicators of drinking and integrate diverse participant samples are most likely to hold utility.

Conclusion

We have entered an era of extraordinary growth for measurement tools in behavioral research, with measurement options expanding exponentially across assessment, research, prevention, and intervention domains (Fairbairn and Bosch, in press; Harari et al., 2016; Mehl et al., 2024). In the current article, we took a “deep dive” into the case example of alcohol—a substance for which traditional self-report measures can be accompanied by a particularly thorny collection of challenges. However, while representing a useful illustration, alcohol consumption

is far from being the only clinically relevant behavior for which objective measurement tools have recently expanded. Data from a range of sources, from physiological sensors to smartphones to language and video analysis, have been leveraged to measure behaviors of central interest within clinical psychological science (Fairbairn and Bosch, in press). Behaviors assessed via these new tools have included exercise (Puterman et al., 2021), sleep (Boe et al., 2019), eating (Thomaz et al., 2015), social interaction (Wang et al., 2014), and affiliative behaviors (Gurrieri et al., 2021). The exact accuracy levels of many such tools across measurement contexts are yet to be determined, as is the advisability of widespread integration of these novel tools in place of traditional measures. However, in approaching such decisions, the specific form of error linked with assessment type is a factor worthy of consideration—whereas error linked with AI-based assessment may at times be sizable, *systematic* forms of error can result from over-reliance on self-reports.

In sum, clinical psychology has long been a science of self-reports, with self-reports used to assess constructs well beyond those that might be considered inherently internal and experiential. In the past, self-reports represented a choice based in part on necessity. Still now, with a growing array of measurement alternatives, the draw to favor a flawed yet familiar survey measure can be strong. However, in light of the potential for common methods bias, the excuse of "devil we do know" increasingly hangs thin. In a field where nearly 90% of studies explore primary study aims using at least one self-report measure, and false-positive findings represent a growing area of concern, measurement diversity should emerge as a priority moving ahead.

Literature Cited

Adjerid, I., Kelley, K., 2018. Big data in psychology: A framework for research advancement. American Psychologist 73, 899–917.

Alcohol Monitoring Services, 2023. SCRAM Systems Announces Milestone of One Million Clients Monitored with SCRAM CAM.

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

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

Allport, F.H., 1927. Self-evaluation: A problem in personal development. Mental Hygiene 11, 570–583.

American Psychiatric Association (Ed.), 2013. Diagnostic and statistical manual of mental disorders: DSM-5, 5th ed. ed. American Psychiatric Association, Washington, D.C.

Amianto, F., Ottone, L., Abbate Daga, G., Fassino, S., 2015. Binge-eating disorder diagnosis and treatment: a recap in front of DSM-5. BMC Psychiatry 15. https://doi.org/10.1186/s12888-015-0445-6

Ancoli-Israel, S., Cole, R., Alessi, C., Chambers, M., Moorcroft, W., Pollak, C.P., 2003. The Role of Actigraphy in the Study of Sleep and Circadian Rhythms. Sleep 26, 342–392. https://doi.org/10.1093/sleep/26.3.342

Anderson, J.C., Hlastala, M.P., 2006. The kinetics of transdermal ethanol exchange. Journal of Applied Physiology 100, 649–655.

Ariss, T., Caumiant, E.P., Fairbairn, C.E., Kang, D., Bosch, N., Morris, J.K., in press. Exploring associations between drinking contexts and alcohol consumption: An analysis of photographs. Journal of Psychopathology and Clinical Science.

Ariss, T., Fairbairn, C.E., Bosch, N., 2023. Examining new-generation transdermal alcohol biosensor performance across laboratory and field contexts. Alcoholism: Clinical & Experimental Research.

Ash, G.I., Gueorguieva, R., Barnett, N.P., Wang, W., Robledo, D.S., DeMartini, K.S., Pittman, B., Redeker, N.S., O'Malley, S.S., Fucito, L.M., 2022. Sensitivity, specificity, and tolerability of the BACTrack Skyn compared to other alcohol monitoring approaches among young adults in a field-based setting. Alcoholism: Clinical and Experimental Research 46, 783–796. https://doi.org/10.1111/acer.14804

Bahri, Y., Dyer, E., Kaplan, J., Lee, J., Sharma, U., 2024. Explaining neural scaling laws. Proceedings of the National Academy of Sciences 121, e2311878121. https://doi.org/10.1073/pnas.2311878121

Bakeman, R., Gottman, J.M., 1997. Observing interaction: An introduction to sequential analysis, 2nd ed. Cambridge University Press, New York.

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

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

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

Barnett, N.P., Wei, J., Czachowski, C., 2009. Measured alcohol content in college party mixed drinks. Psychology of addictive behaviors 23, 152–156.

Baumeister, R.F., Vohs, K.D., Funder, D.C., 2007. Psychology as the science of self-reports and finger movements: Whatever happened to actual behavior? Perspectives on Psychological Science 2, 396–403.

Beck, A.T., 1976. Cognitive therapy and emotional disorders. Meridian, New York.

Berry, D.T.R., Baer, R.A., Harris, M.J., 1991. Detection of malingering on the MMPI: A meta-analysis. Clinical Psychology Review 11, 585–598.

Boe, A.J., McGee Koch, L.L., O'Brien, M.K., Shawen, N., Rogers, J.A., Lieber, R.L., Reid, K.J., Zee, P.C., Jayaraman, A., 2019. Automating sleep stage classification using wireless, wearable sensors. NPJ Digital Medicine 2, 131.

Box, G.E.P., Jenkins, G.M., Reinsel, G.C., 1994. Time series analysis: Forecasting and control. Prentice-Hall, Englewood Cliffs, N.J.

Brobbin, E., Deluca, P., Hemrage, S., Drummond, C., 2022. Accuracy of wearable transdermal alcohol sensors: Systematic review. Journal of Medical Internet Research 24, e35178.

Caddy, G.R., Sobell, M.B., Sobell, L.C., 1978. Alcohol breath tests: Criterion times for avoiding contamination by "mouth alcohol." Behavior Research Methods & Instrumentation 10, 814–818. https://doi.org/10.3758/BF03205406

Campbell, D.T., Fiske, D.W., 1959. Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological Bulletin 56, 81–105.

Chen, T., Guestrin, C., 2016. Xgboost: A scalable tree boosting system, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 785–794.

Chlingaryan, A., Sukkarieh, S., Whelan, B., 2018. Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review. Computers and Electronics in Agriculture 151, 61–69.

Christ, M., Braun, N., Neuffer, J., Kempa-Liehr, A.W., 2018. Time series feature extraction on basis of scalable hypothesis tests (tsfresh–a python package). Neurocomputing 307, 72–77.

Courtney, J.B., Russell, M.A., Conroy, D.E., 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.

Crunelle, C.L., Yegles, M., van Nuijs, A.L.N., Covaci, A., De Doncker, M., Maudens, K.E., Sabbe, B., Dom, G., Lambert, W.E., Michielsen, P., 2014. Hair ethyl glucuronide levels as a marker for alcohol use and abuse: A review of the current state of the art. Drug and Alcohol Dependence 134, 1–11.

