

Personal Sensing of Smartphone Communications to Support Recovery from Alcohol Use
Disorder

Kendra Wyant¹

¹ Department of Psychology, University of Wisconsin-Madison

Author Note

The authors made the following contributions. Kendra Wyant: Conceptualization, Writing - Original Draft Preparation, Writing - Review & Editing.

Correspondence concerning this article should be addressed to Kendra Wyant, 1202 West Johnson St., Madison, WI 53706. E-mail: kpaquette2@wisc.edu

Abstract

Alcohol use disorder is a chronic relapsing disease. People with alcohol use disorder must often monitor their risk of relapsing for years. Lapses occur when a person with a goal of abstinence has a drink. These are often referred to as a slip and in isolation are not indicative of relapse. Lapses are, however, often an early warning sign of relapse. People with alcohol use disorder can have difficulty recognizing when they are at a high probability of lapsing. Current digital therapeutics aimed to treat and monitor alcohol and substance use disorders provide treatment resources for people, but they are not able to be helpful in identifying periods of high risk in real time. Personal sensing can be used to improve digital therapeutics by adding this temporal element - when is someone likely to lapse. Our study uses a novel method by combining a passive personal sensing data stream (cellular communication logs) with a more active data stream (self-reported context about communications). Here we show that cellular communication logs likely contain signal related to lapse risk. Additionally, our findings suggest that passively sensed data streams can potentially be as effective as more active data streams. This work sets the foundation for future research into optimizing digital therapeutics for long term lapse risk monitoring.

Keywords: personal sensing, digital phenotyping, alcohol use disorder, cellular communication logs, text messages, digital therapeutics, digital health

Word count: 6542

Personal Sensing of Smartphone Communications to Support Recovery from Alcohol Use Disorder

Alcohol Use Disorder

According to the National Survey on Drug Use and Health 14.1 million adults in the US met the criteria for alcohol use disorder in 2019 [1]. This is roughly 6% of the US adult population. When left untreated, alcohol use disorder can result in physical, mental, financial, and social consequences [2]. Unfortunately, many people do not receive treatment, and those who do typically only receive treatment for a short period of time and at times they are in acute distress (i.e., detoxification in a hospital setting, court-mandated treatment).

Yet, alcohol use disorder is a chronic relapsing disease [3,4]. People can relapse days, weeks, months, or even years after achieving abstinence [5–8]. Lapses, an initial setback or slip, are often an early warning sign of relapse, a full return to previous drinking behavior [9]. Studies show that lapses predict future lapses, with more frequent lapses resulting in increased chances of relapse [3]. Likewise, longer durations of abstinence is associated with decreased chances of lapse, suggesting the stability of a patient’s recovery itself as an important predictor of abstinence [3]. In fact, the most important predictor of relapse is whether the individual has already had a lapse during treatment [10]. This high correlation between an initial lapse and relapse is the abstinence violation effect, which states that people who internalize feelings of loss of control, guilt, and hopelessness after violating a self-imposed rule (i.e., abstinence) have a greater risk of relapse compared to those who view the lapse as external and controllable [10,11]. Thus, identifying when an initial lapse will occur is an important goal in preventing lapses, repeated lapses, and relapse.

Like most mental health disorders, alcohol use disorder can be characterized by fluctuations in affective and behavioral states that covary with the severity of the underlying disorder and impact its treatment. The ability to detect these changes offers the opportunity to selectively deliver different interventions to patients that match their needs in that

moment. For example, during periods of stability, patients in recovery from alcohol use disorder might benefit from interventions that help them modify their social network and daily activities to include more time with family and friends who support their recovery. In contrast, during times of peak stress and associated alcohol craving, they might need focused interventions that prevent an imminent lapse back to alcohol use.

Lapse Risk Factors

Lapses may seem to come out of nowhere, but they are often preceded by changes in cognitive (e.g., negative affect, craving), situational (e.g., lack of social support, risky situations), and behavioral (e.g., decreased social interactions, coping strategies) factors [3,6,7].

These risk factors are known as proximal risk factors that are fluid and change over time (much like how mental health disorder symptoms can ebb and flow through periods of stability and periods of lapse) [3,5]. Proximal risk factors contrast distal risk factors - stable states or traits that contribute to a mental health disorder, but do not change much over time. For example, one's genetic makeup, the presence of co-occurring psychopathology, or alcohol use disorder severity may contribute to whether one is predisposed to alcohol use disorder, but they are less involved in the cyclical nature of lapses [12]. On the other hand, fluctuations in proximal factors often precipitate a lapse [9,13,14]. But due to the dynamic nature of these factors, traditional treatments (e.g., monthly therapy or counseling sessions) may not be best suited for monitoring for changes in lapse risk.

Not all proximal risk factors are detectable within the same time window. Situational factors, such as passing a bar or being around people who are drinking may immediately precede a lapse. Other stressors, such as getting upsetting news from a family member or increased financial strain, may slowly increase lapse risk over hours, days or weeks. Thus, a system for continuously monitoring a person's lapse risk is critical in detecting when

someone will lapse.

Personal Sensing

With new advances in technology, we can now accomplish this through personal sensing. Personal sensing is a longitudinal in situ measurement approach for collecting data via smartphone sensors, logs, and social media apps [15]. Current digital therapeutics, platforms designed to deliver evidenced-based treatments and interventions, do not yet integrate these data [16]. Yet there is much potential for these therapeutics to harness personal sensing data to capture fluctuations in risk in real time.

The first step to improving digital therapeutics with personal sensing is to be able to build accurate and effective predictive algorithms. Research in the area of mental health broadly has begun to use personal sensing data to predict constructs such as social anxiety [17], depression [18–22], bipolar disorder [23,24] and schizophrenia [25–27].

In the area of substance use the focus has been on predicting drinking or substance use episodes. Some studies have been able to use smartphone-sensors to predict current drinking episodes [28,29]. However, this is too late to deliver preventative interventions. Two recent studies have been able to predict future drug craving with geolocation data [30] and future smoking lapses with EMA data [31]. These predictive algorithms, however, are not yet meeting the threshold for clinical implementation.

It is also important to note that geolocation and EMA place differing amounts of burden on the individual providing the data. Active methods, like EMA, require frequent (i.e., at least daily) interactions with the personal sensing platform and as such place more effort on the individual using the platform. Passive methods, like geolocation, require minimal interactions (i.e., installing an app and setting permissions once) with the personal sensing platform and operate passively in the background.

Additionally, different personal sensing data could have different utility in personal sensing-assisted digital therapeutics. For example, geolocation may be useful for identifying features predictive of an immediate lapse (i.e., being in or near a bar) and could be used to deploy harm-reduction strategies. On the other hand, EMA could be used to identify features more predictive of a slow gradual risk of lapse (i.e., how they slept last night, how much pain they are in, and whether they expect to face any upcoming stressors this week) and recommend connecting with a therapist.

Cellular Communication Logs

One understudied personal sensing method in the lapse risk literature is cellular communication logs. Cellular communication logs are both passive and they have the potential to capture risk factors on varying timetables. We may be able to capture immediate risk based on who someone is calling or what time of day it is. Additionally, slower acting risk indicators may also be detected. Decreased interactions may signify isolation common with depressive symptoms, reaching out to people in one's social network could signify a positive coping strategy, or changes in patterns between a single person in one's social network could indicate conflict [8,12,13].

Not only does it have potential predictive power, but it also appears to be a feasible source of personal sensing data. In a smartphone-based sensing platform the primary expense on the individual is the smartphone. Smartphone usage is already widespread. Eighty-five percent of US adults have a smartphone and this number is consistent across all sociodemographic groups, including those in recovery programs for substance use [32,33]. Additionally, people generally find personal sensing of cellular communication logs to be acceptable [34]. That is people are no less likely to provide cellular communication logs than other personal sensing data streams (i.e., EMA, Geolocation, Sleep).

Social Context

Many proximal lapse predictors are inherently social [35]. For example, decreased interactions may signify isolation common with depressive symptoms, reaching out to people in one's social network could signify a positive coping strategy, or changes in patterns between a single person in one's social network could indicate conflict. Accordingly, these factors are intrinsically linked to social interactions [35–37].

On its own, the nature of people's social interactions with others appears to be a salient factor relevant to lapse risk. Positive social interaction has been associated with positive alcohol treatment outcomes [38–41]. It is thought that the social component of self-help groups like Alcoholics Anonymous (i.e., increasing pro-abstinence relationship ties) play a major contributing role in their efficacy [39,42–44]. The buffering hypothesis of social support suggests that social relationships may buffer individuals against the negative effects of stress by providing a source of resources to promote adaptive responses to stress [45,46]. With stress being an often-cited precipitant of alcohol lapse, it is likely these buffering effects extrapolate to lapse risk [47].

There is also much literature on the importance of social network influences on drinking behavior [35,48]. A social network made up of heavy drinkers or abstainers can ultimately influence one's recovery towards maintaining abstinence or lapsing [48–50]. Individuals in recovery who maintain interactions with people who do not support their recovery is associated with decreased likelihood of maintaining abstinence. Social interactions with negative peer influences (e.g., heavy drinkers) may have more of an effect on treatment outcome than positive peer influences (e.g., abstainers) [48]. This relationship between social support and lapses is robust and extends to substances other than alcohol [51].

Since many social interactions occur or are planned over the phone, cellular communication logs offer insight into patterns of social interaction. We know that social

relationships are dynamic [52–55]. So, with time-stamped logs there is potential to build a sensing system that tracks how these interactions fluctuate over time. For example, changes in their top contacts (i.e., who they are interacting with the most) or changes in patterns of activities (e.g., is an individual is consistently ignoring phone calls from a contact they normally interact often with?).

However, these communications may become much more powerful when they are contextualized (e.g., pleasantness of interactions with the contact, whether the contact is supportive of their recovery, drinking history with contact). For example, context in concert with communication logs can tell us that someone is reaching out to other people in recovery or that their top contacts have recently shifted to include only people supportive of their recovery. So, we can imagine that providing this context to cellular communication logs could greatly increase the predictive signal. Take one more example - an outgoing phone call late at night might be related to lapse risk, but the direction will be very different if they are calling their sponsor vs. a former drinking buddy.

Present Study

The present study seeks to build a personal sensing algorithm to predict next hour lapse probability from contextualized cellular communication logs. Identifying when someone will lapse with hour-by-hour temporal precision is a critical first step in improving digital therapeutics. If we can detect when someone is most likely to lapse we can deliver interventions directly to the individual instead of expecting them to engage with the therapeutic on their own.

Since people with alcohol use disorder may require long-term monitoring, it is important that a digital therapeutic does not place too much burden on the individual. As such we will evaluate whether any incremental benefit exists from the more active context (e.g., pleasantness of interactions with the contact, whether the contact is supportive of their

recovery, drinking history with contact) when compared to a passive only configuration.

This study will extend our current understanding of distal risk factors by identifying additional proximal risk factors over varying time windows. We will evaluate the importance of specific features within our predictive algorithm. Identifying the top performing features will set the foundation for future studies to begin to not only predict *when* someone will lapse but also *which* treatment should be delivered.

Method

This project performs the initial analyses on a subset of data collected from 2017 – 2019 as part of a larger grant funded through the National Institute of Alcohol Abuse and Alcoholism (R01 AA024391). A full list of measures used in the parent project can be found at [<https://osf.io/brs68/>].

