

**Performance and Equity of Contextualized Geolocation Data for Lapse Prediction in
Alcohol Use Disorder**

Claire Punturieri^{aff-1}, Christopher Janssen^{aff-1}, and John J. Curtin^{aff-1}

^{aff-1}Department of Psychology, University of Wisconsin-Madison

Abstract

Insert here.

Keywords: Substance use disordersPrecision mental healthGeolocationAlgorithmic Fairness

Introduction

About 1 in 10 adults in the United States met diagnostic criteria for alcohol use disorder (AUD) in 2022

([samhsacenterforbehavioralhealthstatisticsandqualityHighlights2022National2022?](#)). Of these individuals, many will experience AUD as a chronic, relapsing disorder marked by periods of recovery interspersed with returns back to harmful use (McKay & Hiller-Sturmhofel, 2011; Moos & Moos, 2006). Lapses, or single instances of goal-inconsistent use, necessarily precede a relapse period ([witkiewitzRelapsePreventionAlcohol2004b?](#)). This temporal precedence, combined with their clear definition and ease of observation, make them a suitable target for early intervention. Yet, even when someone anticipates an oncoming lapse, it may be difficult to pinpoint its specific driving forces. Moreover, the precipitants to a lapse will vary between and within people. These two factors can make maintaining recovery goals difficult and motivate the need for lifelong monitoring and support.

One way to provide ongoing assistance to individuals in recovery is through the development of a continuous risk monitoring and support system. An ideal version of this system consists of two core components. First, the system must be able to collect risk-relevant data with sufficient temporal precision. Next, the system should communicate factors driving (or mitigating) this risk to provide personalized recommendations (e.g., behavior modification or continuation, seeking out supports). Not only does this system need to be developed outright, it also needs to be designed to be both sustainable (i.e., can be used over an extended period of

time) and scalable (i.e., can be effectively used by a maximal amount of people). These goals can be accomplished by integrating personal sensing data and machine learning within such a system.

Personal sensing data are data derived via embedded sensors in technology such as smartphones, smartwatches, or wearables (**mohrPersonalSensingUnderstanding2017a?**), and can also include information collected from applications downloaded onto one of these devices like ecological momentary assessment (EMA) surveys. Other examples of these data are location information, text messages, and social media behavior. Because these devices are already ubiquitous within daily life, these data can viably be collected unobtrusively and continuously for clinical purposes. Many of these data do not require individuals to significantly change their behavior or routines in any way, making data collection both sustainable and ecologically valid. Machine learning models can then uncover relationships between antecedent behaviors present in these data and true lapse events. Importantly, the use of machine learning models enables scalability. Continuous and long-term lapse risk detection and accompanying recommendations cannot realistically be provided by clinicians in an already overburdened addiction treatment system (McLellan et al., 2003). Irrespective of burden, *real time* risk detection and recommendations cannot reasonably be done by clinicians.

Previous work has established the viability of brief daily surveys (also called “ecological momentary assessments” or EMAs) in predicting future alcohol use (**wyantMachineLearningModels2023?**). EMAs typically consist of questions pertaining to lapse-relevant constructs such as self-efficacy, past use, stress, and mood. However, EMAs cannot feasibly capture all contributors to lapse. Reliance on self-report is likely only an effective

tool for individuals who have a good level of insight and therefore are better able to identify these lapse precipitants. Moreover, completing several surveys per day might not be realistic for individuals with multiple jobs or comorbid conditions (e.g., depression). Predictions may be less reliable for individuals who provide either inaccurate or fewer responses.

Crucially, recovery is a dynamic process while EMA, by its very nature, collects data at discrete sampling periods. This may lead to a loss of information, as it assumes that responses remain constant between these periods (e.g., if positive mood is reported at one survey, no shifts in mood are captured until the next survey). Factors that contribute to both maintenance of recovery and lapse events change from person-to-person and from moment-to-moment. A shift in affect may precede a lapse for one individual but not another. Time spent experiencing craving may precede a lapse, but only after a certain threshold is met (i.e., a certain amount of time has been spent craving). In order to best capture this fluidity, the ideal information garnered from personal sensing to be used within a continuous risk monitoring and support system should be able to provide a correspondingly appropriate level of granularity. Geolocation data are one such promising source.