Davis-Martin, R.E., Alessi, S.M., Boudreaux, E.D., 2021. Alcohol use disorder in the age of technology: A review of wearable biosensors in alcohol use disorder treatment. Frontiers in Psychiatry 12, 246.

Del Boca, F.K., Darkes, J., 2003. The validity of self-reports of alcohol consumption: State of the science and challenges for research. Addiction 98, 1–12.

Delgado, M.K., Huang, Y., Wanner, K., Goldberg, E., Hemmons, J., Spencer, E., Wetherill, R., 2017. Test accuracy of smartphone-paired breathalysers: A validation study. Injury Prevention 23.

DiClemente, C.C., Prochaska, J.O., 1998. Toward a comprehensive, transtheoretical model of change: Stages of change and addictive behaviors., in: Miller, W.R., Heather, N. (Eds.), Treating Addictive Behaviors. Plenum Press, New York, NY, pp. 3–24.

Didier, N.A., King, A.C., Polley, E.C., Fridberg, D.J., 2023. Signal processing and machine learning with transdermal alcohol concentration to predict natural environment alcohol consumption. Experimental and Clinical Psychopharmacology 32, 245–254. https://doi.org/10.1037/pha0000683

Ditzer, J., Wong, E.Y., Modi, R.N., Behnke, M., Gross, J.J., Talmon, A., 2023. Child maltreatment and alexithymia: A meta-analytic review. Psychological Bulletin 149, 311–329.

Dooley, N., Kennelly, B., Arseneault, L., Zammit, S., Whelan, R., Mosley, O., Cotter, D., Clarke, M., Cotter, D.R., Kelleher, I., McGorry, P., Healy, C., Cannon, M., 2023. Functional Outcomes Among Young People With Trajectories of Persistent Childhood Psychopathology. JAMA Netw Open 6, e2336520. https://doi.org/10.1001/jamanetworkopen.2023.36520

Dougherty, D.M., Hill-Kapturczak, N., Liang, Y., Karns, T.E., Cates, S.E., Lake, S.L., Mullen, J., Roache, J.D., 2014. Use of continuous transdermal alcohol monitoring during a contingency management procedure to reduce excessive alcohol use. Drug & Alcohol Dependence 142, 301–306.

Dumett, M.A., Rosen, I.G., Sabat, J., Shaman, A., Tempelman, L., Wang, C., Swift, R.M., 2008. Deconvolving an estimate of breath measured blood alcohol concentration from biosensor collected transdermal ethanol data. Applied Mathematics and Computation 196, 724–743. https://doi.org/10.1016/j.amc.2007.07.026

Edwards, J.R., 2008. To prosper, organizational psychology should ... overcome methodological barriers to progress. J Organ Behavior 29, 469–491. https://doi.org/10.1002/job.529

Ekman, P., Friesen, W.V., Hager, J.C., 2002. Facial Action Coding System. Network Information Research, Salt Lake City, UT.

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

Fairbairn, C.E., Bosch, N., in press. Applying Artificial Intelligence to expand the measurement toolkit in clinical psychological science: Moving beyond self-reports. Clinical Psychological Science.

Fairbairn, C.E., Bresin, K., Kang, D., Rosen, I.G., Ariss, T., Luczak, S.E., Barnett, N.P., Eckland, N.S., 2018. A multimodal investigation of contextual effects on alcohol's

emotional rewards. Journal of Abnormal Psychology 127, 359–373. https://doi.org/10.1037/abn0000346

Fairbairn, C.E., Han, J., Caumiant, E.P., Benjamin, A.S., Bosch, N., 2025. A wearable alcohol biosensor: Exploring the accuracy of transdermal drinking detection. Drug and Alcohol Dependence 266, 112519.

Fairbairn, C.E., Kang, D., 2020. Transdermal alcohol monitors: Research, applications, and future directions, in: Frings, D., Albery, I. (Eds.), The Handbook of Alcohol Use and Abuse. Elsevier.

Fairbairn, C.E., Kang, D., 2019. Temporal dynamics of transdermal alcohol concentration measured via new-generation wrist-worn biosensor. Alcoholism: Clinical and Experimental Research 43, 2060–2069. https://doi.org/10.1111/acer.14172

Fairbairn, C.E., Kang, D., Bosch, N., 2020. Using machine learning for real-time BAC estimation from a new-generation transdermal biosensor in the laboratory. Drug and Alcohol Dependence 216, 108205. https://doi.org/10.1016/j.drugalcdep.2020.108205

Fairbairn, C.E., Kang, D., Federmeier, K.D., 2021. Alcohol and neural dynamics: A meta-analysis of acute alcohol effects on event-related brain potentials. Biological Psychiatry 89, 990–1000.

Fakhari, S., Waszkiewicz, N., 2023. Old and new biomarkers of alcohol abuse: Narrative review. Journal of Clinical Medicine 12, 2124.

Farber, B.A., 2020. Client deception about substance use: Research findings and a case study. J Clin Psychol 76, 277–285. https://doi.org/10.1002/jclp.22894

Garcia, J., Gustavson, A.R., 1997. The science of self-report. APS Observer 10.

Gibb, K.A., Yee, A.S., Johnston, C.C., Martin, S.D., Nowak, R.M., 1984. Accuracy and usefulness of a breath alcohol analyzer. Annals of Emergency Medicine 13, 516–520. https://doi.org/10.1016/s0196-0644(84)80517-x

Giles, H.G., Meggiorini, S., Renaud, G.E., Thiessen, J.J., Vidins, E.I., Compton, K.V., Saldivia, V., Orrego, H., Israel, Y., 1987. Ethanol vapor above skin: Determination by a gas sensor instrument and relationship with plasma concentration. Alcoholism: Clinical and Experimental Research 11, 249–253.

Grant, B.F., Goldstein, R.B., Saha, T.D., Chou, S.P., Jung, J., Zhang, H., Pickering, R.P., Ruan, W.J., Smith, S.M., Huang, B., Hasin, D.S., 2015. Epidemiology of DSM-5 Alcohol Use Disorder: Results From the National Epidemiologic Survey on Alcohol and Related Conditions III. JAMA Psychiatry 72, 757–766. https://doi.org/10.1001/jamapsychiatry.2015.0584

Griswold, M.G., Fullman, N., Hawley, C., Arian, N., Zimsen, S.R., Tymeson, H.D., Venkateswaran, V., Tapp, A.D., Forouzanfar, M.H., Salama, J.S., 2018. Alcohol use and burden for 195 countries and territories, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. The Lancet 392, 1015–1035.

Gurrieri, L., Fairbairn, C.E., Sayette, M.A., Bosch, N., 2021. Alcohol narrows physical distance between strangers. Proceedings of the National Academy of Sciences 118, e2101937118. https://doi.org/10.1073/pnas.2101937118

Harari, G.M., Lane, N.D., Wang, R., Crosier, B.S., Campbell, A.T., Gosling, S.D., 2016. Using smartphones to collect behavioral data in psychological science: Opportunities,

practical considerations, and challenges. Perspect Psychol Sci 11, 838–854. https://doi.org/10.1177/1745691616650285

Harris, J.C., Leggio, L., Farokhnia, M., 2021. Blood biomarkers of alcohol use: A scoping review. Current Addiction Reports 8, 500–508.

