Evaluating Cellular Communication Sensing for Lapse Risk Prediction During Early Recovery from Alcohol Use Disorder

Kendra Wyant   Coco Yu   John J. Curtin

2026-01-16

Background: Alcohol Use Disorder (AUD) is a chronic, relapsing disease. An automated recovery support system using personal sensing and machine learning may help identify when individuals are at elevated lapse risk. Cellular communication sensing may detect dynamic changes in lapse risk and can be contextualized with self-reported, risk-relevant information about contacts. Objective: We evaluated a machine learning model predicting next-day alcohol lapse among individuals in early recovery from AUD using contextualized cellular communication data and baseline demographic and AUD characteristics. Methods: A total of 144 participants (49% male; mean age=40; 87% non-Hispanic White) with a goal of abstinence provided cellular communication data and alcohol use reports via a 4x daily EMA for up to three months. Models were trained and evaluated using repeated k-fold cross-validation. Results: The best-performing model used an elastic net algorithm and retained 13 features (median posterior auROC = 0.68, 95% Bayesian credible interval (CI; [0.64, 0.71]). A baseline comparison model including only demographic and self-report measures retained five features and demonstrated nearly identical performance (median auROC = 0.68, 95% CI [0.64, 0.71]). Conclusions: Cellular communication data capture some risk-relevant signal for alcohol lapse but do not provide incremental predictive value beyond baseline measures. Several communication features were retained in the final model with moderately sized coefficients, suggesting that aspects of social communication may be important for understanding lapse risk. Although, limitations inherent to cellular communication as a sensing method may outweigh their added value.

Source: Article Notebook

1

Introduction

Alcohol Use Disorder (AUD) is a chronic, relapsing disease[1–3]. Lapses, single episodes of alcohol use, are among the strongest predictors (and a necessary precursor) for relapse, a full return to harmful drinking[4,5]. While lapses can occur at any point in recovery, they are particularly risky during early recovery[6]. Protective coping mechanisms and socio-environmental resources that support recovery are dynamic and accumulate over time[7]. As a result, early recovery represents a critical window of vulnerability during which a lapse is more likely to escalate into relapse.

An automated recovery support system powered by personal sensing and machine learning may assist with the inherently difficult task of identifying when and why someone is at increased risk for lapse. Personal sensing of densely sampled data from individuals' day-to-day lives can provide the inputs necessary for temporally dynamic lapse predictions[8]. Early machine learning models using ecological momentary assessment (EMA) data have achieved excellent accuracy in predicting future lapses back to alcohol use in treatment seeking populations[9–11].

Despite the high predictive success of EMA, questions remain about the long-term feasibility of a self-report sensing method. EMA has been shown to be well-tolerated among substance use populations over relatively short periods of time[12,13]. It is unclear whether individuals would be willing and able to adhere to an extensive EMA protocol (e.g., 4 prompts per day) indefinitely. Moreover, EMA items are chosen using domain expertise from decades of research on the self-report factors that predict lapse. It is possible, however, that there are several alternative precipitators of lapse not yet discovered due to small subtle changes in one's environment, social circle, or lifestyle that cannot be easily pinpointed in self-report.

Cellular communication sensing may be a promising alternative to EMA. Whereas EMA is limited to, at most, several assessments per day, communication sensing is mostly passive and can be monitored moment-by-moment. Cellular communication patterns also appear to capture clear, risk-relevant constructs. Late-night phone calls could indicate an emergency, "drunk dialing," or interpersonal conflict. A decrease in the number of contacts an individual communicates with could reflect a shrinking social circle, isolation, or disengagement.

These data may become even more powerful when communication patterns are contextualized with participant-specific meaning. For instance, knowing a participant's relationship to their contacts, whether they have previously drank alcohol with a given contact, or whether that contact supports their recovery goals could substantially alter interpretation. In the examples above, contextualized communication data might reveal that late-night calls are made to a sponsor, or that a shrinking social circle reflects reduced contact with individuals unsupportive of their recovery. In this way, the same communication patterns may reflect protective processes rather than increased lapse risk.