Research Transparency

We value the principles of research transparency that are essential to the robustness and reproducibility of science [56]. Consequently, we maximized transparency by reporting all data exclusions, manipulations, and available measures in the study [57]. We also made the data, analysis scripts and annotated results, self-report surveys, and other study materials associated with this study publicly available [<https://osf.io/brs68/>].

Participants

We recruited a community sample of people in initial stages of recovery for alcohol use disorder from the Madison area. Avenues of recruitment included referrals from clinics, self-help groups, Facebook, radio, and television. Those who were interested in participating in our three-month study were given a brief description (i.e., told that our research focused on learning how mobile health technology can be used to provide individual support to anyone recovering from alcohol addiction) a short phone screen to determine initial eligibility

(i.e., At least 18 years old, ability to read and write in English, and an eligible smartphone with existing cellphone plan).

One hundred ninety-two participants met criteria for enrollment after completing a more detailed in-person screening visit (Figure 1). We excluded potential participants if they did not meet the criteria for moderate or severe alcohol use disorder (as defined by the DSM-5), did not have a goal of long-term abstinence, had not abstained from alcohol for at least one week, already had over two months of abstinence, or had severe symptoms of psychosis or paranoia. One hundred ninety-one participants consented to participate in the study at the screening session and 169 subsequently enrolled in the study at the enrollment visit which occurred approximately one week later. Of the 169 participants enrolled in the study, 154 provided at least one month of communication logs and EMA surveys. This was an additional requirement for the current study because cellphone communications were collected at each one month follow-up. We dropped data from three participants due to concerns about the quality of their data. These exclusions occurred during the cleaning process and prior to any analyses. Our final analyses were on a sample of 151 participants.

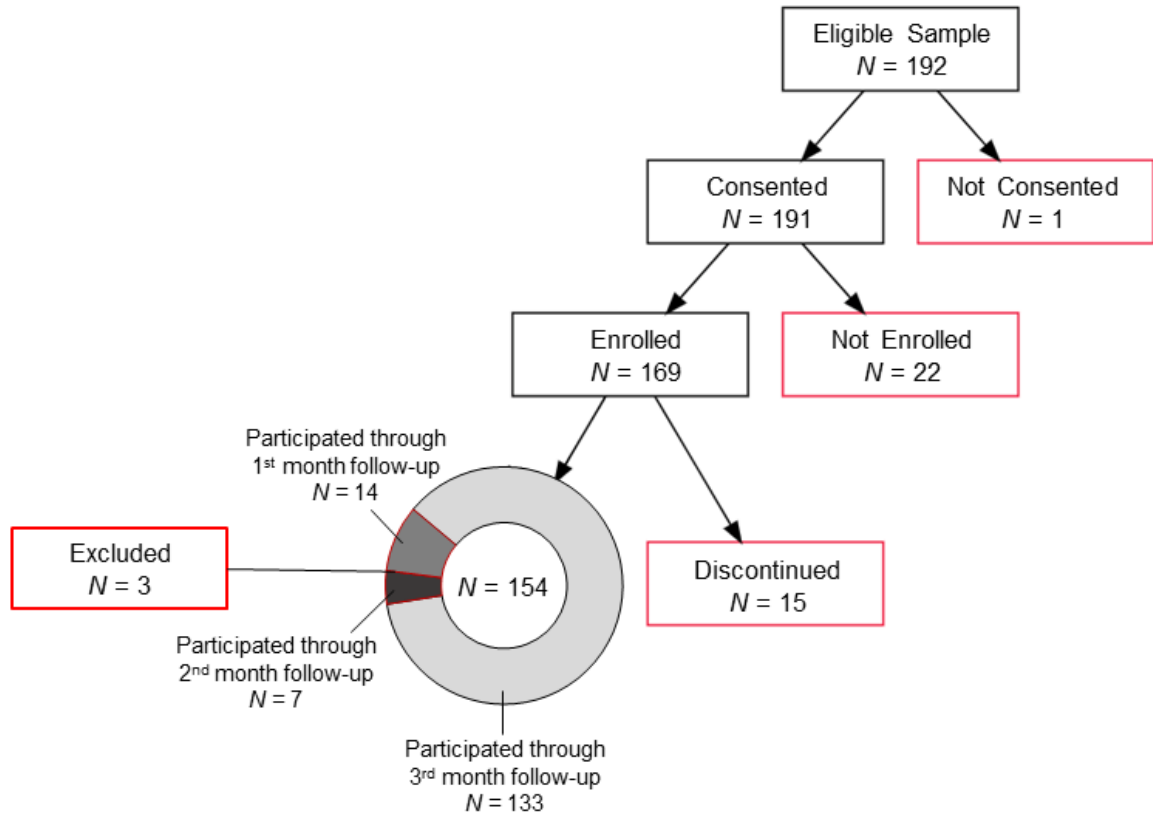


Figure 1. Flowchart of participant retention. The red boxes show all points participants were discontinued or withdrawn. 154 participants completed through the first month follow-up. Of these 154, three participants' data were excluded leaving us with an analysis sample of 151.

Procedure

Our study involved five in-person visits (screening, enrollment, and three follow-up visits). It also required completion of daily EMA surveys to document alcohol lapses, and access to non-deleted text message and call logs (i.e., cellular communication logs). All procedures were approved by the University of Wisconsin-Madison Institutional Review Board.

During the screening session we obtained informed consent, determined eligibility, and documented demographic information, history of lifetime drug and alcohol use, and mental health traits. Participants that consented and were deemed eligible came back for a second

enrollment visit. During enrollment, participants were briefed on how to delete log entries they did not want to share with us and completed a practice EMA survey. Participants also reported contacts they frequently communicated with and answered a series of questions documenting context information about their interactions with each contact (type of relationship with contact, whether they drank with contact in past, drinking status of contact, whether contact would drink in their presence, whether the contact is in recovery, the level of supportiveness the contact provides, and the pleasantness of their interactions with the contact). Participants returned for three follow-up visits, each one month apart.

At each follow-up visit, we downloaded participants' cellular communication logs. These communication logs included the phone number of the other party, whether the call or message was incoming or outgoing, the duration of the call, and the date and time of the call or message. Participants also provided context information for newly identified frequent contacts (i.e., at least two communications in the past month). At the third follow-up visit, participants were debriefed and thanked for their participation.

Feature Engineering and Model Training

We used various feature engineering methods to maximize model performance. Features were selected over several different windows of data preceding the lapse prediction (6, 12, 24, 48, 72, and 168 hours). For example if we were using a window of 24 hours and wanted to predict probability of a lapse at 6:00 PM then we would use all available features since 6:00 PM the day before up until the hour we were predicting for. We believed it was important to include varying time windows because it is likely that some features, like communicating with an old drinking friend, may be more temporally close to a lapse (i.e., 6 hours) and other features, like a slow increase in number of unanswered incoming calls, may reflect likelihood of a lapse over a longer period of time (i.e., 168 hours).

To condense several log entries into a single feature we wrote several functions that

scaled each feature on two metrics (raw and percent). Raw features included things such as the raw proportion of incoming calls, the raw mean duration of calls, the raw rate of text messages per hour, and the raw count of incoming text messages from supportive contacts. Percent features converted the raw features into a percent difference. To do this we first calculated the baseline (proportion, rate, etc.) over all available features (i.e., not restricted to a specific time window) prior to the date and time of the label we were predicting. We then calculated whether the current raw feature was higher or lower than their baseline using the formula $(raw - baseline)/baseline$. We did not know a priori which type of feature would be best (raw, percent, or raw and percent) so we trained models using all three combinations of features.

Phone number and participant ID were used jointly as the unique identifier to merge the communication logs with the context information. Communication logs were from several OS versions of both Android and iPhone smartphones. The phone numbers had to be carefully cleaned to match the context data. This included standardizing numbers by removing US but not non-US country codes, removing special characters, and deriving missing area codes from a participant’s log and context data. I wrote a series of functions that cleaned, extracted, and validated the phone numbers [<https://osf.io/d72c9/>]. US numbers were valid if they had a registered North-American area code and were 10 digits long. Non-US numbers were validated based on country code, length of number, and when applicable area code. Once we were able to match the communication logs to the context data we could derive context features (e.g., number of communications with drinkers) and combination features (e.g., proportion of incoming calls from people who are not in recovery).

When engineering features, we distinguished between active features (features derived from context) and passive features (features derived from communication logs). This distinction was used during model training to specify which feature set we wanted to use. The active feature set used all available features (active and passive) and the passive feature

set was restricted to only passive features.

Since we had an imbalanced outcome variable, we also considered different class resampling methods during model training. We used an up-sampling method where our positive labels (lapses) were bootstrapped until they equaled the number of negative labels (no lapses). We used a down-sampling method where our negative labels were randomly sampled from until they equaled the number of positive labels. We also used the Synthetic Minority Over-Sampling Technique (SMOTE) to produce new synthetic positive labels until they equaled the number of negative labels. These resampling methods were only used on the training data and not applied to the held-out validation sets.

Our candidate model configurations differed on statistical algorithm (glmnet, random forest, Knn), feature set (active, passive), feature type (percent, raw, percent and raw), class resample (up, down, SMOTE). Each model configuration was also tuned on the hyperparameters associated with each statistical algorithm (i.e., lambda and alpha for glmnet, mtry, min n and trees for random forest, and neighbors for Knn). This resulted in us training over 20,000 unique configurations.

Data Analysis Plan

We conducted all analyses in R version 4.0.3 [58] using RStudio [59] and the tidyverse and tidymodels ecosystem of packages [60,61].

We characterized our sample of participants across standard demographics (age, sex, race, ethnicity, income, education level, employment status, and marital status). We also characterized our participants according to life history of drug and alcohol use. This includes alcohol use disorder milestones (age of first drink, age of regular drinking, age of problematic drinking, age of first quit attempt), number of quit attempts, average number of days spent drinking prior to quitting, treatment history for alcohol use disorder (long-term residential, short-term residential, outpatient, individual counseling, group counseling, self-help groups),

whether they have received medication for alcohol use disorder, the number of DSM-5 criteria for alcohol use disorder reported, number and types of drugs used over their lifetime, and whether they have ever injected drugs. Finally we characterized various mental health outcomes of our participants using validated measures such as the Symptom Checklist-90 [62], Intolerance of Uncertainty Scale [63], Anxiety Sensitivity Index-3 [64], Distress Tolerance Scale [65], McMaster Family Assessment Device [66], and the Multidimensional Personality Questionnaire Short Form [[67].

We collected lapse start and end times from participants with a four-time daily EMA. Our outcome labels for training our model consisted of lapse or no lapse at the level of the hour. This means each participant had 24 labels per day for up to 90 days. We treated each label as an independent observation. As expected our negative labels (no lapses) were much more prevalent than positive labels (lapses). To lessen this imbalance we randomly selected a subset of negative labels that were removed so that our labels were down-sampled to about 19 negative labels for every 1 positive label. Our feature engineering was then performed on this subset of observations. We characterized the frequency of lapses across individual participants.

Our first study aim was to train and evaluate the best performing machine learning model to predict alcohol lapse onset from contextualized cellular communication data. We built, trained, and evaluated models with several statistical learning algorithms including penalized parametric linear classification algorithms (LASSO, ridge regression, glmnet), non-parametric classification algorithms (k nearest neighbor), and ensemble methods (random forest). Candidate statistical learning algorithms were trained on a subset of the data (i.e., training sets) using combinations of features derived from participants cellular communications and context information.