Haynes, S.N., Smith, G., Hunsley, J., 2019. Scientific foundations of clinical assessment, 2nd edition. ed. Routledge, New York, NY.

Hunsley, J., Mash, E.J., 2007. Evidence-Based Assessment. Annu. Rev. Clin. Psychol. 3, 29–51. https://doi.org/10.1146/annurev.clinpsy.3.022806.091419

Jones, A.W., 1996. Measuring alcohol in blood and breath for forensic purposes: A historical review. Forensic Science Review 8, 13–44.

Jones, N., 2017. How machine learning could help to improve climate forecasts. Nature 548.

Jones, R.L., 2021. Last Chance Texaco: Chronicles of an American Troubadour. Grove Press.

Keller, P.S., El-Sheikh, M., Keiley, M., Liao, P., 2009. Longitudinal relations between marital aggression and alcohol problems. Psychology of Addictive Behaviors 23, 2–13.

Kendler, K.S., 2005. Toward a Philosophical Structure for Psychiatry. AJP 162, 433–440. https://doi.org/10.1176/appi.ajp.162.3.433

Kerr, W.C., Patterson, D., Koenen, M.A., Greenfield, T.K., 2008. Alcohol content variation of bar and restaurant drinks in Northern California. Alcoholism: Clinical and Experimental Research 32, 1623–1629.

Kim, J., Jeerapan, I., Imani, S., Cho, T.N., Bandodkar, A., Cinti, S., Mercier, P.P., Wang, J., 2016. Noninvasive alcohol monitoring using a wearable tattoo-based iontophoretic-biosensing system. ACS Sens. 1, 1011–1019. https://doi.org/10.1021/acssensors.6b00356

Kranzler, H.R., Soyka, M., 2018. Diagnosis and pharmacotherapy of alcohol use disorder: A review. JAMA 320, 815–824.

Krendl, A.C., Perry, B.L., 2023. Stigma toward substance dependence: Causes, consequences, and potential interventions. Psychol Sci Public Interest 24, 90–126. https://doi.org/10.1177/15291006231198193

Krystal, J.H., State, M.W., 2014. Psychiatric Disorders: Diagnosis to Therapy. Cell 157, 201–214. https://doi.org/10.1016/j.cell.2014.02.042

Lansdorp, B., Ramsay, W., Hamid, R., Strenk, E., 2019. Wearable enzymatic alcohol biosensor. Sensors 19, 2380.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444.

Lee, H., Pham, P., Largman, Y., Ng, A.Y., 2009. Unsupervised feature learning for audio classification using convolutional deep belief networks, in: Advances in Neural Information Processing Systems. pp. 1096–1104.

Lee, S.I., Nam, H.S., Garst, J.H., Huang, A., Campion, A., Arnell, M., Ghalehsariand, N., Park, S., Chang, H., Lu, D.C., Sarrafzadeh, M., Park, E., 2017. Unobtrusive and Continuous Monitoring of Alcohol-impaired Gait Using Smart Shoes. Methods Inf Med 56, 74–82. https://doi.org/10.3414/me15-02-0008

Lilienfeld, S.O., Fowler, K.A., 2006. The self-report assessment of psychopathy, in: Patrick, C. (Ed.), Handbook of Psychopathy. Guilford Press, New York, pp. 107–132.

Luczak, S.E., Ramchandani, V.A., 2019. Special issue on alcohol biosensors: Development, use, and state of the field. Alcohol 81, 161–165.

Luczak, S.E., Rosen, I.G., 2014. Estimating BrAC from transdermal alcohol concentration data using the BrAC estimator software program. Alcoholism: Clinical and Experimental Research 38, 2243–2252. https://doi.org/10.1111/acer.12478

Ma, Z., Yang, G., Yang, Y., Gao, Z., Wang, J., Du, Z., Yu, F., Chen, Q., Zheng, S., Zhang, S., 2025. Speech recognition meets large language model: Benchmarking, models, and exploration, in: Proceedings of the AAAI Conference on Artificial Intelligence. pp. 24840–24848.

Maisto, S.A., Saitz, R., 2003. Alcohol use disorders: Screening and diagnosis. American J Addict 12. https://doi.org/10.1111/j.1521-0391.2003.tb00493.x

Marlatt, G.A., 1996. Taxonomy of high-risk situations for alcohol relapse: Evolution and development of a cognitive-behavioral model. Addiction 91, 37–50. https://doi.org/10.1046/j.1360-0443.91.12s1.15.x

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

Mehl, M.R., Eid, M., Wrzus, C., Harari, G.M., Ebner-Priemer, U., 2024. Handbook of Mobile Sensing in Psychology: Methods and Applications.

Murgia, S., Fairbairn, C.E., Barnett, N.P., Flanagan, J.C., Bosch, N., in press. Acceptability and feasibility of a new-generation alcohol biosensor: A mixed-methods evaluation in a large community sample. Experimental and Clinical Psychopharmacology.

NIAAA, 2015. Wearable alcohol biosensor challenge [WWW Document]. URL https://www.nih.gov/news-events/news-releases/niaaa-selects-winners-its-wearable-alcohol-biosensor-challenge

Nisbett, R.E., Wilson, T.D., 1977. Telling more than we can know: Verbal reports on mental processes. Psychological Review 84, 231–259.

Norman, T., Peacock, A., Ferguson, S.G., 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, 1112–1121. https://doi.org/10.1111/dar.13215

Norvig, P., Russell, S.J., 2021. Artificial intelligence: A modern approach, 4th ed. Pearson.

Nyman, E., Palmlöv, A., 1936. The elimination of ethyl alcohol in sweat. Skandinavisches Archiv Für Physiologie 74, 155–159. https://doi.org/10.1111/j.1748-1716.1936.tb01150.x

Oszkinat, C., Luczak, S.E., Rosen, I.G., 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 1–8. https://doi.org/10.1109/TNNLS.2022.3140726

Pew Research Center, 2022. Internet, smartphone and social media use.

Pew Research Center, 2018. Mobile Fact Sheet.

Piasecki, T.M., 2019. Assessment of alcohol use in the natural environment. Alcoholism: Clinical and experimental research 43, 564–577.

Podsakoff, P.M., MacKenzie, S.B., Lee, J., Podsakoff, N.P., 2003. Common method biases in behavioral research: A critical review of the literature and recommended remedies. Journal of Applied Psychology 88, 879–903.

Podsakoff, P.M., Podsakoff, N.P., Williams, L.J., Huang, C., Yang, J., 2024. Common method bias: It's bad, it's complex, it's widespread, and it's not easy to fix. Annu. Rev. Organ. Psychol. Organ. Behav. 11, 17–61. https://doi.org/10.1146/annurev-orgpsych-110721-040030

Puterman, E., Pauly, T., Ruissen, G., Nelson, B., Faulkner, G., 2021. Move more, move better: A narrative review of wearable technologies and their application to precision health. Health Psychology 40, 803–810.

Rash, C.J., Petry, N.M., Alessi, S.M., Barnett, N.P., 2019. Monitoring alcohol use in heavy drinking soup kitchen attendees. Alcohol 81, 139–147.

Rees, L.M., Tombaugh, T.N., Gansler, D.A., Moczynski, N.P., 1998. Five validation experiments of the Test of Memory Malingering (TOMM). Psychological Assessment 10, 10–20.