In this study, we assessed whether contextualized cellular communication features contain clinically meaningful signals for predicting next-day alcohol lapse risk among individuals in early recovery from AUD. Using a machine learning model, we evaluated the predictive utility

of these features and identified the most important communication features, with the goal of uncovering new, clinically meaningful predictors of lapse risk.

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 a goal of abstinence from alcohol, had been abstinent 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[15], and presence of psychological symptoms[14]). At intake we collected additional self-report data on abstinence self-efficacy[16], craving[17], and recent recovery efforts.

At each monthly follow-up, we downloaded backups of participants' cellular communication metadata directly from their smartphones. Metadata included the phone number of the other party, the date and time of the communication, the origin of call or message (i.e., incoming or outgoing), whether the call was answered (voice calls only), and the duration of the call (voice calls only). During each follow-up visit, study staff identified important contacts. Contacts that participants communicated with at least twice by call or text in the past month were deemed important. For each important contact, participants answered 7 contextual questions about their type of relationship, whether they ever drank alcohol with this person, the drinking status of the contact, expectations about whether the contact would drink in their presence, recovery status of contact, level of supportiveness of contact, and affective experiences with the contact.

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. 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/).

---

[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[14].

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. Participants reported the date and hour of the start and end time of any alcohol use on the first item of the EMA. Prediction windows were labeled as lapse if any alcohol use was reported in the 24-hour window. In total, there were 11,507 labeled prediction windows across all participants. Positive lapse labels were underrepresented (7.5%; 861/11,507).

We filtered the data to include only communications with known context (i.e., people with whom they communicated with at least twice in a month and whom they provided self-report context about). Cellular communication features were engineered from all available data up to the start of each window. We used six scoring epochs (6, 12, 24, 48, 72, and 168 hours before the start of the prediction window) to create features. Within each scoring epoch we calculated two types of features: raw and difference features. Raw features represent the raw feature value calculated within a scoring epoch (e.g., the rate count of text messages during the 48 hours immediately preceding the start of the prediction window). Difference features capture participant-level changes from their baseline scores. Specifically, we subtracted each participant's mean score for each feature (using all available data prior to the prediction window) from the associated raw feature (e.g., the participant's average rate count of text messages across all time on study subtracted from the rate count in the preceding 48 hours).

The full model included 406 features from cellular communication data plus 24 numeric or dummy-coded features from baseline self-report measures. We also evaluated a comparison model that used only the baseline features. Table 1 details the raw predictors, feature engineering procedures, and features included in the full vs. baseline models. Other feature engineering steps performed during cross-validation included imputing missing values (median imputation for numeric features and mode imputation for nominal features) and removing zero and near-zero variance features as determined from held-in data.

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 so that all folds contained comparable proportions of individuals who lapsed frequently (i.e., 10+ times).

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.[3] Random intercepts were included for repeat and fold (nested within repeat). auROCs were

---

[3]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).

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/).

Raw Predictor
Originated
Call duration
Call answered
Date/time of communication
Phone number
Type of Relationship
Have you drank alcohol with this person?
What is their drinking status?
Would you expect them to drink in your presence?
Are they currently in recovery from drugs or alcohol?
Are they supportive about your recovery goals?
How are your typical experiences with this person?
DSM-5 symptom count
Past year alcohol problems
Craving
Abstinence self-efficacy: Negative affect, social, physical, and craving subscales
Number of individual alcohol counseling sessions attended (past 30 days)
Number of group alcohol counseling sessions attended (past 30 days)
Number of self-help group meetings attended (past 30 days)
Number of other mental health counseling sessions attended (past 30 days)
Number of days in contact with supportive people (past 30 days)
Number of days in contact with unsupportive people (past 30 days)
Taken prescribed medication for alcohol use disorder (past 30 days)
Taken prescribed medication for other mental health disorder (past 30 days)
Satisfaction with progress toward recovery goals (past 30 days)
Confidence in abstinence ability (next 30 days)
Has a goal of abstinence
Age
Sex at birth
Race
Education
Income

Raw Predictor					
Marital Status					

Note:
 Cellular communication features were scored over six scoring epochs (6, 12, 24, 48, 72, and 168 hours before t