Balanced accuracy (i.e., the average of sensitivity and specificity) was our performance metric for both model selection and evaluation. However, during model evaluation we also

report other performance metrics (sensitivity, specificity, positive predictive value, negative predictive value) to fully characterize the performance of our best model configuration. We also visually inspected lapse probability predictions by participant to further probe model performance.

During model selection, we used grouped 10-fold cross-validation (grouped 10X CV) without repeats (i.e., 1 set of 10 grouped folds). We selected the top performing model configuration (statistical algorithm and combination of features) according to highest balanced accuracy. All cross-validation folds were grouped by participant ID so that a single participant’s data were not being used to predict future data on the same participant. In other words, all data from a participant was either included in the training or held-out validation set, but not both.

We evaluated our top performing model configuration using grouped 10-fold cross validation with 10 repeats (grouped 10x10-fold CV). We opted to use resampling for model evaluation instead of an independent held-out test set because, with our limited sample size, averaging balanced accuracy over 100 held-out validation sets gives us a more stable (i.e., low variance) and less biased estimate of model performance than other available options using a single test set.

Finally, we used a Bayesian correlated t-test [68] to compare our top model configuration’s balanced accuracy estimates across the 100 validation sets to a balanced accuracy of .50 (i.e., the expected performance of a null model with no predictive signal). In this analysis, we defined a meaningful difference to be more than a difference of 1% in either direction (i.e., Region of Practical Equivalence [ROPE] set to .49 - .51).

Our study’s second aim was to compare the best model configuration using all available features from both passive signals from communications logs and actively measured context (i.e., active) with the best model configuration restricted to only passive signals (i.e.,

passive). To do this we employed a model comparison approach. Our top passive and active model configurations were selected using grouped 10-fold CV and then evaluated using grouped 10x10-fold CV. We then used a Bayesian correlated t-test to compare balanced accuracy from each of the 100 validation sets of our top active model configuration with our top passive model configuration. We defined a meaningful difference between model configurations to be more than a 1% difference in either direction (i.e., ROPE set to -.01 - .01). Through this relative comparison, we quantified any performance benefit from adding features based on the the active component of context.

Our third aim of the present study was to evaluate the importance of features within the top performing model configuration. The most predictive features were identified based on an algorithm agnostic feature importance index. This approach involves permuting each individual feature to determine how much predictive signal was lost when removed from the model.

Results

Participant Characteristics

Participants were on study for an average of 85 days out of the possible 90 days. All participants provided at least one month of data. Table 1 shows the demographic breakdown for the 151 participants used in our analyses. The majority of participants identified as white (86.8%). Just over half of participants identified as male (51%) and 49% identified as female. Table 2 shows the relevant drug and alcohol history for these participants and table 3 characterizes our participants mental health. Participants had endorsed using on average 4 other types of drugs (not including alcohol) over their lifetime. Additionally, participants on average scored a 9 on a self-report version of the DSM-5 symptom criteria for alcohol use disorder. Generally, scores of 2-3 are considered mild, 4-5 are considered moderate, and 6+ considered severe alcohol use disorder.

On average participants had about 32 contacts ($SD = 22$) they frequently communicated with. Figures 2 - 7 show the distribution of participant responses about their frequent contacts to each of our six context questions. Seventy-two percent of our participants ($N = 109$) have at least one person they frequently communicate with who they used to drink with regularly (always/almost always). Additionally, 99% of our participants ($N = 150$) regularly communicate with at least one person they know to be a drinker. In fact, each participant communicate with on average 18 people ($M = 17.7$, $SD = 15.0$) who are known drinkers and 95% of participants ($N = 143$) said that at least one of these contacts are likely to drink in front of them in the future. On the recovery-positive side of these communications, 99% of participants ($N = 149$) have a supportive contact they frequently communicate with and 80% ($N = 120$) have regular communications with someone else in recovery. Finally, on average across participants, 75% of communications ($SD = .17$) were reported to be pleasant.

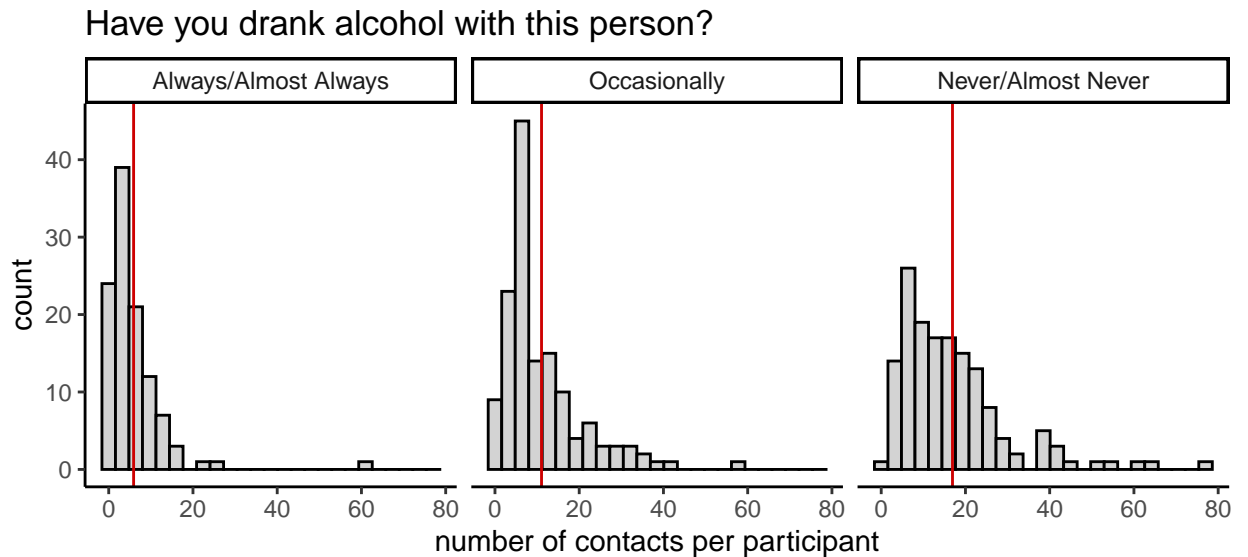


Figure 2. Have you drank alcohol with this person? Distribution of number of contacts each participant has for each response option (Always/Almost Always, Occassionally, Never/Almost Never). Mean number of contacts is depicted as the solid red line.

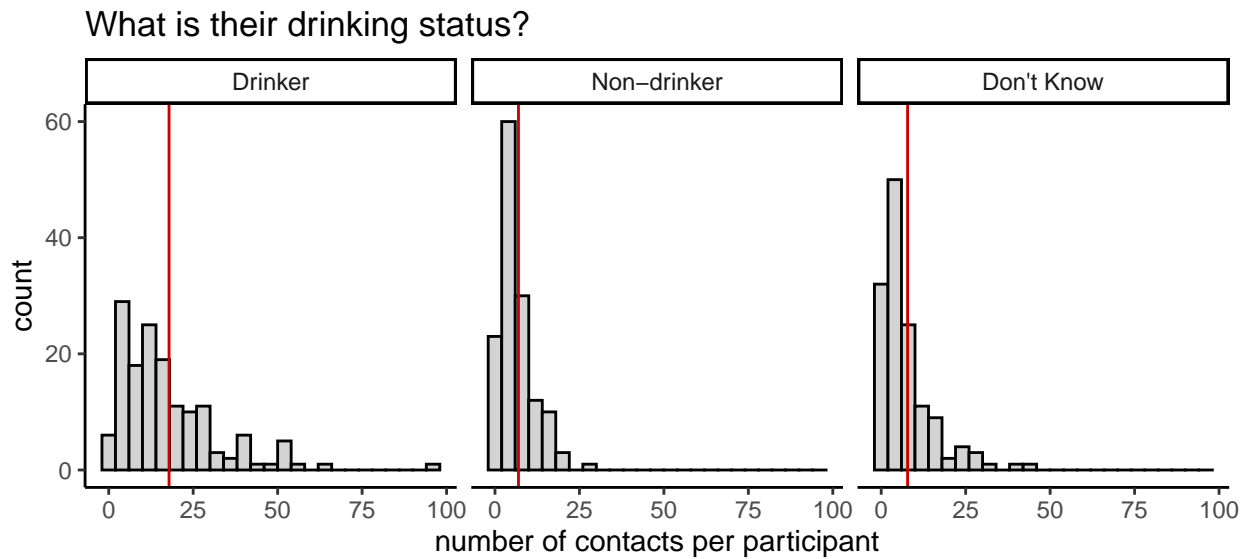


Figure 3. What is their drinking status? Distribution of number of contacts each participant has for each response option (Drinker, Non-drinker, Don't Know). Mean number of contacts is depicted as the solid red line.

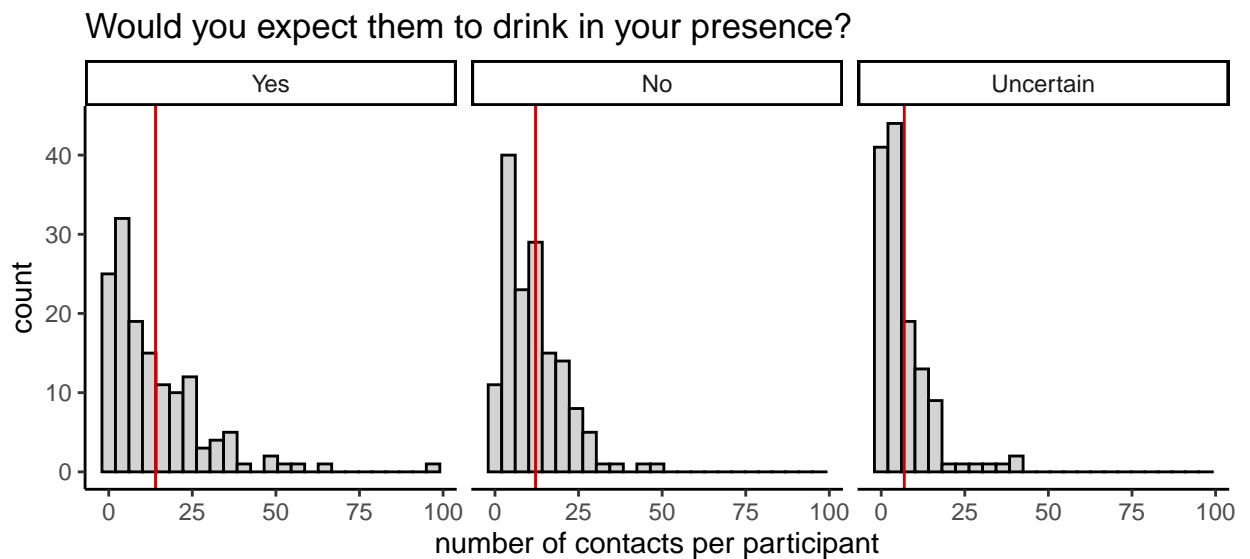


Figure 4. Would you expect them to drink in your presence? Distribution of number of contacts each participant has for each response option (Yes, No, Uncertain). Mean number of contacts is depicted as the solid red line.