Richards, V.L., Glenn, S.D., Turrisi, R.J., Mallett, K.A., Ackerman, S., Russell, M.A., 2024. Transdermal alcohol concentration features predict alcohol-induced blackouts in college students. Alcohol, Clinical and Experimental Research 48, 880–888. https://doi.org/10.1111/acer.15290

Ridder, T., Ver Steeg, B., Laaksonen, B.D., 2009. Comparison of spectroscopically measured tissue alcohol concentration to blood and breath alcohol measurements. Journal of Biomedical Optics 14, 054039-1-054039–11.

Roache, J.D., Karns, T.E., Hill-Kapturczak, N., Mullen, J., Liang, Y., Lamb, R.J., Dougherty, D.M., 2015. Using transdermal alcohol monitoring to detect low-level drinking. Alcoholism: Clinical and Experimental Research 39, 1120–1127.

Roberts, W., McKee, S.A., 2019. Mobile alcohol biosensors and pharmacotherapy development research. Alcohol 81, 149–160.

Robins, L.N., 1981. National Institute of Mental Health Diagnostic Interview Schedule: Its History, Characteristics, and Validity. Arch Gen Psychiatry 38, 381. https://doi.org/10.1001/archpsyc.1981.01780290015001

Romano, E., Tremblay, R.E., Vitaro, F., Zoccolillo, M., Pagani, L., 2001. Prevalence of Psychiatric Diagnoses and the Role of Perceived Impairment: Findings from an Adolescent Community Sample. Child Psychology Psychiatry 42, 451–461. https://doi.org/10.1111/1469-7610.00739

Rosenberg, M., Kianersi, S., Luetke, M., Jozkowski, K., Guerra-Reyes, L., Shih, P.C., Finn, P., Ludema, C., 2023. Wearable alcohol monitors for alcohol use data collection among college students: feasibility and acceptability. Alcohol 111, 75–83.

Russell, M.A., Turrisi, R.J., Smyth, J.M., 2022. Transdermal sensor features correlate with ecological momentary assessment drinking reports and predict alcohol-related consequences in young adults' natural settings. Alcoholism: Clinical and Experimental Research 46, 100–113.

Sanislow, C.A., Pine, D.S., Quinn, K.J., Kozak, M.J., Garvey, M.A., Heinssen, R.K., Wang, P.S.-E., Cuthbert, B.N., 2010. Developing constructs for psychopathology research:

Research domain criteria. Journal of Abnormal Psychology 119, 631–639. https://doi.org/10.1037/a0020909

Schuckit, M.A., Clarke, D.F., Smith, T.L., Mendoza, L.A., 2020. Characteristics associated with denial of problem drinking among two generations of individuals with alcohol use disorders. Drug and Alcohol Dependence 217, 108274. https://doi.org/10.1016/j.drugalcdep.2020.108274

Schwarz, N., 2007. Cognitive aspects of survey methodology. Applied Cognitive Psychology 21, 277–287. https://doi.org/10.1002/acp.1340

Schwarz, N., 1999. Self-reports: How the questions shape the answers. American Psychologist 54, 93–105. https://doi.org/10.1037/0003-066X.54.2.93

Sher, K.J., Trull, T.J., 1996. Methodological Issues in Psychopathology Research. Annu. Rev. Psychol. 47, 371–400. https://doi.org/10.1146/annurev.psych.47.1.371

Shiffman, S., 2009. Ecological momentary assessment (EMA) in studies of substance use. Psychological Assessment 21, 486–497. https://doi.org/10.1037/a0017074

Sobell, L.C., Sobell, M.B., 1992. Timeline follow-back, in: Litten, R.Z., Allen, J.P., Totowa, N.J. (Eds.), Measuring Alcohol Consumption. Humana Press, Totowa, MJ, pp. 41–72.

Stubbs, R.J., O'Reilly, L.M., Whybrow, S., Fuller, Z., Johnstone, A.M., Livingstone, M.B.E., Ritz, P., Horgan, G.W., 2014. Measuring the difference between actual and reported food intakes in the context of energy balance under laboratory conditions. Br J Nutr 111, 2032–2043. https://doi.org/10.1017/S0007114514000154

Swift, R.M., 2003. Direct measurement of alcohol and its metabolites. Addiction 98, 73–80.

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

Swift, R.M., 1993. Transdermal measurement of alcohol consumption. Addiction 88, 1037–1039.

Tackett, J.L., Brandes, C.M., King, K.M., Markon, K.E., 2019. Psychology's replication crisis and clinical psychological science. Annu. Rev. Clin. Psychol. 15, 579–604. https://doi.org/10.1146/annurev-clinpsy-050718-095710

Thomaz, E., Essa, I., Abowd, G.D., 2015. A practical approach for recognizing eating moments with wrist-mounted inertial sensing, in: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing. Presented at the UbiComp '15: The 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing, ACM, Osaka Japan, pp. 1029–1040. https://doi.org/10.1145/2750858.2807545

Tourangeau, R., Yan, T., 2007. Sensitive questions in surveys. Psychological Bulletin 133, 859–883.

Tucker, J.S., Ellickson, P.L., Collins, R.L., Klein, D.J., 2006. Does solitary substance use increase adolescents' risk for poor psychosocial and behavioral outcomes? A 9-year longitudinal study comparing solitary and social users. Psychology of Addictive Behaviors 20, 363.

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

Vazire, S., 2010. Who knows what about a person? The self–other knowledge asymmetry (SOKA) model. Journal of Personality and Social Psychology 98, 281–300.

Vinu Mohan, A.M., Windmiller, J.R., Mishra, R.K., Wang, J., 2017. Continuous minimally-invasive alcohol monitoring using microneedle sensor arrays. Biosens Bioelectron 91, 574–579. https://doi.org/10.1016/j.bios.2017.01.016

Vollmer, N., 2020. Recital 71 EU General Data Protection Regulation (EU-GDPR) [WWW Document]. URL http://www.privacy-regulation.eu/en/recital-71-GDPR.htm (accessed 2.8.21).

Wang, R., Chen, F., Chen, Z., Li, T., Harari, G., Tignor, S., Zhou, X., Ben-Zeev, D., Campbell, A.T., 2014. StudentLife: Assessing mental health, academic performance and behavioral trends of college students using smartphones, in: Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing. pp. 3–14.

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

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

Watkins, C.E., Campbell, V.L., Nieberding, R., Hallmark, R., 1995. Contemporary practice of psychological assessment by clinical psychologists. Professional Psychology: Research and Practice 26, 54–60. https://doi.org/10.1037/0735-7028.26.1.54

Webster, G.D., Gabler, H.C., 2007. Feasibility of transdermal ethanol sensing for the detection of intoxicated drivers. Annu Proc Assoc Adv Automot Med 51, 449–464.

Weissenborn, R., Duka, T.A., 2003. Acute alcohol effects on cognitive function in social drinkers: Their relationship to drinking habits. Psychopharmacology 165, 306–312. https://doi.org/10.1007/s00213-002-1281-1

White, A.M., 2003. What happened? Alcohol, memory blackouts, and the brain. Alcohol Research and Health 27, 186–196.