Results

Participants

We screened 192 participants. Of these, 169 enrolled and 154 completed the first follow-up visit. Data from 1 participant was excluded due to not reporting a goal of abstinence and lapsing multiple times a day every day on study. Data from 1 participant was excluded due to evidence of careless responding. Data from 1 participant was excluded due to poor compliance with EMA resulting in questionable lapse labels. Data from 7 participants were excluded due to poor compliance providing communication data (i.e., deleting logs prior to download and/or not providing context information about important contacts). The final analytic sample included 144 participants. Table 2 provides the demographic characterization of our sample. 56% of participants reported at least one lapse while on study.

	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					

	N	%	M	SD	Range
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			
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: Demographic Characterization

Note:
N = 144

Communications

Participants had an average of 26 important contacts (range 2-113) that were contextualized with self-report information. We obtained a total of 375,912 contextualized communications

across participants. Participants had, on average, 2,610 contextualized communications (range = 109-14,225) averaging to about 33 communications per day (range 3-278).

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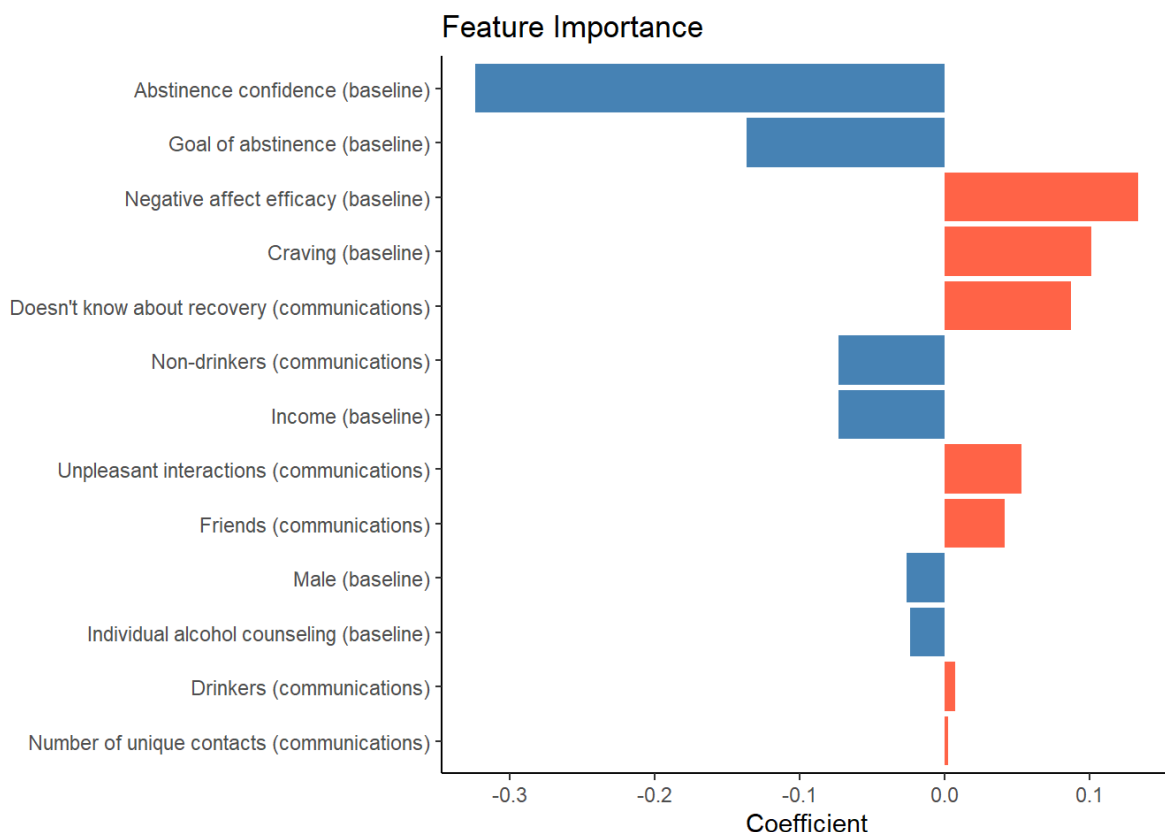
Figure 1: Global feature importance (elastic net coefficient) for the full model. Features are ordered by absolute coefficient value. Blue bars indicate higher feature values, on average, lower lapse risk. Red bars indicate higher feature values, on average, increase risk. Baseline features were collected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