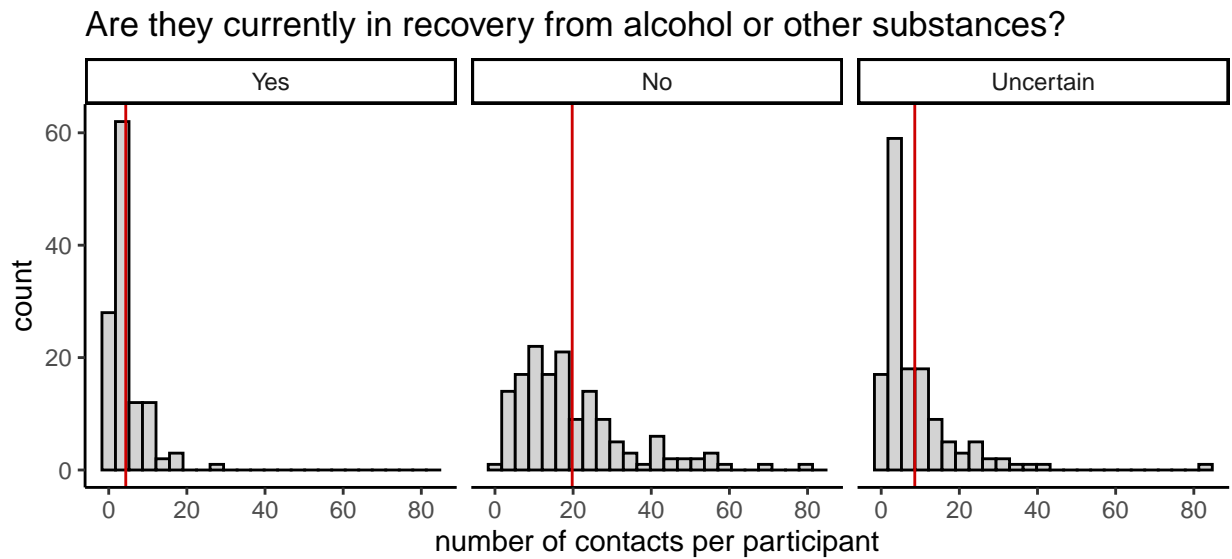


Figure 5. Are they currently in recovery from alcohol or other substances? Distribution of number of contacts each participant has for each response option (Yes, No, Uncertain). Mean number of contacts is depicted as the solid red line.

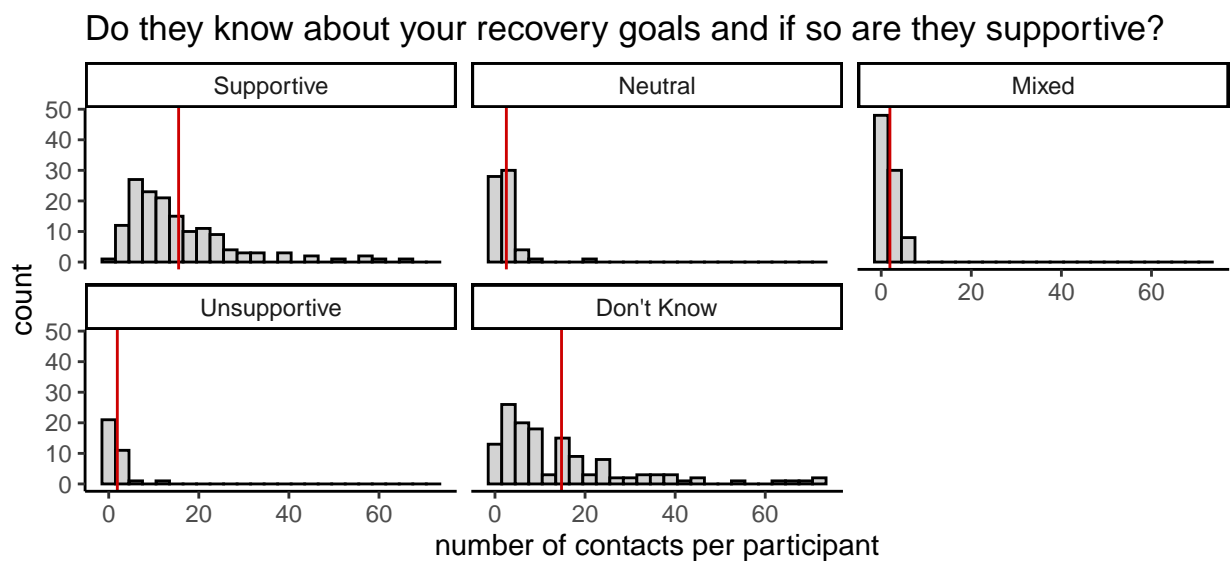


Figure 6. Do they know about your recovery and if so are they supportive? Distribution of number of contacts each participant has for each response option (Supportive, Neutral, Mixed, Unsupportive, Don't Know). Mean number of contacts is depicted as the solid red line.

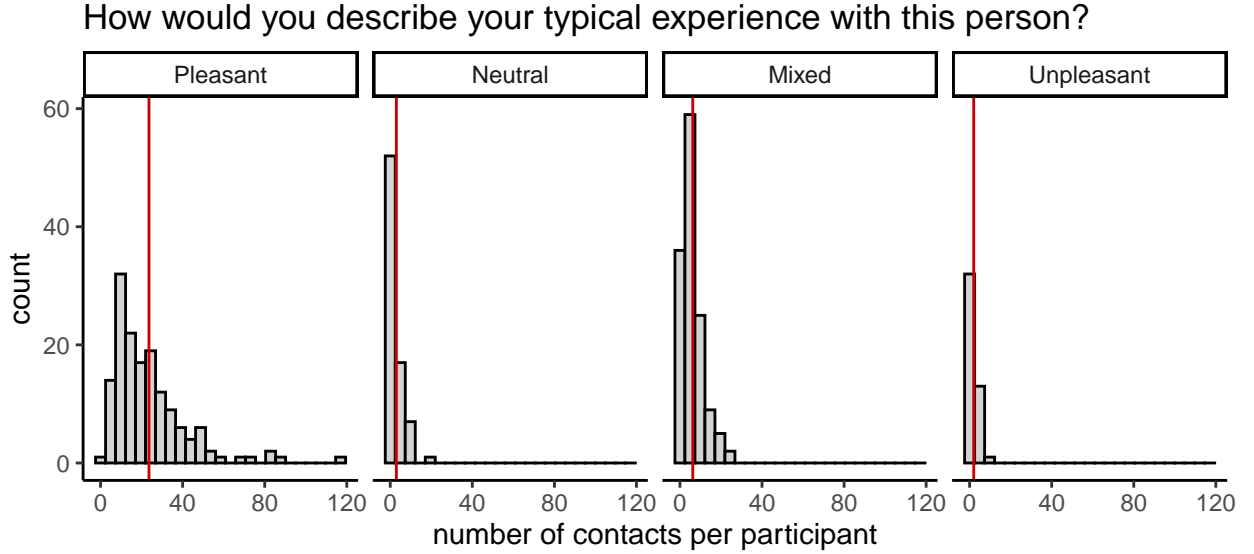


Figure 7. How would you describe your typical experience with this person?

Distribution of number of contacts each participant has for each response option (Pleasant, Neutral, Mixed, Unpleasant). Mean number of contacts is depicted as the solid red line.

Lapse characterization

Across participants there were a total of 1029 positive lapses. After down-sampling non-lapses this made up 5% of our labels. There was variation in the frequency of lapses, ranging from 0-75 lapses per participant ($M = 6.8$, $SD = 12.0$). Only 56% of participants ($N = 84$) reported a lapse. However, this was expected since our participants all had a goal of abstinence from alcohol. Figure 8 shows a histogram of frequency of lapses for each participant.

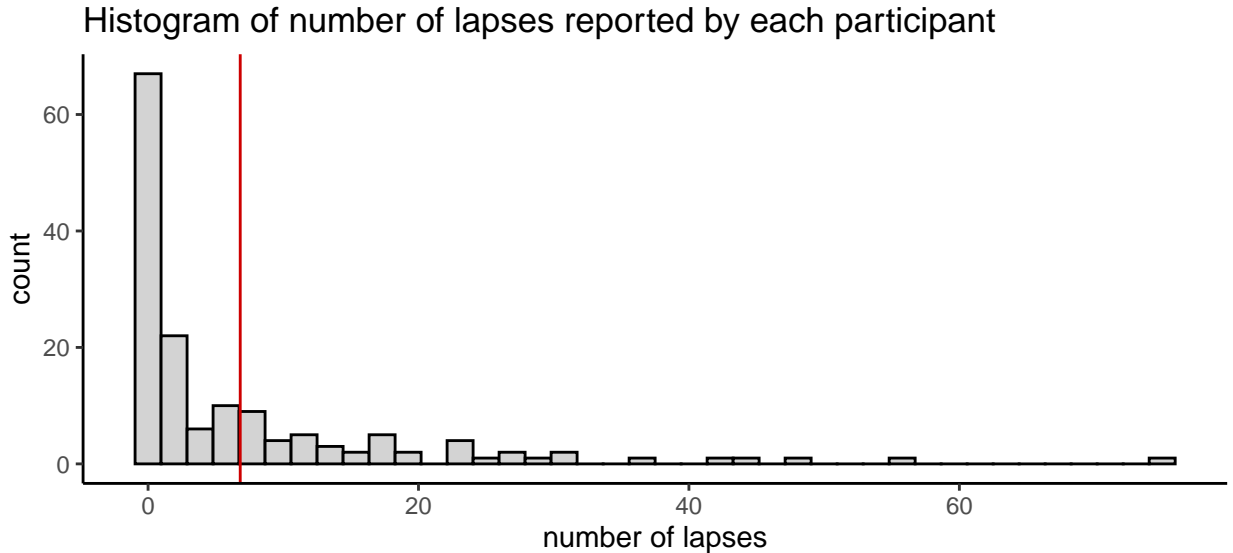


Figure 8. Histogram of number of lapses reported by each participant. Fifty-six percent of participants ($N = 84$) reported at least one lapse while on study. The mean number of lapses reported is depicted as the solid red line.

Best Model Performance

Model configurations differed by statistical algorithm (glmnet, random forest, Knn), feature set (active, passive), feature type (percent, raw, percent and raw), and class resample (up, down, SMOTE). We also tuned each statistical algorithm to select the optimal hyperparameter values. We then selected the best model configuration using a grouped 10-fold cross-validation (grouped 10-fold CV) resampling method. We selected our top candidate configurations on balanced accuracy. Table 4 shows the best performing model configuration (i.e., highest balanced accuracy) for each statistical algorithm (glmnet, random forest, Knn). We selected the statistical algorithm with highest balanced accuracy to be our top performing model configuration.

Our top performing model configuration was a random forest statistical algorithm using passive features and down-sampling. To reduce the effects of optimization bias on our evaluation of our top model configuration’s predictive performance, we re-estimated the

performance of this top performing configuration using grouped 10-fold cross validation with 10 repeats (grouped 10x10 CV). This means we trained our model configuration on 100 different subsets of training data and predicted on the 100 held out validation sets. We then averaged across the 100 performance estimates to get a single estimate with low variance. This method gave us a balanced accuracy estimate of .60.

Table 5 shows a confusion matrix where we can see how well the model configuration predicts on new data for negative cases (i.e., no lapse) compared to positive cases (i.e., lapses). Table 6 characterizes our best model configuration over several metrics appropriate for classification. Our best model had a specificity of .73 and a sensitivity of .46. This means out of all the no lapses we are correctly identifying 73% of them and out of all the lapses we are correctly identifying 46% of them.

Figure 9 shows a distribution of posterior probabilities for the 100 balanced accuracy estimates. We performed a Bayesian correlated t-test to compare these estimated to a balanced accuracy of .50 (i.e., what we would expect from a model with no signal). We defined the Region of Practical Equivalence (ROPE) to be within 1% higher or lower than .50. The ROPE is depicted as the vertical dashed lines. The test showed there is more than a 99% posterior probability that our model is performing better than .51. This suggests there is a meaningful difference between our selected model configuration and a model with no signal.