Geolocation Data for Risk Monitoring

Geolocation data consist of latitude and longitude coordinates and can be sampled continuously at regular intervals using applications on smartphones, and therefore have greater temporal sensitivity and specificity compared to EMAs. Furthermore, the collection of these data requires little to no input from the user beyond initial set-up, while EMAs necessitate repeated daily

engagement. In fact, many smartphones and smartwatches automatically collect geolocation data by default. Incorporating these data within a continuous risk monitoring and support system could mean reduced patient burden (i.e., fewer surveys to fill out) and a potentially more equitable system (i.e., easier for people who cannot fill out multiple surveys per day to engage with and, therefore, benefit from). Taken together, there is high potential for geolocation data to be feasibly harnessed both to improve upon past work and to eventually be integrated within a larger monitoring and support system.

Beyond the potential benefits geolocation data offer relative to EMA, the importance of location, such as environmental cues or one's perceived riskiness of a setting, has been shown to play an important role in lapse (Janak & Chaudhri, 2010; Walton et al., 1995; Walton et al., 2003). This link with lapse risk has led to the creation of coping skills that target substance-associated contexts in several treatment strategies like mindfulness-based relapse prevention (LeCocq et al., 2020). These findings underscore not only the potential wealth of information relating to lapse risk that an individual's location can provide, but also demonstrate the utility of incorporating location information into treatment. Furthermore, geolocation data have been identified as being of particular use in both understanding the precipitants to harmful substance use and its effective treatment (Stahler et al., 2013).

Research at the intersection of geolocation and mental health has generated a number of promising features which are derived from basic movement patterns (hereafter called “raw” features). These raw features quantify individual activity patterns as a way to examine behavioral change. For example, geolocation data have been used to estimate loneliness and isolation

through measures of circadian rhythmicity, movement speed, location variance, and clusters of frequently visited locations (Doryab et al., 2019). Many of these same features, in addition to others such as location entropy (the variability of time spent across significant location clusters), amount of time spent at home, and time spent in transit, have also been used to quantify symptoms of depression (Saeb et al., 2015; **saebMobilePhoneSensor2015b?**) and negative symptoms of schizophrenia (Raugh et al., 2020). This research has also linked experiential diversity in movement patterns to positive mood (Heller et al., 2020). These data have not only been applied to the measurement of mood symptoms, but to also predict their emergence (for review, see Shin & Bae, 2023).

In order to implement these features in the context of a monitoring and support system, it is important that these features be intervenable. In other words, the system has to be able to provide a recommendation to the individual using it. From an algorithmic perspective, it is important to balance both prediction and explanatory goals. Certain features detailed in the psychological literature, such as circadian rhythmicity and entropy, have clearer face validity in terms of intervenability as compared to others, like movement speed. For example, changes in the regularity of someone's movement (e.g., circadian rhythmicity) might be indicative of irregular sleep or significant life disruptions (e.g., losing one's job). Encouraging the re-establishment of a regular routine to maintain recovery goals might be a helpful solution, given that disruptions of circadian rhythms appear related to increased alcohol use (Nelson et al., 2024).

A benefit of these raw features is that they require little to no additional information on the part of both researchers and participants in order to identify meaningful movement patterns.

While this makes them easy to implement, it does mean that these features are stripped of important context that exists in our everyday environments. For example, two individuals could exhibit relatively regular movement patterns (like both attending work) while having a vastly different experience en route to those regular locations. For one individual, the route to the office might consist of coffee shops, greenspace, and affluent neighborhoods, whereas for another it might include bars, liquor stores, and higher crime areas.