Widmark, E.M.P., 1932. Die theoretischen Grundlagen und die praktische Verwendbarkeit der gerichtlich-medizinischen Alkoholbestimmung. Urban & Schwarzenberg.

Williams, D., Botting, N., Boucher, J., 2008. Language in autism and specific language impairment: Where are the links? Psychological Bulletin 134, 944–963.

World Health Organization, 2022. ICD-11: International classification of diseases.

Xu, Y., Kong, Q., Wang, W., Plumbley, M.D., 2018. Large-scale weakly supervised audio classification using gated convolutional neural network, in: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, pp. 121–125.

Yu, J., Fairbairn, C.E., Gurrieri, L., Caumiant, E.P., 2022. Validating transdermal alcohol biosensors: A meta-analysis of associations between BAC and transdermal alcohol sensor output. Addiction 117.

Zwaigenbaum, L., Penner, M., 2018. Autism spectrum disorder: advances in diagnosis and evaluation. BMJ k1674. https://doi.org/10.1136/bmj.k1674

Terms and Definitions List

Noise: Unpredictable fluctuations in data with the potential to obscure a true relationship between variables (“random” variability)

Confound: An extraneous factor that correlates with key study variables, so distorting (inflating or deflating) observed effects (“systematic” variability)

Common/Shared Methods Bias: A systematic form of error with the potential to bias associations between variables measured using the same method of measurement

Blood Alcohol Concentration (*BAC*): Measure of alcohol intoxication assessed either directly within blood/plasma or, more often, via breathalyzer

Transdermal Alcohol Concentration (*TAC*): Vaporized alcohol present in the air pocket immediately above the skin. Includes a combination of evaporated sweat and insensible perspiration

Accuracy: Operationalized as congruence between transdermal estimates of drinking and “true” drinking (e.g., as measured using *BAC*, standard drinks, or broader abstinence/risk-level)

Temporal Specificity: Operationalized as temporal displacement between transdermal estimates of drinking and "true" drinking

Time Series Analysis: Used to examine patterns in series of observations collected over time. Useful for forecasting future trends.

Machine Learning: A family of analytic approaches wherein algorithms “learn” patterns and associations through direct exposure to data

Methodless Findings Fallacy: The misconception that scientific results can be judged independent of the specific analytic and device tools used to achieve them

Perfect Solution Fallacy: The concept that an ideal method of measurement exists well-suited for all potential applications

Common Methods Bias

Measures that use the *same method* to measure *different constructs* often produce results more similar than do measures that assess the *same construct* using *different methods* (Campbell and Fiske, 1959). Such effects arise from *shared methods bias* (also called *common methods bias*)—a systematic form of error variance that can represent a substantial source of bias in behavioral research (Campbell and Fiske, 1959; Edwards, 2008; Podsakoff et al., 2003) (see Figure 1). While in theory common methods bias might either magnify (inflate) or diminish effect sizes, results of meta-analyses indicate influences are overwhelmingly inflationary, enlarging effect sizes by an estimated 120-160% in behavioral research (Podsakoff et al., 2024).

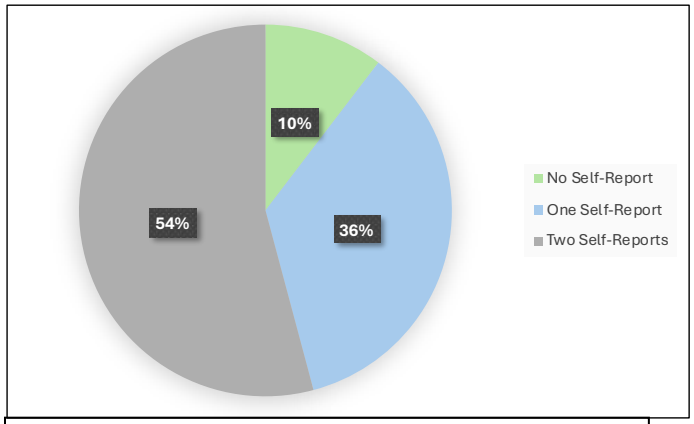
Factors that have been established or theorized as potential contributors to common methods bias in survey studies include strong affective states among participants, implicit theories, response style, and stigma/social desirability concerns (Fairbairn and Bosch, in press; Podsakoff et al., 2024). Each of these elements has the potential for exacerbation within clinical populations and assessment contexts, wherein self-report assessment is pervasive, so leading to a potentially outsized influence for common methods bias in clinical psychological research (Fairbairn and Bosch, in press).

Table 1. List of wearable alcohol biosensor applications together with predicted application-specific requirements for accuracy and temporal specificity

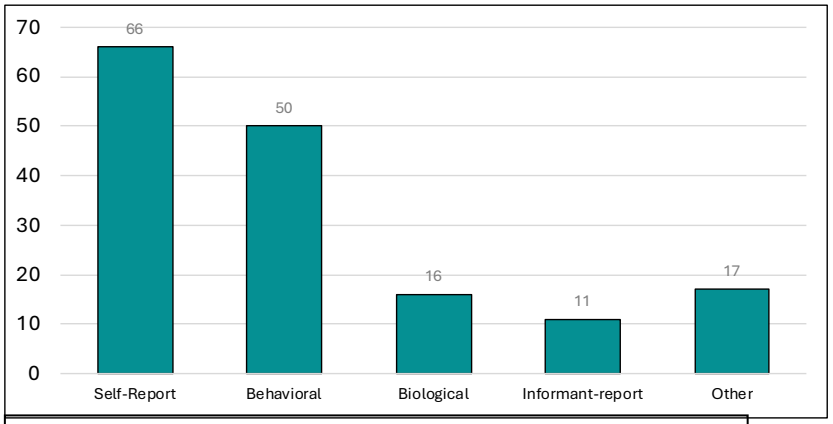
Application		Performance Requirements	
Domain	**Description**	**Accuracy**	**Temporal Spec**
Prevention	**Health Tracker:** Day-level record of abstinence and/or high-risk drinking episodes for longer term wear among everyday drinkers (e.g., Fitbit for alcohol)	Mod/Low	Low
Intervention	**Treatment-Adherence/Relapse Tracker**: Objective, day-level record of abstinence and/or high-risk drinking episodes for patients enrolled in MI, HR, or CM programs	Mod/Low	Low
	JITAI/Relapse Prevention: Real-time detection of drinking among those in recovery, prompting intervention and thus diminishing relapse severity/duration	Low	High
Medical	**Medical Monitor**: Recording objective drinking risk-level for providers treating patients with medical diagnoses contraindicating high-level consumption (e.g., heart disease, diabetes, etc.)	Mod/Low	Low
	***BAC* Monitor**: Non-invasive, real-time *BAC* monitor for patients presenting to emergency medical facilities with clinically relevant intoxication	High	High
Motor Vehicle	**Driving Safety Monitor**: Passive real-time *BAC* estimation prompting alarms or contacting a ride sharing service when *BAC* approaching higher levels	High/Mod	High
Research	**Longitudinal Research**: Objective day-level record of changes in abstinence and/or high-risk drinking over months/years	Mod/Low	Low
	Laboratory *BAC* Monitor: Continuous, passive, real-time *BAC* monitor for participants in laboratory studies involving alcohol	High	High
	Basic Field Research: Assessing immediate causes/consequences of alcohol use in real-world contexts	Mod/Low	High/Mod
	Intervention Outcome: Providing an objective day-level record of drinking following participation in trials of novel addiction interventions	Mod/Low	Mod/Low

Note: displays predictions regarding the minimum level of accuracy and temporal specificity (low/moderate/high) required for useful application of a wearable alcohol sensor in identified domain.
Accuracy: Congruency between sensor-based estimates of alcohol use and "true" alcohol use
Temporal Specificity: Temporal displacement between sensor-based estimates of alcohol use and "true" alcohol use
BAC=Blood alcohol concentration; e*BAC*=Sensor-based estimates of *BAC*; Mod=Moderate; MI=Motivational interviewing; HR=Harm Reduction; CM=Contingency management; JITAI=Just-in-time adaptive interventions.