In the appendix we show that our best performing model configuration has variation in predictions for each individual participant. Figure A1 contains predictions that are predicted probabilities of a lapse. Each observation was held out 10 times (grouped 10x10 CV) and the figure shows the averaged probability across these 10 predictions. Actual lapses, are depicted in red.

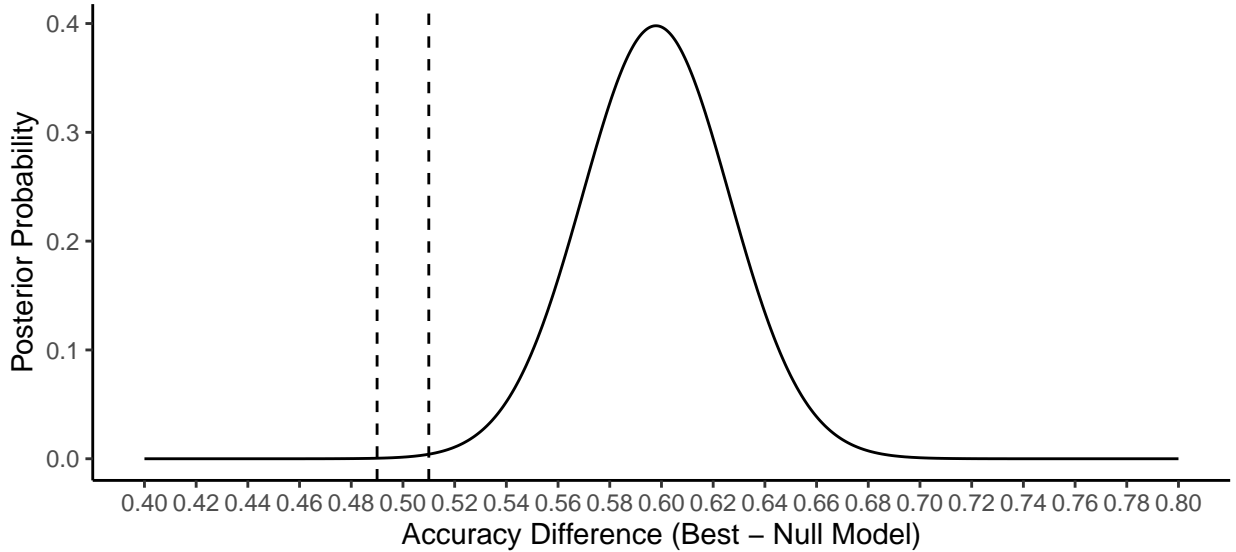


Figure 9. Posterior probabilities of balanced accuracy estimates for the best model (passive random forest) compared to a balanced accuracy of .50.

Model Comparison between Active and Passive

Our best performing active model configuration (i.e., using all the features available) was a glmnet algorithm using up-sampling. We retrained this model configuration on 100 subsets of data and predicted on the held out validation set 100 times (grouped 10x10 CV). This gave us an averaged balanced accuracy of .60.

Figure 10 shows the balanced accuracy estimates for passive and active model configurations over the 100 held-out validation sets. A Bayesian correlated t-test revealed there were no differences between the best active and best passive model configuration. The posterior probability that our active model performed meaningfully better (i.e., to the right of the ROPE) than the passive model was .37 (Figure 11). There is a 20% posterior probability that the two models are equivalent and a 43% posterior probability that the passive model is better. Thus, there is not strong evidence to support the collection of context to calculate active features at this point.

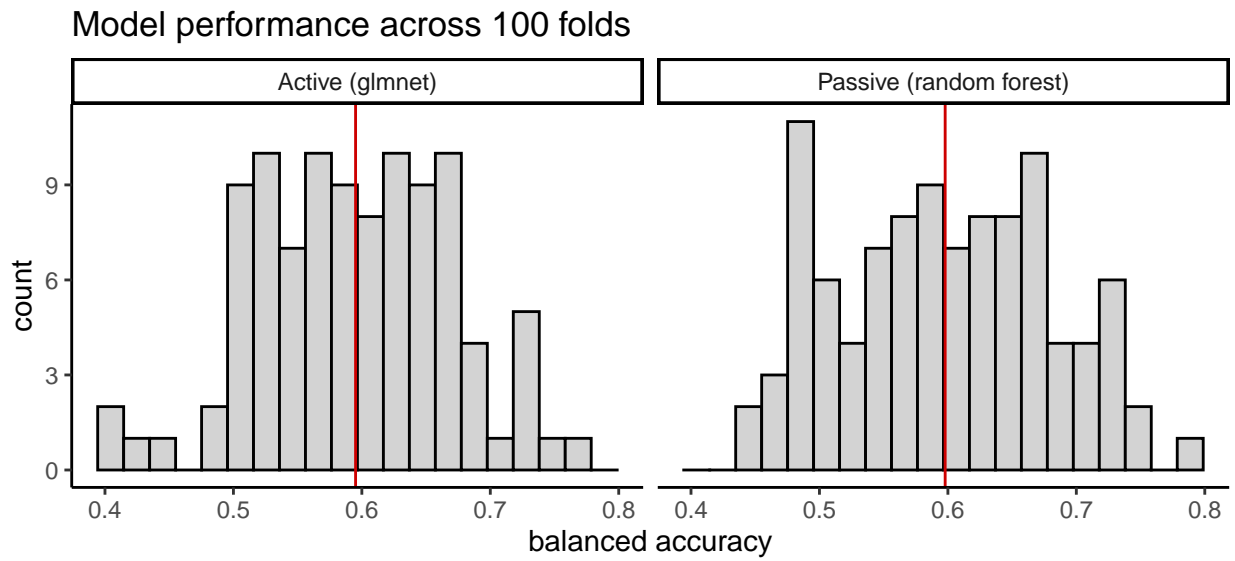


Figure 10. Histogram of best active (glmnet) and passive (random forest) model configuration balanced accuracies across 100 validation sets.

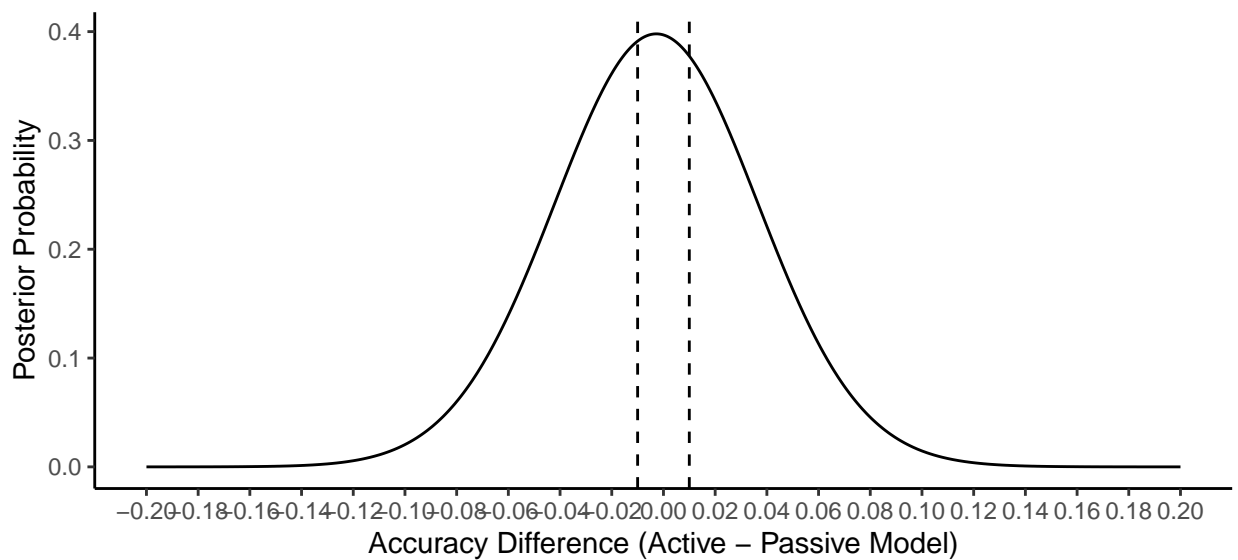


Figure 11. Model comparison of the best active (glmnet) and best passive (random forest) model configurations.

Top Features

We permuted each feature within the top performing model configuration (passive random forest) to determine each feature’s importance. Figure 12 shows which features, when permuted, resulted in the greatest reduction in balanced accuracy.

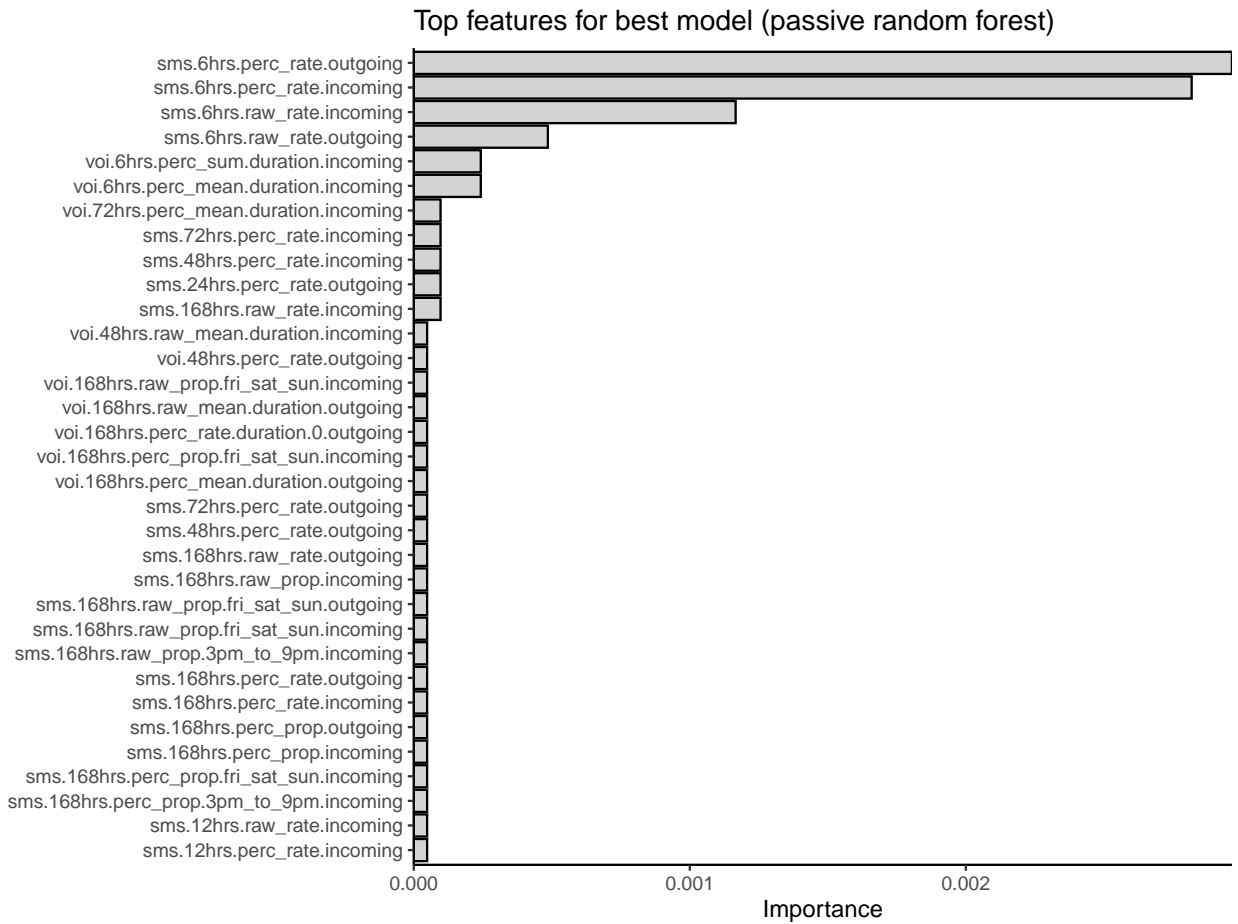


Figure 12. Feature importance scores for best performing model configuration (Passive Random Forest).