Context-based features rectify this, in part, by capturing information about the qualities of the surrounding environment. In past work, geolocation data have been used to examine psychosocial stress exposure among substance users by using location data to obtain composite scores of community socioeconomic status and crime (Kwan et al., 2019). In a study of adolescents (both substance-using and not), Mennis and Mason leveraged domain expertise to identify local risky (e.g., pawn shops, corner stores, bars) and protective (e.g., recreational centers, after school programs) locations in a major U.S. city (Mennis & Mason, 2011). Yet, one can imagine further problems which might arise from simply pulling geographic features without individual-specific context. In the Mennis and Mason study, adolescents in this study, counter to researcher's predictions, consistently identified a given recreation center as risky because of the increased violence in that location through a structured qualitative interview [“Ecological Interview”; Mennis & Mason (2011)]. Thus, broad-strokes characterization of locations only based on domain expertise can obscure crucial person-specific differences.

Perhaps a middle ground between these approaches which leverages the benefit of personalized information seen in EMA with the relative low-burdensomeness of geolocation data

is to ask participants to provide researchers with insight into what a given location means to them. Geolocation features could be combined and further enriched with brief, intermittent surveys probing specific information about frequently visited locations (referred to as “contextualized” features). For example, geolocation and EMA data have been used to examine the relationship between location and mood in polydrug users (termed “geographical momentary assessment”; (Epstein et al., 2014)). Geolocation data have also been leveraged to alert individuals when they are approaching a self-identified risky location, such as a previously-frequented bar, within recovery-based smartphone applications (Attwood et al., 2017; Carreiro et al., 2021; Gustafson et al., 2014). Other recovery-based apps have expanded upon this by utilizing geolocation data to create “geo-fences” around areas of past use, like former smoking locations, such that individuals receive real-time notifications as they move through the environment (e.g., a pop-up message on a smartphone which reads “*You are entering a high-risk zone*”; (Naughton et al., 2016)).

Despite promising results suggesting that geolocation data can be capitalized on to improve our understanding of mental health outcomes generally and substance use patterns specifically, research has not been done to understand the *predictive* value of geolocation data in understanding lapses in AUD. Leveraging both raw and contextualized geolocation features, some of the more nuanced facets captured within location can be uncovered, such as associations with others (or lack thereof, e.g., social isolation), associations with previous or anticipated drinking behaviors (e.g., whether or not alcohol is present), and associations with affect (i.e., negative versus positive emotions tied to a given location). Using a combination of these insights

in building a model for use within a continuous risk monitoring and support system will result in our ability to identify a wider variety of potential lapse precipitants and, theoretically, more accurately capture heterogeneous experiences of recovery.

The Current Study

This study used geolocation data and machine learning to predict next-day lapse in individuals with a diagnosis of AUD and a recovery goal of abstinence in order to address several key gaps in the literature. First, we pursued a novel line of research by using geolocation data to predict lapses, expanding on previous risk monitoring work from our lab using EMA (**wyantMachineLearningModels2023?**). Second, we leveraged geolocation features from across the literature that are both raw and contextualized in order to cast a wide net across potential precipitants to lapse. To understand the utility of geolocation data broadly and specific ways of calculating geolocation-based features more specifically, we evaluated several different model configurations and compare them below. We further investigated what specific features were most predictive within our model to address our explanatory goal of uncovering relevant, actionable lapse risk factors which could be used to provide personalized recommendations in future work. This study constitutes an initial evaluation of a model designed to predict lapse back to alcohol use using minimally burdensome data that has the potential to be integrated within a continuous risk monitoring and support system.