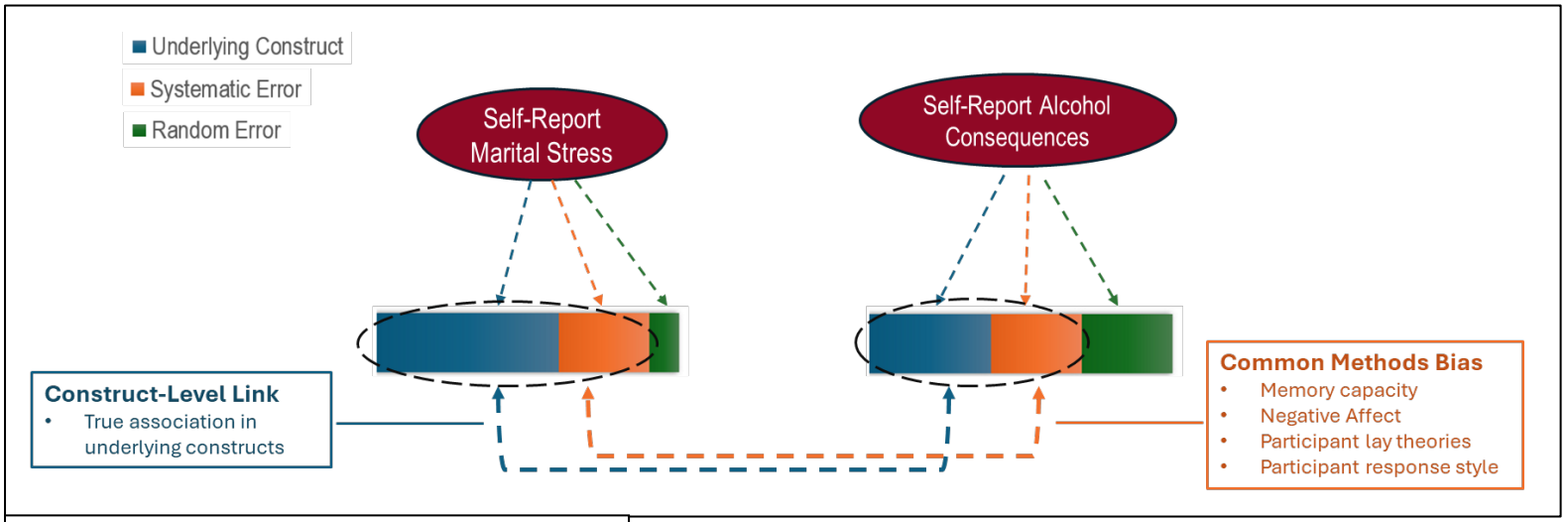
Figure 1. Measure Types Used in Highly Ranked Clinical Psychology Outlets 2021-24 (Top Panel) and Illustration of Common Methods Bias (Bottom Panel)



A. Proportion Studies Employing Self-Reports



B. Measure Types Employed for Assessing Behavior



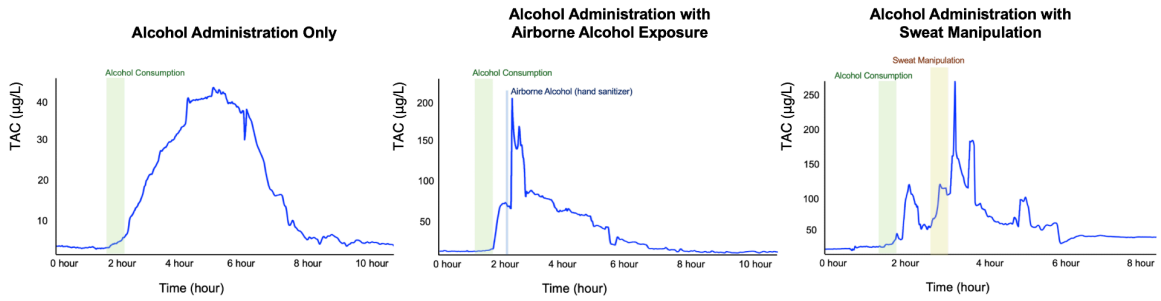
C. Illustration of Common Methods Bias

Note: [Top Panel] A literature review was conducted to assess measure-type employed in tests of primary study aims in articles published in the journals *Clinical Psychological Science*, *Journal of Consulting and Clinical Psychology*, and *Journal of Psychopathology and Clinical Science* from July 1, 2021-July 1, 2024 (Fairbairn & Bosch, in press). A total of 386 correlational and 160 randomized studies were identified and coded. Self-reports were defined as questionnaire or interview-type measures. Review procedures and hypotheses were pre-registered at Open Science Framework (see https://osf.io/cvnk8). The top lefthand panel (A) depicts the proportion of non-experimental studies that tested associations between primary variables of interest using one vs two vs no closed-ended self-reports. The top righthand panel (B) reflects counts at the level of the measure, reflecting measure-type employed specifically for assessing behavioral constructs.

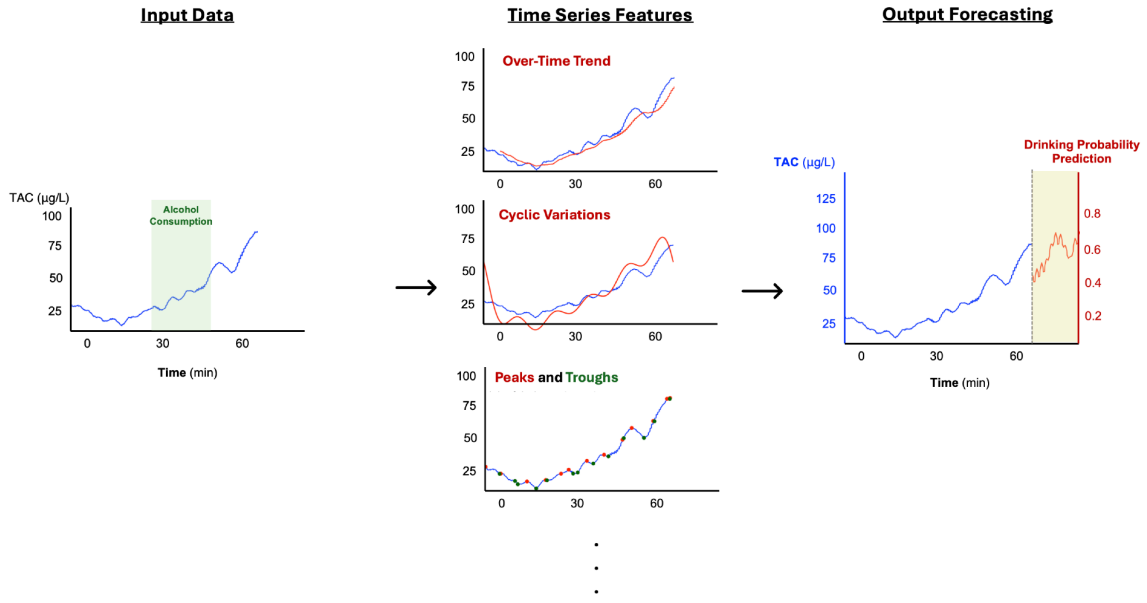
[Bottom Panel (C)] In the examination of relationships between multiple variables, significant correlations can emerge attributable to not only shared variation in the underlying constructs, but also to shared systematic forms of error linked with the method of measurement (see also Podsakoff et al., 2024).Variability in measured constructs can be divided into three components: variability in the true underlying construct, random error variance, and systematic error variance. A meta-analysis suggested that inflationary effects of self-report common methods bias for effect sizes in behavioral research can reach levels as high as 250% (Podsakoff et al., 2024).