Discussion

The analyses in this paper make up the initial analyses for our study into how cellular communications and context can be used for alcohol lapse probability prediction. As of now, our top performing model configuration shows there is signal in the passive communication

logs. However, it is far from achieving the level of precision necessary for clinical implementation. We especially see this with its poor positive predictive value. We know that positive predictive value is greatly impacted by imbalanced outcomes. Predicting at the level of the hour dramatically increases this imbalance. It is possible that with our data we may be more successful building models that predict lapse probability at the level of the day.

Surprisingly we found that active model configurations did not perform better than passive model configurations. One possible explanation for this is the high dimensional space that the active model configurations were being trained in. Whereas the passive models were only given 122-244 features, the active models were given up to 5,488 features. This is an especially large ratio of predictors to observations when we consider we are training our models on only about 18,500 observations (9/10ths of 20,580). It is possible that using this large number of features resulted in the model configurations overfitting to noise in the data. Thus, when making predictions on new data the performance suffered.

The feature importance plot shows that there are only a handful of features with substantial impact. With the features that do rank as highly important, we have no way of knowing which direction it is predictive of. For example, our top two features were the percent rate of incoming and outgoing text messages over the past six hours. This suggests that changes in frequency of text messaging over short time windows are highly related to lapse probability. What we don't know is whether increased text messaging is a protective factor or a risk factor when it comes to lapses. This is one difficulty that comes with trying to interpret less interpretable algorithms like random forest. One way we might be able to evaluate the direction of these predictors is to use partial regression plots. This would allow us to see the residuals of a model regressing the outcome label on all predictors except for the one of interest against the residuals of the predictor of interest regressed on all other predictors.

The most notable area for improving our models, however, is at the feature engineering

stage. There are several additional ways we might be able to capitalize on the data available to us. We opted to not use demographic, drug and alcohol use, and mental health characteristics of our participants (Tables 1-3) in an effort to reduce the number of features and focus on the predictive value of communication logs and context. Feature sets quickly become high dimensional and complex when adding these data. Therefore, the solution may be to put more thought into specific combinations of features or use other algorithms that may be more adept at handling this high dimensionality. In addition to participant characterization. We can likely gain good signal from temporal characteristics of the label. We might hypothesize that lapses are more common at night or on weekends. Therefore, the hour we are predicting on itself might serve as a predictor.

Although the present findings suggest there was no added value from active context. It is possible we are not using context to its full potential. Additional feature engineering that involves condensing response options to fewer groups, adding interactions between context and demographic or other static variables, examining different temporal windows, and using alternative statistics to describe the features could all serve to more efficiently extract lapse prediction signal.

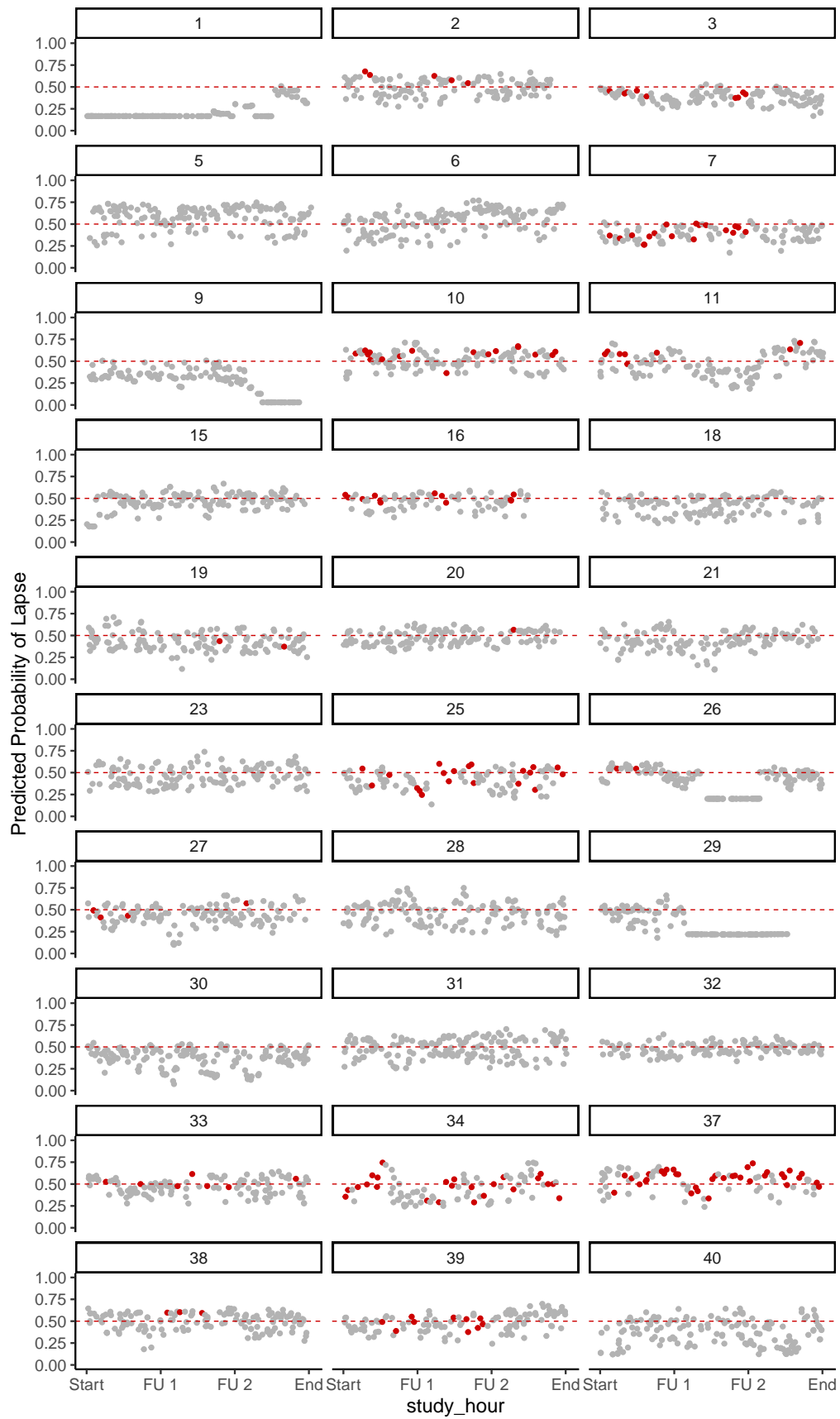
We can also imagine that cellular communication signal could become much stronger if we also used the participant's own language (i.e., text message content) or we knew where the participant had recently been (i.e., geolocation). These are signals that we currently have access to and plan to implement as we build and train similar models. If we find that a combination of raw data streams are most effective for lapse probability prediction it would be no more difficult on the individual providing the data. All three of these data sources (communication logs, text message content, and geolocation) are passively sensed suggesting it is no more burdensome to collect all three data streams compared to just communication logs.

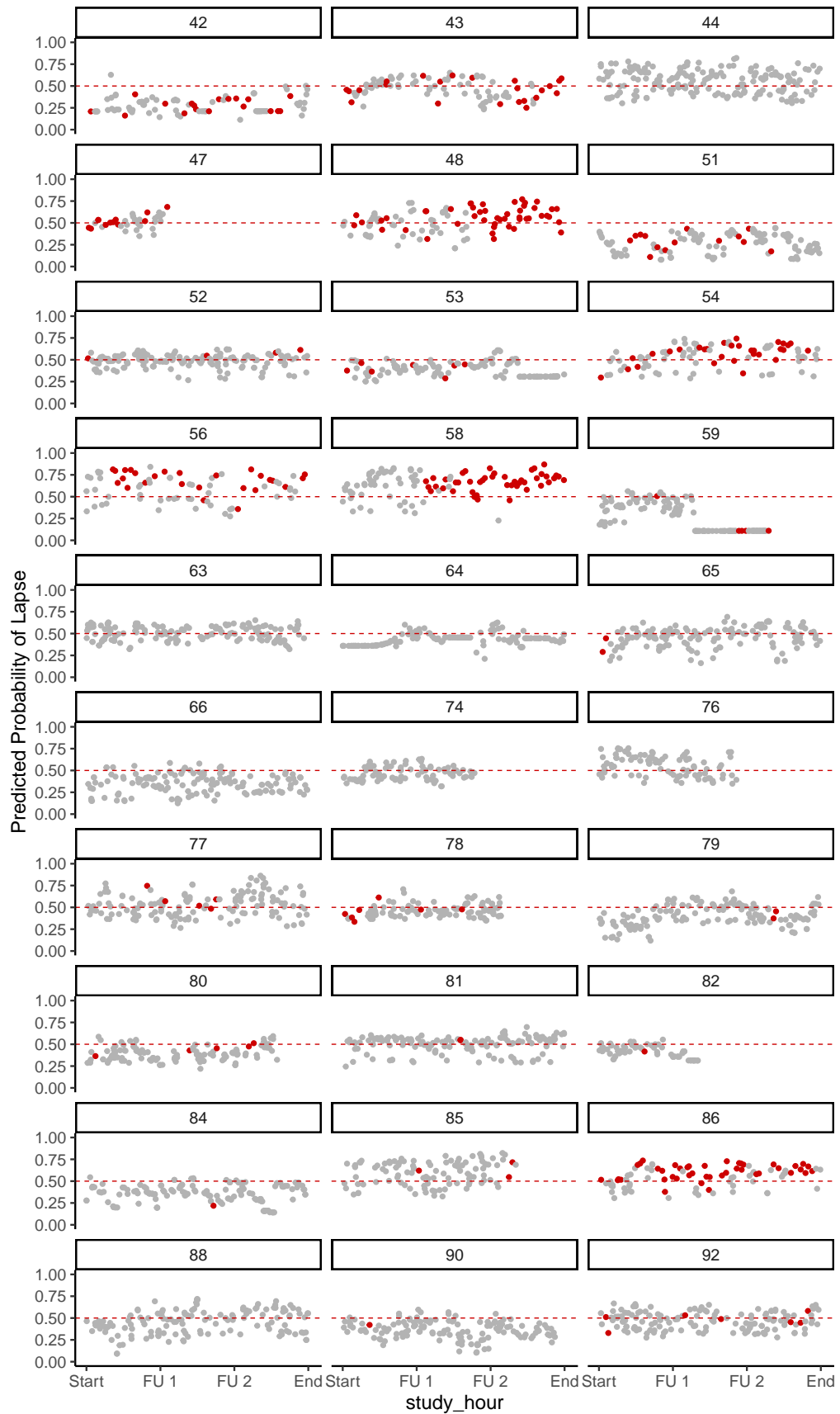
As we have implied, a successful predictive algorithm that is able to detect when a