Discussion

Our model achieved fair performance, with an auROC of 0.68, indicating that some predictive signal was present. However, it did not offer incremental value beyond a baseline model that included only demographic and self-report measures. Consistent with this, the four most important predictors in our model were all self-report variables: abstinence confidence, abstinence goal, negative affect efficacy, and craving.

Nonetheless, several communication features were retained in the final model with moderately sized coefficients. These included communications with people unaware of the participant's recovery status, non-drinkers, friends, and individuals who were unpleasant to interact with. 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 communication may be less informative than the quality and social significance. Future research may benefit from collecting richer contextual data about communication contacts to better understand the social dynamics contributing to lapse risk.

Even with highly contextualized communication data, however, prediction may be limited by data sparsity. Many participants had few daily communications, and some had extended periods with no recorded interactions at all. Our study design may have further contributed to this limitation. We collected only phone and SMS text communications through the native smartphone app. In recent years, many individuals use private messaging apps (e.g., WhatsApp, Signal) or social media platforms (e.g., Facebook Messenger, Instagram) as their primary communication method[18]. Therefore, our dataset likely missed a substantial portion of participants' communications. Notably, the communication features that were retained in the final model were scored over the longest scoring epochs (72 and 168 hours), suggesting that when more data are available (i.e., more communications) these features may be become more important. Future studies could explore whether increasing communication data from these alternative platforms yield stronger predictive signals.

We cannot entirely dismiss the potential value of cellular communication data for risk prediction. For example, researchers have successfully incorporated communication data into models with other sensing data (e.g., accelerometer, geolocation, and device usage) to detect current[19] and predict future[20] heavy drinking episondes in non-treatment seeking young adult populations. It is possible, that in certain populations cellular communications may hold more signal. Young adults may have more frequent communications reducing sparsity concerns. Additionally, non-treatment seeking populations may be less likely to sensor their data (i.e., deleting communications) when the drinking behavior is not at odds with their goals and/or values. However, even in these instances, the unique contribution of cellular communications above and beyond other sensing methods is unclear. Some communication features, such as outgoing call duration and the number of outgoing calls emerged in the top 20 important features for detecting current drinking episodes. Conversely, when predicting future drinking episodes, no communication features appeared in the top 20. Other sensing methods, like geolocation and accelerometer data, appeared to be more robustly important for both detection and prediction.

Other practical challenges in collecting call and text message data further limit the feasibility of this sensing method. For example, we obtained participants' cellular communication data by downloading backups of their communication logs in person during their monthly follow-up visits. However, Apple heavily restricts apps in its app store from accessing call and text message data, making real-time sensing of communications challenging (if not impossible) for IOS users. We conclude that other forms of social interaction characterization (e.g., engineering

time spent with supportive contacts from geolocation data) are more worthwhile to pursue in future research.

Author Contributions

KW contributed to conceptualization, data curation, formal analysis, methodology, visualization, writing – original draft, and writing - review and editing. CY contributed to conceptualization, data curation, methodology, writing - original draft. JJC contributed to conceptualization, data curation, methadology, writing – review and editing, supervision of analysis, funding acquisition.

Acknowledgments

The authors wish to thank Susan E. Wanta for her role as the project administrator.

Statements and Declarations

Ethical Considerations

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

Consent to Participate

All participants provided written informed consent.

Consent for Publication

Not applicable.

Declaration of Conflicting Interest

The authors declared no potential conflicts of interest with respect to the research, authorship, and publication of this article.

Funding

Data Availability

Our de-identified data, questionnaires, and other study materials are publicly available on our OSF page (https://osf.io/wgpz9/).

References

1.   McLellan AT, Lewis DC, O'Brien CP, et al. Drug dependence, a chronic medical illness: Implications for treatment, insurance, and outcomes evaluation. *JAMA* 2000; 284: 1689–1695.