Figure 2. Illustration of Utility of Over-Time Data for Parsing Contextual Confounds and Future Forecasting

A. Characteristic *TAC* Curves Linked with Drinking, Sweating, and Airborne Alcohol



B. Example Application of TAC Times Series Features for Future Forecasting

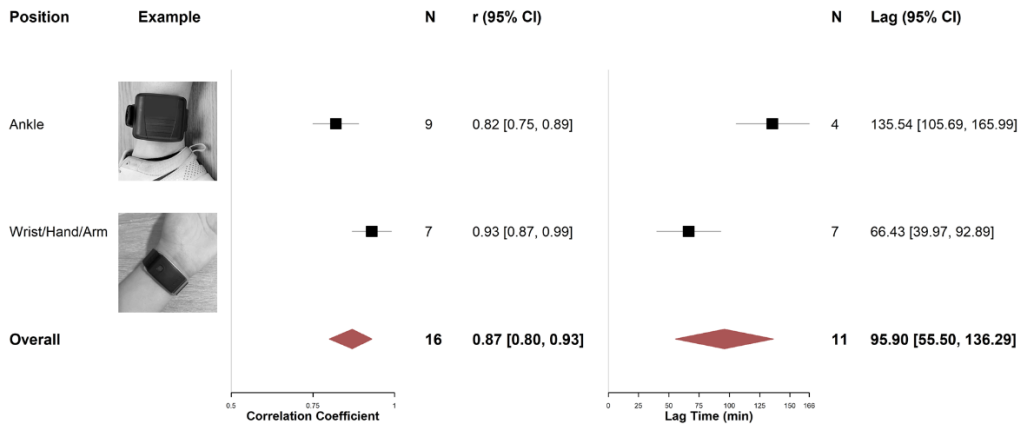


Note. The top panel (A) represents new-generation transdermal sensor data derived from a laboratory-based alcohol-administration paradigm. Curves illustrate how dense time series can differentiate characteristic over-time patterns linked with drinking (gradual *TAC* rise and gradual fall) vs airborne alcohol (sharp rise and sharp fall) and sweating (sharp rise followed by gradual fall). The bottom panel (B) provides examples of features that can be extracted from time-series data, and the application of these features for addressing temporal lags in the relationship between TAC and ingested alcohol to produce estimates of drinking (i.e., via future forecasting).

Figure 3. New- and Old- Generation Transdermal Alcohol Sensors

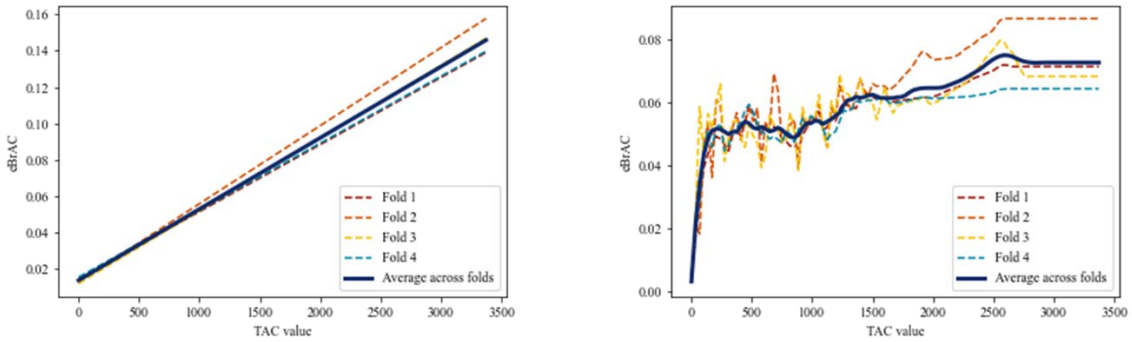
A. Images of New-Generation Transdermal Wrist Sensors

B. Meta-Analysis Comparing Effects of Device Type/Body Position on Transdermal Sensor Performance



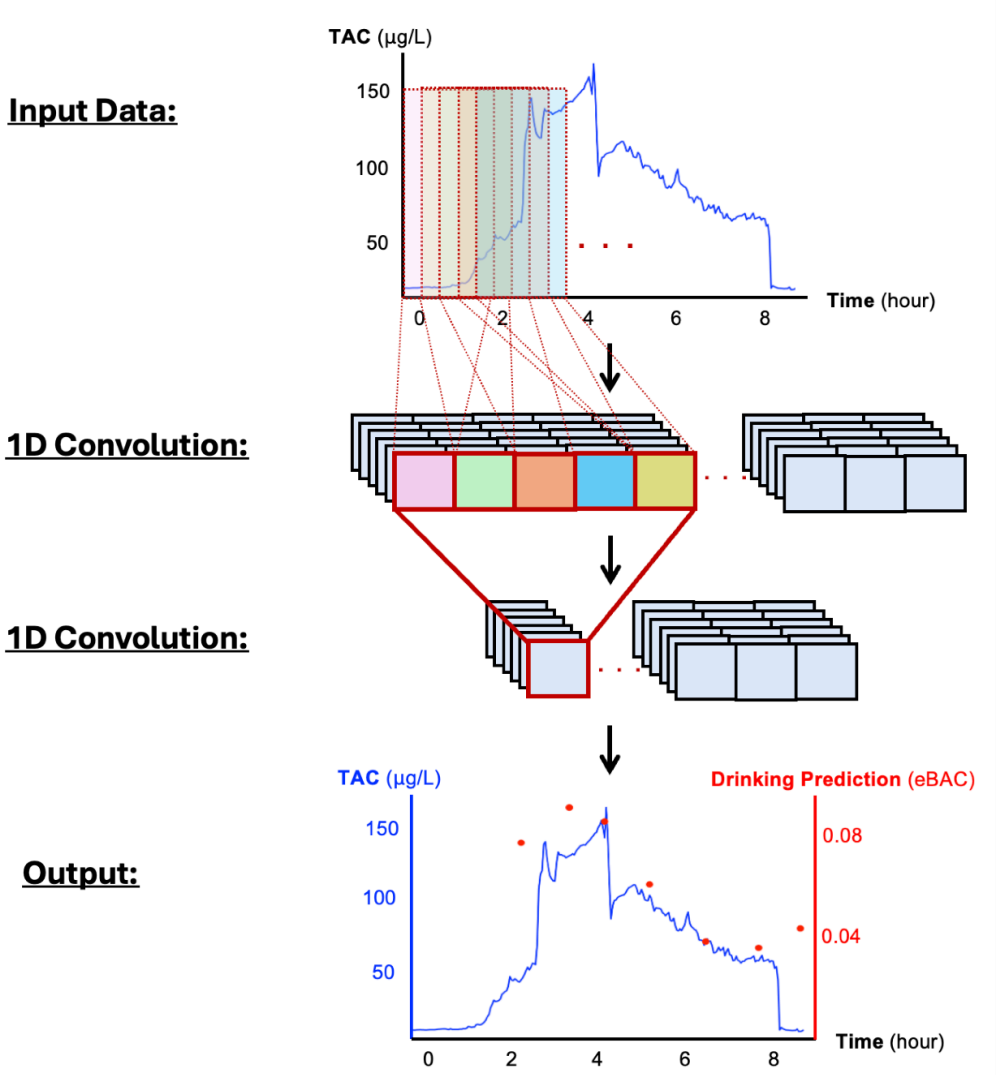
Note. The top panel (A) displays images of four new-generation, smartphone-connected transdermal wrist sensors. The bottom panel (B) displays results of a meta-analysis of 30 studies of transdermal sensors, mainly based in laboratory settings. Results indicated that lag times for ankle-worn sensors were approximately double the magnitude of lags for sensors worn on the arm, wrist, or hand, and *TAC-BAC* correlations tended to be lower for ankle-worn devices (Yu et al., 2022)