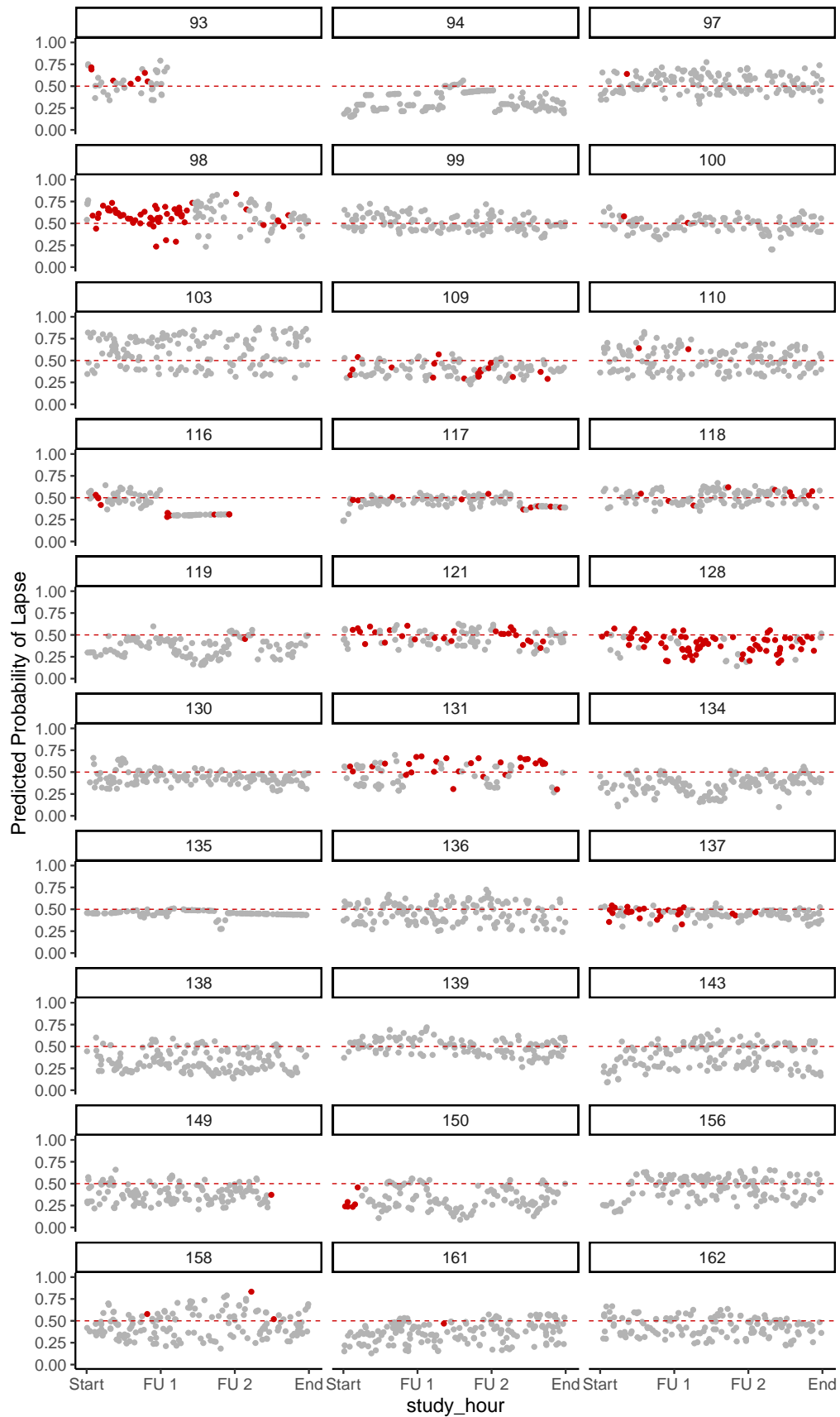
lapse will occur using personal sensing is an important first step in improving digital therapeutics. The second step is to determine when to intervene and what intervention to provide. The present study predicts the onset of a lapse to the hour (0 lag). While this may be an important algorithm for deploying immediate interventions (e.g., urge surfing, connection to a sponsor, setting drink limits) it may be too late for other interventions (e.g., seeing your therapist). Future research should look at varying lag windows that can predict a day or week out from a lapse. Once we know the optimal time window to intervene, we can then begin to look at treatment matching. Contextual bandit machine learning models can be used to help solve the issue of which treatment is right for an individual given what we know about them from their data. Micro-randomized trials can then be used to systematically evaluate the selected treatment's efficacy at a series of decision points [69].

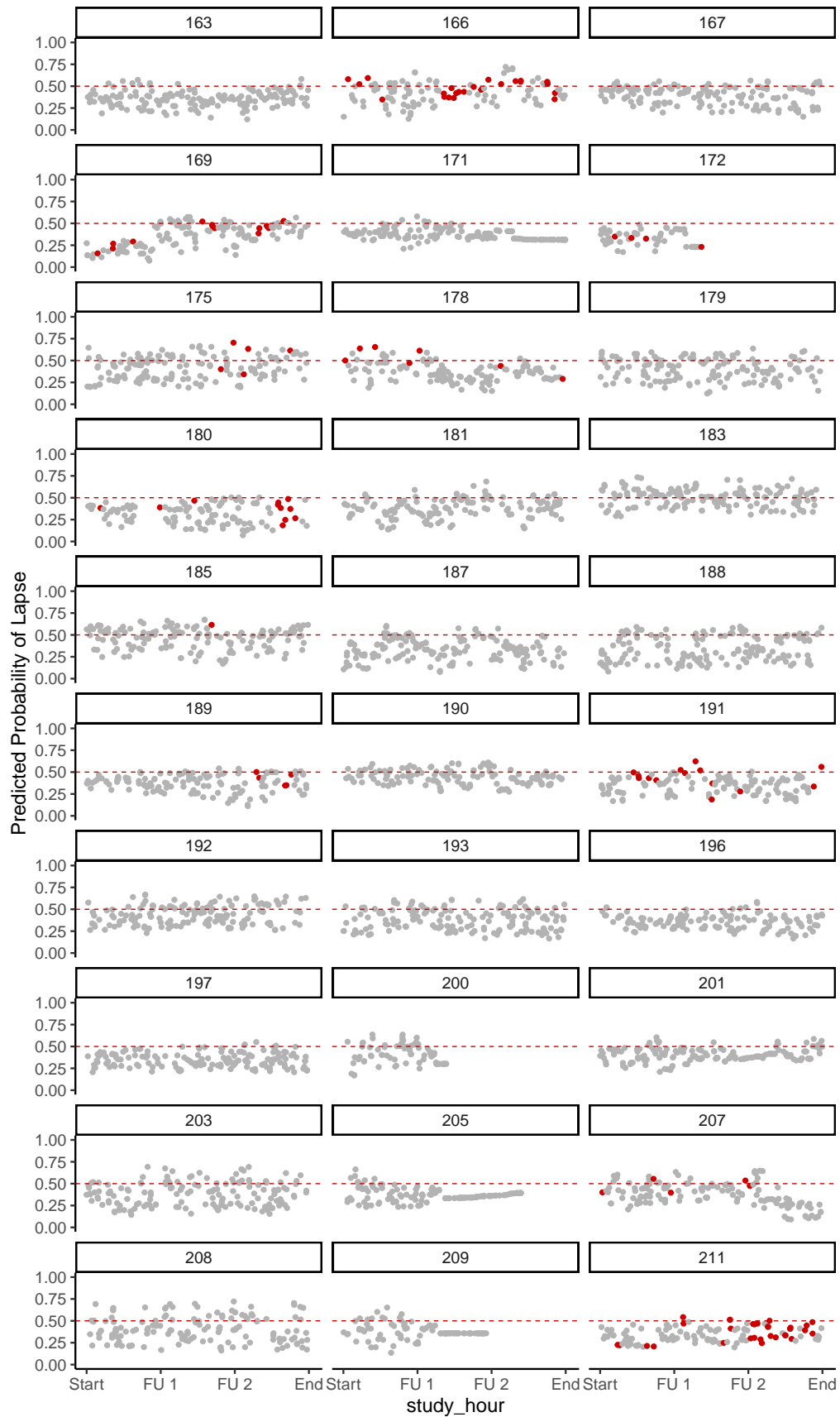
Clearly there is much more work to be done in this area. However, our study provides a starting point that establishes cellular communication logs have predictive signal and that passive features can potentially be as effective as more active features.

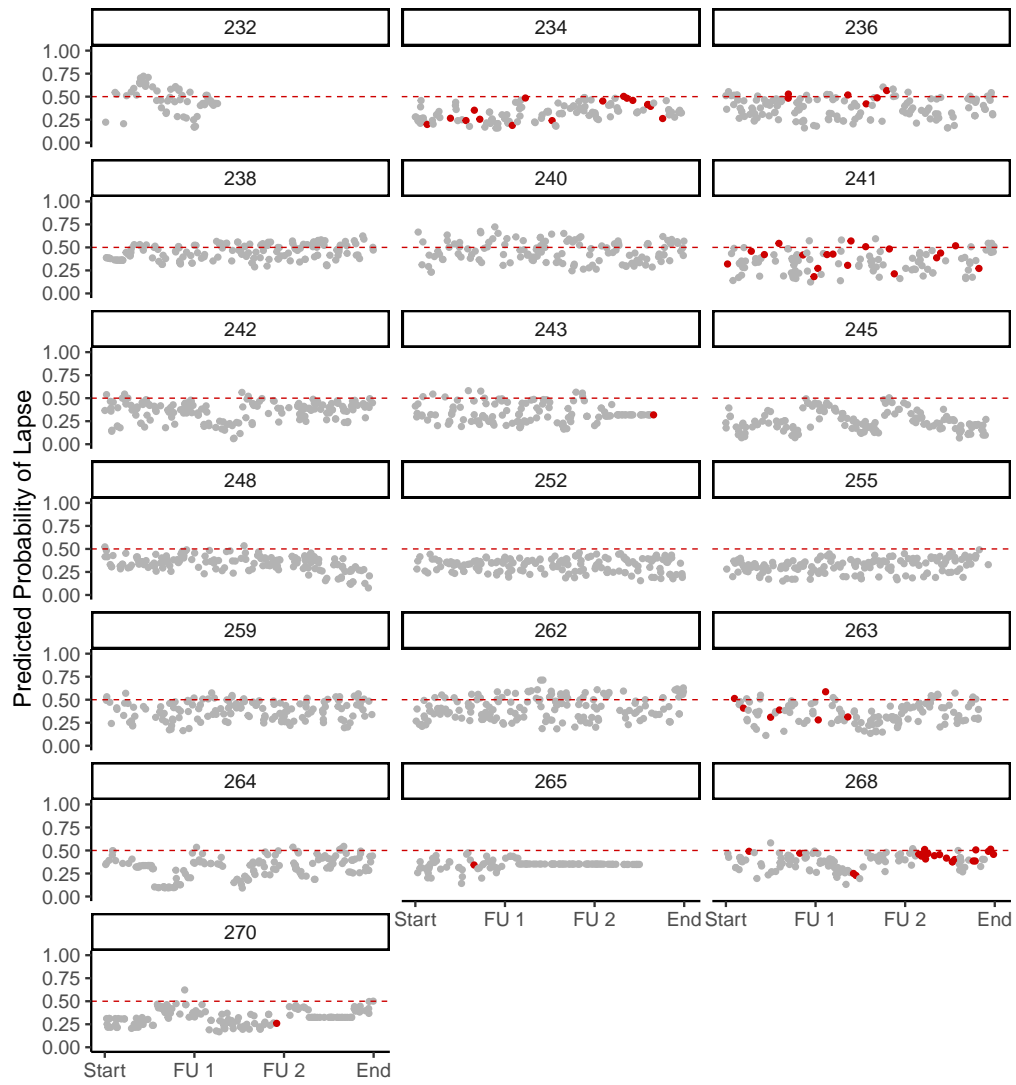
Appendix











study_hour

Figure A1. Predicted probabilities of lapse for each participant. A grouped 10x10 resampling method was used to obtain these probabilities. Known lapses are in red. The red dashed line represents the threshold for classifying a probability as a lapse (i.e., everything above the line was predicted to be a lapse).