Methods

Participants

One hundred and forty six individuals in early recovery (1-8 weeks of abstinence) from AUD were recruited from the Madison area to take part in a three-month study on how mobile health technology can provide recovery support between 2017 and 2019 (R01 AA024391). Recruitment approaches included social media platforms (e.g., Facebook), television and radio advertisements, and clinic referrals. Prospective participants completed a phone screen to assess match with eligibility criteria (Table 1). Participants were excluded if they exhibited severe symptoms of paranoia or psychosis (a score ≤ 2.24 on the SCL-90 psychosis scale or a score ≤ 2.82 on the SCL-90 paranoia scale administered at screening). Participants completed a baseline measure of demographics and other constructs relevant to lapse at the screening visit, which was used for fairness assessments (Table 2).

Procedure

Participants enrolled in a three-month study consisting of five in-person visits, daily surveys, and continuous passive monitoring of geolocation data. Following screening and enrollment visits in which participants consented to participate, learned how to manage location sharing (i.e., turn off location sharing when desired), and reported frequently visited locations, participants completed three follow-up visits one month apart. At each visit, participants were asked questions about frequently visited (>2 times during the course of the previous month) locations (Table 3). Participants were debriefed at the third and final follow-up visit. Participants were expected to

provide continuous geolocation data while on study. Other personal sensing data streams (EMA, cellular communications, sleep quality, and audio check-ins) were collected as part of the parent grant's aims (R01 AA024391).

To enable collection of geolocation data, participants downloaded either the Moves app or the FollowMee app during the intake visit. Moves was bought-out and subsequently deprecated while the study was ongoing (July 2018) and data collection continued using FollowMee until the end of the study. Both apps continuously tracked location via GPS and WiFi positioning technology.

After completion of the study, data were processed to filter out duplicated points, fast movement speeds (>100mph), sudden positional jumps, and periods of long duration suggesting sampling error issues (>24 hours with no movement or >2 hours with a positional jump of more than 0.31 miles or 500 meters). Data points were classified as “in transit” when spacing between individual positions suggested a movement speed of greater than 4mph per NIH health guidelines ([u.s.departmentofhealthandhumanservicesPhysicalActivityGuidelines2018?](#)). Participants were considered to be at a known contextual location if they were within 0.031 miles (50 meters) of a reported frequently visited location.

Data analytic strategy

Data preprocessing, modeling, and Bayesian analyses were done in R using the tidyverse ecosystem (Kuhn & Wickham, 2020). Models were trained using high-throughput computing resources provided by the University of Wisconsin Center for High Throughput Computing

(Center for High Throughput Computing, 2006). In the interest of transparency, all analysis scripts and results can be found publicly on our study website (https://jjcurtin.gitub.io/study_gps).

Outcome variable: Lapses

Alcohol lapses were used as the outcome variable in this study and were used to provide labels for model training, for testing model performance, and for testing issues of algorithmic fairness. Future lapse occurrence (here conceptualized as next-day lapse) was predicted in 24-hour windows which advanced hour-by-hour. *Lapse* and *no lapse* occurrences were identified from the daily survey question, “*Have you drank any alcohol that you have not yet reported?*”. Participants who responded *yes* to this question were then asked to report the date and hour of the start and the end of the drinking episode. In this case, the prediction window was labeled *lapse*. Prediction windows were labeled *no lapse* if no alcohol use was reported within that window.

Input variables: Feature engineering

Feature engineering is the process of creating variables (or “*features*”) from unprocessed data and was used to transform raw geolocation data. Separate feature categories were created for the six contextualized geolocation categories (presented in Table 3) and for six raw geolocation categories: variability in location, time spent outside of the home in the evening, time spent at home, entropy, normalized entropy, and circadian movement .

Raw geolocation features were selected from the literature on account of their ability to be tied to realistic recommendations in a continuous monitoring and support system (see Table 4 for

a detailed description of feature calculations and potential application). Entropy, normalized entropy, and circadian movement were all calculated based on location clusters.

All contextualized features were calculated both as raw (i.e., summed duration over past 6, 12, 24, 48, 72, and 168 hour periods) and change features (i.e., relative to all previous geolocation data) in order to capture individual variation. The calculation windows, as well as whether raw and/or change features were calculated, for raw features are displayed in Table 4.

