Using Cellular Communication Sensing to Support Early Recovery from Alcohol Use Disorder

Kendra Wyant Coco Yu John J. Curtin

2025-10-16

Source: Article Notebook

Introduction

Alcohol Use Disorder (AUD) is a chronic, relapsing disease (Dennis & Scott, 2007; McLellan et al., 2000; Rounsaville, 2010). Lapses, single episodes of alcohol use, and relapse, a full return to harmful drinking, can occur at any point in recovery (Kirshenbaum et al., 2009; Nguyen et al., 2020; Scott et al., 2005; Witkiewitz, 2011). As with other chronic health conditions where symptoms fluctuate, sometimes unexpectedly, sustained AUD recovery requires ongoing monitoring of lapse risk.

Machine learning–guided recovery systems may now assist with the inherently difficult task of identifying when and why someone is at increased risk. Personal sensing of densely sampled data from individuals' day-to-day lives can provide the inputs necessary for temporally dynamic lapse predictions (Mohr et al., 2017). Early models using ecological momentary assessment data have achieved excellent accuracy (Chih et al., 2014; Wyant et al., 2024, under review). Still, questions remain about the long-term feasibility of self-report sensing methods and whether new, important risk factors might emerge from sensing methods that passively collect smartphone data without user input.

Cellular communication sensing may be one promising method. It offers the potential for greater temporal specificity in capturing fluctuations in risk compared with self-report data. Collecting communication data in near real time could allow an algorithm to detect potential triggers as they occur, without prompting users to reflect on their feelings or waiting for users to report about their environement at a later point. For example, late night phone calls could indicate an emergency, "drunk dialing", or other risk-relevant interactions, while an expanding or shrinking social circle could be characterized by the number of unique contacts someone has communicated with.

These data may become even more powerful when communication contacts are contextualized with personal meaning for a given participant (e.g., Who is this contact to them? How pleasant or unpleasant is a typical interaction with them? Have they drank with them in the past?). In this scenario, contextualized communication logs might reveal that the late-night phone call was to a sponsor, or that the shrinking social circle was due to reduced contact with people who are unsupportive of their recovery.

In this study, we evaluated the performance of a machine learning model that predicts the probability of a next-day lapse using contextualized cellular communication data. We also describe the most important features contributing to these predictions, with the goal of identifying new, clinically meaningful features emerging from communication-based sensing.

Methods

Participants and Procedure

We recruited adults in early recovery from AUD in Madison, Wisconsin, through print and digital advertisements and partnerships with treatment centers. Eligibility criteria required that participants were age 18 or older, able to read and write in English, had moderate to severe AUD [1], had been abstinent from alcohol for 1–8 weeks, were willing to use a single smartphone, and were not exhibiting severe psychosis or paranoia.[2]

Participants completed up to 5 study visits over approximately 3 months: a screening visit, intake visit, and 3 monthly follow-up visits. At screening we collected demographic information (age, sex at birth, race, ethnicity, education, marital status, employment, and income) and clinical characteristics (DSM-5 AUD symptom count, alcohol problems (Hurlbut & Sher, 1992), and presence of psychological symptoms (Derogatis, L.R., 2000)). At intake we collected additional self-report data on abstinence self-efficacy (McKiernan et al., 2011), craving (Flannery et al., 1999), and recent recovery efforts. At each monthly follow-up, we downloaded cellular communication metadata (voice calls and SMS text message logs) from participants' smartphones. We identified important contacts (i.e., individuals they had communicated with at least twice by call or text in the past month) and asked 7 contextual questions about these contacts.

While enrolled, participants completed 4 brief daily ecological momentary assessments (7-10 questions). The first item assessed alcohol use (date and time of any unreported drinking episodes). Lapse reports were verified at follow-up visits using a timeline follow-back interview.

[1] (4 self-reported DSM-5 symptoms)

[2] Defined as scores >2.2 or 2.8, respectively, on the psychosis or paranoia scales of the Symptom Checklist–90 (Derogatis, L.R., 2000).