2.   Dennis M, Scott CK. Managing Addiction as a Chronic Condition. *Addiction Science & Clinical Practice* 2007; 4: 45–55.

3.   Rounsaville DB. *Lapse, Relapse, and Chasing the Wagon: Post-Treatment Drinking and Recovery.* PhD Thesis, University of Maryland, Baltimore County, 2010.

4.   Marlatt GA, Gordon JR (eds). *Relapse Prevention: Maintenance Strategies in the Treatment of Addictive Behaviors.* First edition. New York: The Guilford Press, 1985.

5.   Marlatt GA, Donovan DM (eds). *Relapse Prevention, Second Edition: Maintenance Strategies in the Treatment of Addictive Behaviors.* 2nd edition. New York London: The Guilford Press, 2007.

6.   Daley DC, Douaihy A. Reducing the Risk of Relapse. In: Daley DC, Douaihy AB, Daley DC, et al. (eds) *Managing Substance Use Disorder: Practitioner Guide.* Oxford University Press, 2019, p. 0.

7.   Cleveland HH, Brick TR, Knapp KS, et al. Recovery and Recovery Capital: Aligning Measurement with Theory and Practice. In: Croff JM, Beaman J (eds) *Family Resilience and Recovery from Opioids and Other Addictions.* Cham: Springer International Publishing, 2021, pp. 109–128.

8.   Mohr DC, Zhang M, Schueller SM. Personal Sensing: Understanding Mental Health Using Ubiquitous Sensors and Machine Learning. *Annual Review of Clinical Psychology* 2017; 13: 23–47.

9.   Chih M-Y, Patton T, McTavish FM, et al. Predictive modeling of addiction lapses in a mobile health application. *Journal of Substance Abuse Treatment* 2014; 46: 29–35.

10.   Wyant K, Sant'Ana SJ, Fronk GE, et al. Machine learning models for temporally precise lapse prediction in alcohol use disorder. *Journal of Psychopathology and Clinical Science* 2024; 133: 527–540.

11.   Wyant K, Fronk GE, Yu C, et al. Forecasting Risk of Alcohol Lapse up to Two Weeks in Advance using Time-lagged Machine Learning Models.

12.   Jones A, Remmerswaal D, Verveer I, et al. Compliance with ecological momentary assessment protocols in substance users: A meta-analysis. *Addiction (Abingdon, England)* 2019; 114: 609–619.

13.   Wyant K, Moshontz H, Ward SB, et al. Acceptability of Personal Sensing Among People With Alcohol Use Disorder: Observational Study. *JMIR mHealth and uHealth* 2023; 11: e41833.

14.   Derogatis, L.R. *Brief Symptom Inventory 18 - Administration, scoring, and procedures manual.* Minneapolis: NCS Pearson, 2000.

15.   Hurlbut SC, Sher KJ. Assessing alcohol problems in college students. *Journal of American College Health* 1992; 41: 49–58.

16.   McKiernan P, Cloud R, Patterson DA, et al. Development of a Brief Abstinence Self-Efficacy Measure. *Journal of social work practice in the addictions* 2011; 11: 245–253.

17.   Flannery BA, Volpicelli JR, Pettinati HM. Psychometric properties of the Penn Alcohol Craving Scale. *Alcoholism, clinical and experimental research* 1999; 23: 1289–1295.

18.   McDowell B, Dumais KM, Gary ST, et al. Preferences and Attitudes Towards Digital Communication and Symptom Reporting Methods in Clinical Trials. *Patient preference and adherence* 2025; 19: 255–263.

19.   Bae S, Ferreira D, Suffoletto B, et al. Detecting Drinking Episodes in Young Adults Using Smartphone-based Sensors. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2017; 1: 1–36.

20.   Bae SW, Suffoletto B, Zhang T, et al. Leveraging Mobile Phone Sensors, Machine Learning and Explainable Artificial Intelligence to Predict Imminent Same-Day Binge Drinking Events to Support Just-In-Time Adaptive Interventions: A Feasibility Study. *JMIR formative research.* Epub ahead of print February 2023. DOI: 10.2196/39862.