Finally, we also created demographic features based on intake surveys: two quantitative (age, in years; income, in USD) and five dummy-coded (sex assigned at birth, male versus female; race/ethnicity, non-Hispanic White versus non-White and/or Hispanic; marital status, married versus not married versus other; education, high school or less versus some college versus some degree; employment status, employed versus unemployed).

This resulted in a total of XXX features.

Imputation of missing data and removal of zero-variance features are additional general processing steps that were completed during feature engineering.

Algorithm development & performance

We trained and assessed several configurations of an XGBoost machine learning algorithm. The choice of using an XGBoost algorithm was motivated by two main reasons: 1) the calculation of Shapley values, used to understand the relative contributions of features in predictions, is optimized for XGBoost; and 2) previous work in our lab has made use of XGBoost algorithms in

model development (**wyantMachineLearningModels2023?**) and the ability to eventually consolidate features across models in future work is of high priority.

We created four model types which varied across three feature sets: demographics only (baseline model); demographics and raw geolocation features (raw model); demographics and contextualized geolocation features (contextualized model); and demographics, raw geolocation, and contextualized geolocation features (full model). These model comparisons were designed to assess the value of raw geolocation features over demographics (raw model versus baseline model), contextualized geolocation features over demographics (contextualized model versus baseline model), and contextualized geolocation features over raw geolocation features (contextualized model versus raw model).

Configurations of models varied across a relevant and appropriate range of model-specific hyperparameters (mtry, tree depth, learning rate) as well as resampling techniques (up-sampling of the positive class, lapse, and down-sampling of the negative class, no lapse, ranging from 1:1 to 5:1) to account for the class imbalance in our outcome variable.

Models were trained and assessed using 10×30 participant-grouped, nested k -fold cross-validation. Grouped cross-validation ensures that all data from a given participant are retained as either held-in or held-out, thereby preventing the introduction of bias from a participant's data being used to predict their own data. Nested cross-validation uses two nested loops for dividing and holding out folds: an outer loop, where held-out folds serve as test sets for model evaluation; and inner loops, where held-out folds serve as validation sets for model selection (Jonathan et al., 2000).

The primary performance metric for model selection and evaluation of the validation sets was area under the Receiver Operating Characteristic (auROC) curve (Kuhn & Johnson, 2018). auROC indexes the probability that the model will predict a higher score for a randomly selected positive case (lapse) relative to a randomly selected negative case (no lapse). The overall percentage of lapses across all observations (each day per participant on study) was 7.9%, motivating the selection of this metric as it is unaffected by class imbalance.

Shapley values were computed in log-odd units in order to evaluate global importance of each feature category. Shapley values measure the unique contribution of features in an algorithm's predictions and therefore identify the relative importance of different features (Lundberg & Lee, 2017). Global feature importance for each broad feature category was calculated by averaging the absolute values of Shapley values across all observations per feature category. Highly important features represent relevant, actionable potential antecedents to lapse (and therefore points of intervention) that will be relevant in the future development of a fully integrated continuous risk monitoring and support system. However, these are descriptive analyses because standard errors or other indices of uncertainty for important scores are not available for Shapley values. Of note, even low global features may be an important contributor to lapse, and therefore motivate intervention, for a specific person at a specific time (i.e., high local importance).

Bayesian hierarchical generalized linear models were used to estimate the posterior probability distributions of auROCs. Median posterior probability for auROC and Bayesian credible intervals (CIs) are reported as an evaluation of each model's overall performance. We

used a threshold of auROC = .5 (chance performance) when examining CIs, such that a CI that does not contain .5 indicates performance above chance and therefore predictive signal in the data.

Finally, Bayesian model comparisons were used to evaluate the four contrasts defined above based on feature sets. The probability that model performance systematically difference was determined for every model comparison. Posterior probabilities for auROC differences and the 95% Bayesian CI are reported.