Additional sensing data streams and self-report measures were collected for the parent grant. The full study protocol is available on our Open Science Framework page (https://osf.io/wgpz9/).

We screened 192 participants. Of these, 169 enrolled and 154 completed the first follow-up. Data from 10 participants were excluded due to loss of abstinence goals, careless responding, or unusually low compliance. The final analytic sample included 144 participants.

Data Analysis Plan

Our models predicted the probability of an alcohol lapse within a 24-hour window. Predictions were generated daily at 4 a.m., beginning on participants' second study day and continuing for up to 3 months. In total, there were 11,507 labeled prediction windows across all participants.

Features were engineered from all available data up to the start of each window.[3] The full model included 406 features from cellular communication data plus 24 features from baseline self-report measures. We also evaluated a comparison model that used only the baseline features. Table 1 details the raw predictors and feature engineering procedures.

Candidate model configurations differed by algorithm (elastic net, random forest, XGBoost), outcome resampling method, and hyperparameter values. The best configuration for each model was selected using 6 repeats of participant-grouped 5-fold cross-validation. Our performance metric was area under the receiver operating curve (auROC). Folds were stratified by a between-subject measure of our outcome (low lapsers: 0-9 lapses; high lapsers: 10+ lapses).

We evaluated model performance with a Bayesian hierarchical generalized linear model. Posterior distributions with 95% credible intervals (CI) were estimated from the 30 held-out test sets using weakly informative, data-dependent priors to regularize and reduce overfitting.[4] Random intercepts were included for repeat and fold (nested within repeat). auROCs were logit-transformed and regressed on model type to estimate the probability that model performances differed systematically.

Our best performing models used an elastic net algorithm. We quantified feature importance by examining the retained features (i.e., coefficient value > 0) in the full model and ordering them by absolute coefficient value. These values provide an estimate of the direction and magnitude of association between each predictor and the outcome, conditional on the other features retained. All our annotated analysis scripts are publicly available on our study website (https://jjcurtin.github.io/study_messages/).

[3]We filtered the data to include only communications with known context prior to feature engineering.

[4]Residual SD ~ normal(0, exp(2)); intercept (centered predictors) ~ normal(2.3, 1.3); window-width contrasts ~ normal(0, 2.69); covariance ~ decov(1,1,1,1).

Raw Predictor	Response Options
Originated	Incoming, outgoing
Call duration	Duration (in minutes)
Call answered	Yes, no
Date/time of communication	Date and time
Phone number	Phone number
Type of Relationship	Family, friend, counselor or so
Have you drank alcohol with this person?	Never/almost never, occasion
What is their drinking status?	Drinker, non-drinker, don't kn
Would you expect them to drink in your presence?	Yes, no, uncertain
Are they currently in recovery from drugs or alcohol?	Yes, no, don't know
Are they supportive about your recovery goals?	Supportive, unsupportive, mi
How are your typical experiences with this person?	Pleasant, unpleasant, mixed,
DSM-5 symptom count	Numeric (4-11)
Past year alcohol problems	Numeric (0-27)
Craving	Numeric (0-30)
Abstinence self-efficacy: Negative affect, social, physical, and craving subscales	Numeric (0-20)
Number of individual alcohol counseling sessions attended (past 30 days)	Numeric
Number of group alcohol counseling sessions attended (past 30 days)	Numeric
Number of self-help group meetings attended (past 30 days)	Numeric
Number of other mental health counseling sessions attended (past 30 days)	Numeric
Number of days in contact with supportive people (past 30 days)	Numeric
Number of days in contact with unsupportive people (past 30 days)	Numeric
Taken prescribed medication for alcohol use disorder (past 30 days)	Yes, no
Taken prescribed medication for other mental health disorder (past 30 days)	Yes, no
Satisfaction with progress toward recovery goals (past 30 days)	Numeric (0-4)
Confidence in abstinence ability (next 30 days)	Numeric (0-4)
Has a goal of abstinence	Yes, no, uncertain
Age	Numeric (years)
Sex at birth	Male, female
Race	Non-Hispanic White, non-Wh
Education	High school or less, some coll
Income	Numeric (dollars)
Marital Status	Married, not married, other

Source: Make All Tables for Main Manuscript

Ethical Considerations

All procedures were approved by the University of Wisconsin-Madison Institutional Review Board (Study #2015-0780). All participants provided written informed consent.