Figure 4. Comparison Between Linear Regression and Machine Learning for Modeling the Relationship between Blood and Transdermal Alcohol Measurement



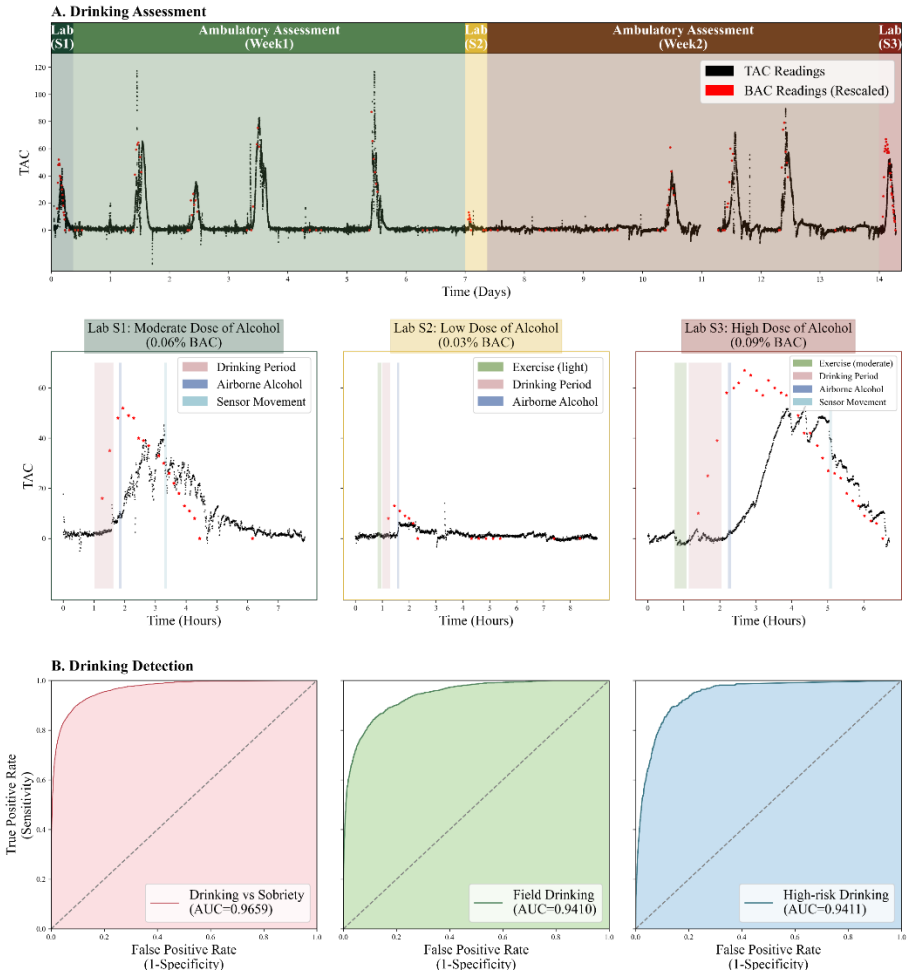
Note. The relationship between transdermal alcohol concentration (*TAC*) and blood alcohol concentration (*BAC*) as estimated using linear regression (lefthand panel) and Extra-Trees machine-learning (righthand panel). "Folds" refer to portions of the data held out for testing during cross-validation (Fairbairn & Bosch, 2021)

Figure 5. Example Convolutional Neural Network Model Application to Transdermal Sensor Time-Series Data for the Estimation of Blood Alcohol Concentration (eBAC)



Note. The top panel displays the input *TAC* signal (blue line) over time, with magnified elements highlighting the sliding time window used to extract temporal patterns. Each window is used as an input example to the model, and multiple overlapping windows are passed through the network. The second panel displays the 1D convolution layer, within which filters are applied to these windows to learn local temporal time-series features within the *TAC* curve signal. These convolutional operations allow the model to learn features that are predictive of alcohol consumption events but not tied to specific moments in time. A second convolutional layer (third panel) further processes these extracted features to form meaningful representations of the *TAC* dynamics. The output layer produces a point prediction of *eBAC* for each input window. The bottom panel shows an example set of predicted *eBAC* values (red dots) aligned with the original *TAC* curve (blue line) over time, illustrating how the model produces estimates of *BAC* based on the shape and trajectory of the *TAC* signal.

Figure 6. Combined Laboratory-Ambulatory Study for Testing Transdermal Alcohol Monitors Introducing Contextual Variability in Dosing Conditions



Note. (A) Example of raw *TAC* and *BAC* data collected from a single participant in both laboratory and field contexts (Fairbairn et al., 2025). Ambulatory assessment lasted 14 days, during which time participants wore transdermal sensors and supplied *BAC* readings in real-world contexts in response to custom prompts. *BAC* data is scaled up by a factor of 1000 for visualization purposes. (B) Results based on N=100 participants to enroll in trial. Leftmost graph reflects area under the receiver operating characteristic curve (AUROC) for real-time transdermal models in differentiating BAC > 0.00% vs BAC = 0.00% within a combined laboratory and field dataset; center graph reflects AUROC in differentiating BAC > 0.00% vs BAC = 0.00% in field-dataset only, and the rightmost graph reflects AUROC in differentiating BAC ≥ 0.08% vs BAC < 0.08% within a combined laboratory and field dataset.