Algorithmic fairness

Classes for fairness analyses were defined on the basis of personal individual characteristics divided such that subgroups reflected coarse dichotomies of groups which experience relatively increased and decreased societal privilege. This resulted in four broad classes: white versus non-white, younger than 55 versus equal to or older than 55, above or below the federal poverty line (Bartels, 2024), and sex at birth (male versus female). A Bayesian hierarchical generalized linear model was used to estimate the posterior probability distributions of auROCs and corresponding 95% Bayesian CIs across these eight subgroups. Finally, Bayesian group comparisons were used in order to identify the likelihood of differential performance of our model between subgroups within each class.

References

- Attwood, S., Parke, H., Larsen, J., & Morton, K. L. (2017). Using a mobile health application to reduce alcohol consumption: A mixed-methods evaluation of the drinkaware track & calculate units application. *BMC Public Health*, 17(1), 394. <https://doi.org/10.1186/s12889-017-4358-9>
- Bartels, T. (2024). 2024 Federal Poverty Rates Published: Why that matters for your student loans. In *VIN Foundation*. <https://vinfoundation.org/2024-federal-poverty-rates-published-why-that-matters-for-your-student-loans/>.
- Carreiro, S., Taylor, M., Shrestha, S., Reinhardt, M., Gilbertson, N., & Indic, P. (2021). Realize, Analyze, Engage (RAE): A Digital Tool to Support Recovery from Substance Use Disorder. *Journal of Psychiatry and Brain Science*, 6, e210002. <https://doi.org/10.20900/jpbs.20210002>
- Center for High Throughput Computing. (2006). *Center for high throughput computing*. Center for High Throughput Computing. <https://doi.org/10.21231/GNT1-HW21>
- Doryab, A., Villalba, D. K., Chikersal, P., Dutcher, J. M., Tumminia, M., Liu, X., Cohen, S., Creswell, K., Mankoff, J., Creswell, J. D., & Dey, A. K. (2019). Identifying Behavioral Phenotypes of Loneliness and Social Isolation with Passive Sensing: Statistical Analysis, Data Mining and Machine Learning of Smartphone and Fitbit Data. *JMIR mHealth and uHealth*, 7(7), e13209. <https://doi.org/10.2196/13209>
- Epstein, D. H., Tyburski, M., Craig, I. M., Phillips, K. A., Jobes, M. L., Vahabzadeh, M., Mezghanni, M., Lin, J.-L., Furr-Holden, C. D. M., & Preston, K. L. (2014). Real-time tracking of neighborhood surroundings and mood in urban drug misusers: Application of a new method to study behavior in its geographical context. *Drug and Alcohol Dependence*, 134, 22–29. <https://doi.org/10.1016/j.drugalcdep.2013.09.007>

- Gustafson, D. H., McTavish, F. M., Chih, M.-Y., Atwood, A. K., Johnson, R. A., Boyle, M. G., Levy, M. S., Driscoll, H., Chisholm, S. M., Dillenburg, L., Isham, A., & Shah, D. (2014). A Smartphone Application to Support Recovery From Alcoholism: A Randomized Clinical Trial. *JAMA Psychiatry*, 71(5), 566. <https://doi.org/10.1001/jamapsychiatry.2013.4642>
- Heller, A. S., Shi, T. C., Ezie, C. E. C., Reneau, T. R., Baez, L. M., Gibbons, C. J., & Hartley, C. A. (2020). Association between real-world experiential diversity and positive affect relates to hippocampal-striatal functional connectivity. *Nature Neuroscience*, 23(7), 800–804. <https://doi.org/10.1038/s41593-020-0636-4>
- Janak, P. H., & Chaudhri, N. (2010). The Potent Effect of Environmental Context on Relapse to Alcohol-Seeking After Extinction. *The Open Addiction Journal*, 3, 76–87. <https://doi.org/10.2174/1874941001003010076>
- Jonathan, P., Krzanowski, W. J., & McCarthy, W. V. (2000). On the use of cross-validation to assess performance in multivariate prediction. *Statistics and Computing*, 10(3), 209–229. <https://doi.org/10.1023/A:1008987426876>
- Kuhn, M., & Johnson, K. (2018). *Applied Predictive Modeling* (1st ed. 2013, Corr. 2nd printing 2018 edition). Springer. <https://doi.org/10.1007/978-1-4614-6849-3>
- Kuhn, M., & Wickham, H. (2020). *Tidymodels: A collection of packages for modeling and machine learning using tidyverse principles*.
- Kwan, M.-P., Wang, J., Tyburski, M., Epstein, D. H., Kowalczyk, W. J., & Preston, K. L. (2019). Uncertainties in the geographic context of health behaviors: A study of substance users' exposure to psychosocial stress using GPS data. *International Journal of Geographical Information Science*, 33(6), 1176–1195. <https://doi.org/10.1080/13658816.2018.1503276>