Results

Participants

Table 2 provides the demographic characterization of our sample. We obtained a total of 375,912 contextualized communications across participants. Participants had, on average, 2,610 communications (range = 109-14,225). 56% of participants reported at least one lapse.

	N	%	M	SD	Range
Age			40.4	11.8	21-72
Sex at Birth					
Female	74	51.4			
Male	70	48.6			
Race					
American Indian/Alaska Native	3	2.1			
Asian	2	1.4			
Black/African American	8	5.6			
White/Caucasian	125	86.8			
Other/Multiracial	6	4.2			
Hispanic, Latino, or Spanish origin					
Yes	3	2.1			
No	141	97.9			
Education					
Less than high school or GED degree	1	0.7			
High school or GED	14	9.7			
Some college	39	27.1			
2-Year degree	13	9.0			
College degree	55	38.2			
Advanced degree	22	15.3			
Employment					
Employed full-time	70	48.6			

	N	%	M	SD	Range
Employed part-time	25	17.4			
Full-time student	7	4.9			
Homemaker	1	0.7			
Disabled	7	4.9			
Retired	8	5.6			
Unemployed	15	10.4			
Temporarily laid off, sick leave, or maternity leave	3	2.1			
Other, not otherwise specified	8	5.6			
Personal Income			$35,050	$32,069	$0-200,000
Marital Status					
Never married	63	43.8			
Married	32	22.2			
Divorced	42	29.2			
Separated	5	3.5			
Widowed	2	1.4			

Table 2: Demographics

Note:
N = 144

Source: Make All Tables for Main Manuscript

Model Evaluation

The median posterior auROC for the full model was 0.68, with relatively narrow 95% CI ([0.64, 0.71]) that did not contain .5. This provides strong evidence that the model is capturing signal in the data. The final model retained 13 features (Figure 1). The top four were baseline measures of abstinence confidence, having a goal of abstinence, abstinence self-efficacy when experiencing negative affect, and craving. Communication frequency with people unaware of the individual's recovery goals also emerged as an important feature associated with increased lapse risk.

We evaluated a comparison model to assess the incremental predictive value of cellular communication features beyond baseline measures. The baseline model retained 5 features and achieved performance nearly identical to the full model (median auROC = 0.68, 95% CI [0.64, 0.71]). The median difference in auROC between the full and baseline models was less than .01, providing no evidence (52% probability) that their posterior distributions were meaningfully different.

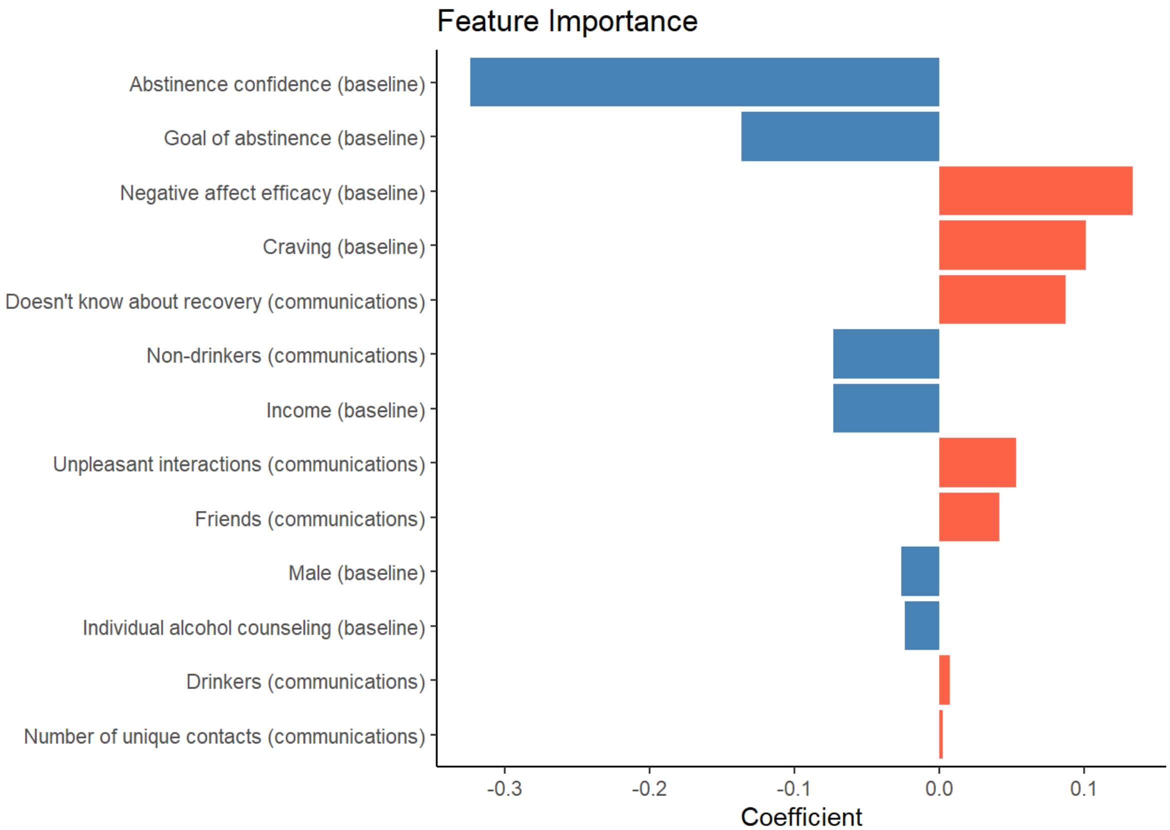

Figure 1: Global feature importance (elastic net coefficient) for the full model. Features are ordered by absolute coefficient value. Blue bars indicate higher feature values, on average, lower lapse risk. Red bars indicate higher feature values, on average, increase risk. Baseline features were collected from self-report measures at the start of the study. Communication features were engineered from the contexualized cellular communications.

Source: Make All Figures for Main Manuscript

Discussion

Our machine learning model incorporating cellular communications achieved fair performance, with an auROC of 0.68, indicating that some signal was present in the data. However, we found no incremental predictive value compared to a baseline model that included only demographic variables and self-report measures. In the full model, which combined cellular communication and baseline measures, the four most important predictors were all self-report variables: abstinence confidence, abstinence goal, negative affect efficacy, and craving.

This finding suggests that cellular communication data may not add substantial information for predicting lapse risk beyond what can already be captured by brief self-report and demographic measures. Nonetheless, it is notable that several communication features were retained in the final model with moderately sized coefficients. These features, including communications with people unaware of the participant's recovery status, with non-drinkers, with friends, and with individuals who were unpleasant to interact with, may still offer some insight into lapse vulnerability.

In contrast, raw counts of calls and text messages and call durations were not retained in the final model. This implies that the quantity of communications alone may be less informative for lapse prediction than the quality and social meaning of the interactions. Future research may benefit from collecting richer contextual information about communication contacts to better understand the social dynamics contributing to lapse risk. Even with highly contextualized communication data, however, prediction may be constrained by data sparsity. Many participants had relatively few communications per day, and some had extended periods with no recorded interactions at all. Such sparsity limits the capacity of these data to capture short-term fluctuations in lapse risk.

Our study design may have further contributed to this limitation. We collected only phone and SMS text communications through the native smartphone app. Yet, in recent years, many individuals have shifted their primary communication to private messaging apps (e.g., WhatsApp, Signal) or social media platforms (e.g., Facebook Messenger, Instagram) (**mcdowellPreferencesAttitudesDigital2025?**). As a result, our dataset likely did not capture the full range of participants' social interactions. Future studies could examine whether communication data from these platforms yield stronger predictive signal.

Overall, our findings suggest that models using passive cellular sensing captured limited incremental signal for lapse prediction beyond demographic and self-report measures. While we do not rule out the potential value of cellular communication data in certain contexts, our results underscore the need to explore other passive sensing methods, such as geolocation, which may provide denser data that captures different aspects of daily life relevant to lapse risk.