- LeCocq, M. R., Randall, P. A., Besheer, J., & Chaudhri, N. (2020). Considering Drug-Associated Contexts in Substance Use Disorders and Treatment Development. *Neurotherapeutics: The Journal of the American Society for Experimental NeuroTherapeutics*, 17(1), 43–54. <https://doi.org/10.1007/s13311-019-00824-2>
- Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 4768–4777.
- McKay, J. R., & Hiller-Sturmhofel, S. (2011). Treating alcoholism as a chronic disease: Approaches to long-term continuing care. *Alcohol Research & Health: The Journal of the National Institute on Alcohol Abuse and Alcoholism*, 33(4), 356–370.
- McLellan, A. T., Carise, D., & Kleber, H. D. (2003). Can the national addiction treatment infrastructure support the public's demand for quality care? *Journal of Substance Abuse Treatment*, 25(2), 117–121.
- Mennis, J., & Mason, M. J. (2011). People, Places, and Adolescent Substance Use: Integrating Activity Space and Social Network Data for Analyzing Health Behavior. *Annals of the Association of American Geographers*, 101(2), 272–291. <https://doi.org/10.1080/00045608.2010.534712>
- Moos, R. H., & Moos, B. S. (2006). Rates and predictors of relapse after natural and treated remission from alcohol use disorders. *Addiction (Abingdon, England)*, 101(2), 212–222. <https://doi.org/10.1111/j.1360-0443.2006.01310.x>
- Naughton, F., Hopewell, S., Lathia, N., Schalbroeck, R., Brown, C., Mascolo, C., McEwen, A., & Sutton, S. (2016). A Context-Sensing Mobile Phone App (Q Sense) for Smoking Cessation: A

Mixed-Methods Study. *JMIR mHealth and uHealth*, 4(3), e106. <https://doi.org/10.2196/mhealth.5787>

Nelson, M. J., Soliman, P. S., Rhew, R., Cassidy, R. N., & Haass-Koffler, C. L. (2024).

Disruption of circadian rhythms promotes alcohol use: A systematic review. *Alcohol and Alcoholism*, 59(2), agad083. <https://doi.org/10.1093/alcalc/agad083>

Raugh, I. M., James, S. H., Gonzalez, C. M., Chapman, H. C., Cohen, A. S., Kirkpatrick, B., &

Strauss, G. P. (2020). Geolocation as a Digital Phenotyping Measure of Negative Symptoms and

Functional Outcome. *Schizophrenia Bulletin*, 46(6), 1596–1607. <https://doi.org/10.1093/schbul/sbaa121>

Saeb, S., Zhang, M., Kwasny, M., Karr, C. J., Kording, K., & Mohr, D. C. (2015). The

relationship between clinical, momentary, and sensor-based assessment of depression. *2015 9th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth)*, 229–232. <https://doi.org/10.4108/icst.pervasivehealth.2015.259034>

Shin, J., & Bae, S. M. (2023). A Systematic Review of Location Data for Depression Prediction.