Chih, M.-Y., Patton, T., McTavish, F. M., Isham, A. J., Judkins-Fisher, C. L., Atwood, A. K., & Gustafson, D. H. (2014). Predictive modeling of addiction lapses in a mobile health application. *Journal of Substance Abuse Treatment*, *46*(1), 29–35. https://doi.org/10.1016/j.jsat.2013.08.004

Dennis, M., & Scott, C. K. (2007). Managing Addiction as a Chronic Condition. *Addiction Science & Clinical Practice*, *4*(1), 45–55.

Derogatis, L.R. (2000). *Brief Symptom Inventory 18 - Administration, scoring, and procedures manual.* NCS Pearson.

Flannery, B. A., Volpicelli, J. R., & Pettinati, H. M. (1999). Psychometric properties of the Penn Alcohol Craving Scale. *Alcoholism, Clinical and Experimental Research*, *23*(8), 1289–1295.

Hurlbut, S. C., & Sher, K. J. (1992). Assessing alcohol problems in college students. *Journal of American College Health.*, *41*(2), 49–58.

Kirshenbaum, A. P., Olsen, D. M., & Bickel, W. K. (2009). A quantitative review of the ubiquitous relapse curve. *Journal of Substance Abuse Treatment*, *36*(1), 8–17. https://doi.org/10.1016/j.jsat.2008.04.001

McKiernan, P., Cloud, R., Patterson, D. A., Wolf, S., Golder, S., & Besel, K. (2011). Development of a Brief Abstinence Self-Efficacy Measure. *Journal of Social Work Practice in the Addictions*, *11*(3), 245–253. https://doi.org/10.1080/1533256X.2011.593445

McLellan, A. T., Lewis, D. C., O'Brien, C. P., & Kleber, H. D. (2000). Drug dependence, a chronic medical illness: Implications for treatment, insurance, and outcomes evaluation. *JAMA*, *284*(13), 1689–1695. https://doi.org/10.1001/jama.284.13.1689

Mohr, D. C., Zhang, M., & Schueller, S. M. (2017). Personal Sensing: Understanding Mental Health Using Ubiquitous Sensors and Machine Learning. *Annual Review of Clinical Psychology*, *13*(1), 23–47. https://doi.org/10.1146/annurev-clinpsy-032816-044949

Nguyen, L.-C., Durazzo, T. C., Dwyer, C. L., Rauch, A. A., Humphreys, K., Williams, L. M., & Padula, C. B. (2020). Predicting Relapse After Alcohol Use Disorder Treatment in a High-Risk Cohort: The Roles of Anhedonia and Smoking. *Journal of Psychiatric Research*, *126*, 1–7. https://doi.org/10.1016/j.jpsychires.2020.04.003

Rounsaville, D. B. (2010). *Lapse, Relapse, and Chasing the Wagon: Post-Treatment Drinking and Recovery* [PhD thesis]. University of Maryland, Baltimore County.

Scott, C. K., Foss, M. A., & Dennis, M. L. (2005). Pathways in the relapse–treatment–recovery cycle over 3 years. *Journal of Substance Abuse Treatment*, *28 Suppl 1*, S63–72. https://doi.org/10.1016/j.jsat.2004.09.006

Witkiewitz, K. (2011). Predictors of heavy drinking during and following treatment. *Psychology of Addictive Behaviors*, *25*(3), 426–438. https://doi.org/10.1037/a0022889

Wyant, K., Fronk, G. E., Yu, C., Punturieri, C. E., & Curtin, J. J. (under review). *Forecasting Risk of Alcohol Lapse up to Two Weeks in Advance using Time-lagged Machine Learning Models.*

Wyant, K., Sant'Ana, S. J., Fronk, G. E., & Curtin, J. J. (2024). Machine learning models for temporally precise lapse prediction in alcohol use disorder. *Journal of Psychopathology and Clinical Science*, *133*(7), 527–540. https://doi.org/10.1037/abn0000901