International Journal of Environmental Research and Public Health, 20(11), 5984. <https://doi.org/10.3390/ijerph20115984>

Stahler, G. J., Mennis, J., & Baron, D. A. (2013). Geospatial technology and the "exposome":

New perspectives on addiction. *American Journal of Public Health*, 103(8), 1354–1356. <https://doi.org/10.2105/AJPH.2013.301306>

Walton, M. A., Blow, F. C., Bingham, C. R., & Chermack, S. T. (2003). Individual and social/environmental predictors of alcohol and drug use 2 years following substance abuse treatment.

Addictive Behaviors, 28(4), 627–642. [https://doi.org/10.1016/s0306-4603\(01\)00284-2](https://doi.org/10.1016/s0306-4603(01)00284-2)

Walton, M. A., Reischl, T. M., & Ramanthan, C. S. (1995). Social settings and addiction relapse. *Journal of Substance Abuse*, 7(2), 223–233. [https://doi.org/10.1016/0899-3289\(95\)90006-3](https://doi.org/10.1016/0899-3289(95)90006-3)

Tables and Figures

Table 1: Eligibility Criteria

Table 1: Eligibility criteria for study enrollment. **Personal or study-provided.

Eligibility Criteria
>= 18 years of age
Ability to read and write in English
Diagnosis of moderate AUD (>= 4 self-reported DSM-5 symptoms)
Abstinent from alcohol for 1-8 weeks
Willing to use only one smartphone** while on study

Table 2: Collected demographic information

Table 2: Demographic and relevant alcohol use history variables sampled at screening visit.

Variable	Measure
Demographics	Age
	Sex
	Race
	Ethnicity
	Employment
	Income
	Marital Status
Alcohol	Alcohol Use History
	DSM-5 Checklist for AUD
	Young Adult Alcohol Problems Test
	WHO-The Alcohol, Smoking and Substance Involvement
	Screening Test

Table 3: Contextual geolocation information

Table 3: Location information collected from frequently visited locations.

Question	Responses
Address	
Type of place	Work, School, Volunteer, healthcare, Home of a friend, Home of a family member, Liquor store, Errands (e.g., grocery store, post office), Coffee shop or cafe, Restaurant, Park, Bar, Gym or fitness center, AA or recovery meeting, Religious location (e.g., church, mosque, temple), Other
Have you drank alcohol here before?	No, Yes
Is alcohol available here?	No, Yes
How would you describe your experiences here?	Pleasant, Unpleasant, Mixed, Neutral
Does being at this location put you at any risk to begin drinking?	No risk, Low risk, Medium risk, High risk
Did the participant identify this place as a risky location they are trying to avoid now that they are sober?	No, Yes

Table 4: Raw geolocation features

Table 4: Raw geolocation features derived from the GPS literature with description of modifications made for current study.

Feature	Original formula	Modification for current study	Interpretation	Theoretical mechanism of action
Location Variance	$\text{LocationVariance} = \log(\sigma_{lat}^2 + \sigma_{long}^2)$	Calculated as both raw and difference scores for past 6, 12, 24, 48, 72, and 168 hours	Spatial variability over defined segments	Lower variance may be indicative of more isolation, less geographical coverage
Entropy	$\text{Entropy} = -\sum_{i=1}^N p_i \log(p_i)$	Calculated as both raw and difference scores for past 24, 48, 72, and 168 hours (PLACEHOLDER but calc will probably be similar)	Distribution of time across location clusters	Higher entropy may be indicative of greater flexibility in routine, which could either indicate experiential diversity or lack of routine
Normalized Entropy	$\frac{\text{Entropy}}{\log(N)}$	Calculated as both raw and difference scores for past 24, 48, 72, and 168 hours (PLACEHOLDER)	Distribution of time across location clusters, accounting for number of clusters	
Circadian Movement	Energy GPS spectrum first	Two-week lead time followed by one week	Measure of how closely	Higher CM may indicate greater temporal regularity
Time	Calculated of Percentage of periods which advance by	Calculated as both raw and	an	