References

1.

Substance Abuse and Mental Health Services Administration (SAMHSA). 2019 National Survey on Drug Use and Health. Center for Behavioral Health Statistics and Quality; 2019.

2.

Rehm J. The Risks Associated With Alcohol Use and Alcoholism. *Alcohol Research & Health* 2011;34(2):135–143. PMID:22330211

3.

Rounsaville DB. Lapse, Relapse, and Chasing the Wagon: Post-Treatment Drinking and Recovery [PhD thesis]. [United States – Maryland]: University of Maryland, Baltimore County; 2010.

4.

Scott CK, Foss MA, Dennis ML. Pathways in the relapse-treatment-recovery cycle over 3 years. *Journal of Substance Abuse Treatment* 2005;28 Suppl 1:S63–72. PMID:15797640

5.

Witkiewitz K. Lapses following alcohol treatment: Modeling the falls from the wagon. *Journal of Studies on Alcohol and Drugs* 2008 Jul;69(4):594–604. PMID:18612576

6.

Jin H, Rourke SB, Patterson TL, Taylor MJ, Grant I. Predictors of relapse in long-term abstinent alcoholics. *Journal of Studies on Alcohol* 1998 Nov;59(6):640–646. [doi: ghqvc8]

7.

Andersson H, Wenaas M, Nordfjærn T. Relapse after inpatient substance use treatment: A prospective cohort study among users of illicit substances. *Addictive Behaviors* 2018 Nov;90. [doi: 10.1016/j.addbeh.2018.11.008]

8.

Miller WR, Walters ST, Bennett ME. How effective is alcoholism treatment in the United States? *Journal of Studies on Alcohol* 2001 Mar;62(2):211–220. PMID:11327187

9.

Witkiewitz K, Marlatt GA. Relapse prevention for alcohol and drug problems: That was zen, this is tao. *American Psychologist* 2004 May;59(4):224–235. [doi: 10.1037/0003-066X.59.4.224]

10.

Högström Brandt AM, Thorburn D, Hiltunen AJ, Borg S. Prediction of single episodes of drinking during the treatment of alcohol-dependent patients. *Alcohol (Fayetteville, NY)* 1999 May;18(1):35–42. PMID:10386663

11.

Marlatt GA, Witkiewitz K. Relapse Prevention for Alcohol and Drug Problems. :44.

12.

Hufford MR, Witkiewitz K, Shields AL, Kodya S, Caruso JC. Relapse as a nonlinear dynamic system: Application to patients with alcohol use disorders. *Journal of Abnormal Psychology* 2003;112(2):219–227. [doi: 10.1037/0021-843X.112.2.219]

13.

Chih M-Y, Patton T, McTavish FM, Isham AJ, Judkins-Fisher CL, Atwood AK, Gustafson DH. Predictive modeling of addiction lapses in a mobile health application. *Journal of Substance Abuse Treatment* 2014 Jan;46(1):29–35. PMID:24035143

14.

Larimer ME, Palmer RS, Marlatt GA. Relapse Prevention. *Alcohol Research & Health* 1999;23(2):151–160. PMID:10890810

15.

Mohr DC, Zhang M, Schueller SM. Personal Sensing: Understanding Mental Health Using Ubiquitous Sensors and Machine Learning. *Annual Review of Clinical Psychology* 2017;13(1):23–47. PMID:28375728

16.

Alliance DT. Digital Therapeutics Definition and Core Principles. 2019.

17.

Jacobson NC, Summers B, Wilhelm S. Digital Biomarkers of Social Anxiety Severity: Digital Phenotyping Using Passive Smartphone Sensors. *Journal of Medical Internet Research JMIR Publications Inc., Toronto, Canada; 2020 May;22(5):e16875. [doi: 10.2196/16875]*

18.

Wang Y, Ren X, Liu X, Zhu T. Examining the Correlation Between Depression and Social Behavior on Smartphones Through Usage Metadata: Empirical Study. *JMIR mHealth and uHealth JMIR Publications Inc., Toronto, Canada; 2021 Jan;9(1):e19046. [doi: 10.2196/19046]*

19.

Razavi R, Gharipour A, Gharipour M. Depression screening using mobile phone usage metadata: A machine learning approach. *Journal of the American Medical Informatics Association* 2020 Apr;27(4):522–530. [doi: 10.1093/jamia/ocz221]

20.

Bai R, Xiao L, Guo Y, Zhu X, Li N, Wang Y, Chen Q, Feng L, Wang Y, Yu X, Xie H. Tracking and Monitoring Mood Stability of Patients With Major Depressive Disorder by Machine Learning Models Using Passive Digital Data: Prospective Naturalistic Multicenter Study. *JMIR MHEALTH AND UHEALTH* 2021;15.

21.

Reece AG, Danforth CM. Instagram photos reveal predictive markers of depression. *EPJ Data Science* 2017 Aug;6(15):1–12. [doi: 10.1140/epjds/s13688-017-0110-z]

22.

De Choudhury M, Counts S, Horvitz E. Predicting postpartum changes in emotion and behavior via social media. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13 Paris, France: ACM Press; 2013. p. 3267.* [doi: 10.1145/2470654.2466447]

23.

Palmius N, Tsanas A, Saunders K, Bilderbeck A, Geddes J, Goodwin G, de Vos M. Detecting Bipolar Depression From Geographic Location Data. *IEEE Transactions on Biomedical Engineering* 2016 Oct;PP:1–1. [doi: 10.1109/TBME.2016.2611862]

24.

Faurholt-Jepsen M, Bauer M, Kessing LV. Smartphone-based objective monitoring in bipolar disorder: Status and considerations. *International Journal of Bipolar Disorders* 2018 Jan;6. PMID:29359252

25.

Mitchell M, Hollingshead K, Coppersmith G. Quantifying the Language of Schizophrenia in Social Media. 2015. p. 20. [doi: 10.3115/v1/W15-1202]

26.

Barnett I, Torous J, Staples P, Sandoval L, Keshavan M, Onnela J-P. Relapse prediction in schizophrenia through digital phenotyping: A pilot study. *Neuropsychopharmacology* 2018 Jul;43(8):1660–1666. [doi: 10.1038/s41386-018-0030-z]

27.

Mcmanus K, Mallory E, Goldfeder R, Haynes W, Tatum J. Mining Twitter Data to Improve Detection of Schizophrenia. *AMIA Joint Summits on Translational Science proceedings AMIA Summit on Translational Science* 2015 Aug;2015:122–6.

28.

Bae S, Ferreira D, Suffoletto B, Puyana JC, Kurtz R, Chung T, Dey AK. Detecting Drinking Episodes in Young Adults Using Smartphone-based Sensors. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2017 Jun;1(2):1–36. [doi: 10.1145/3090051]

29.

Bae S, Chung T, Ferreira D, Dey AK, Suffoletto B. Mobile phone sensors and supervised machine learning to identify alcohol use events in young adults: Implications for just-in-time adaptive interventions. *Addictive Behaviors* 2018 Aug;83:42–47. PMID:29217132

30.

Epstein DH, Tyburski M, Kowalczyk WJ, Burgess-Hull AJ, Phillips KA, Curtis BL, Preston KL. Prediction of stress and drug craving ninety minutes in the future with passively collected GPS data. *npj Digital Medicine* 2020 Dec;3(1):26. [doi: ghqvcw]

31.

Hébert ET, Suchting R, Ra CK, Alexander AC, Kendzor DE, Vidrine DJ, Businelle MS. Predicting the first smoking lapse during a quit attempt: A machine learning approach. *Drug and Alcohol Dependence* 2021 Jan;218:108340. [doi: ghpt54]

32.

Center PR. Mobile Fact Sheet. 2021.

33.

Masson CL, Chen IQ, Levine JA, Shopshire MS, Sorensen JL. Health-related internet use among opioid treatment patients. *Addictive Behaviors Reports* 2019 Jun;9:100157. [doi: ghqvc2]

34.

Wyant K, Moshontz H, Stephanie B. Ward, John J. Curtin. Acceptability of personal sensing among people with alcohol use disorder. in prep.

35.

Hunter-Reel D, McCrady B, Hildebrandt T. Emphasizing interpersonal factors: An extension of the Witkiewitz and Marlatt relapse model. *Addiction* 2009 Aug;104(8):1281–1290. [doi: bqvszsj]

36.

Witkiewitz K, Marlatt GA. Emphasis on Interpersonal Factors in a Dynamic Model of Relapse. *American Psychologist* American Psychological Association; 2005 May;60(4):341–342. [doi: bxvs99]

37.

Stanton M. Relapse Prevention Needs More Emphasis on Interpersonal Factors. *American Psychologist* American Psychological Association; 2005 May;60(4):340–341. [doi: frq49s]

38.

Groh DR, Jason LA, Davis MI, Olson BD, Ferrari JR. Friends, Family, and Alcohol Abuse: An Examination of General and Alcohol-Specific Social Support. 2010;11.

39.

Kelly JF, Stout RL, Magill M, Tonigan JS. The role of Alcoholics Anonymous in mobilizing adaptive social network changes: A prospective lagged mediational analysis. 2012;21.

40.

Zywiak WH, Longabaugh R, Wirtz PW. Decomposing the relationships between pretreatment social network characteristics and alcohol treatment outcome. *Journal of Studies on Alcohol* 2002 Jan;63(1):114–121. PMID:11925053

41.

Longabaugh R, O'Malley SS. Network Support as a Prognostic Indicator of Drinking Outcomes: The COMBINE Study. *JOURNAL OF STUDIES ON ALCOHOL AND DRUGS* 2010;10. [doi: ghqvc7]

42.

Groh D, Jason L, Keys C. Social network variables in alcoholics anonymous: A literature review. *Clinical Psychology Review* 2008 Mar;28(3):430–450. [doi: b86245]

43.

Woff I, Toumbourou J, Herlihy E, Hamilton M, Wales S. Service providers' perceptions of substance use self-help groups. *Substance Use & Misuse* 1996 Aug;31(10):1241–1258. PMID:8879073

44.

Humphreys K, Mankowski ES, Moos RH, Finney JW. Do enhanced friendship networks and active coping mediate the effect of self-help groups on substance abuse? *Annals of Behavioral Medicine: A Publication of the Society of Behavioral Medicine* 1999;21(1):54–60. PMID:18425655

45.

Cohen S, Wills TA. Stress, Social Support, and the Buffering Hypothesis. :48. [doi: b4k]

46.

Holt-Lunstad J, Smith TB, Layton JB. Social relationships and mortality risk: A meta-analytic review. *PLoS medicine* 2010 Jul;7(7):e1000316. PMID:20668659

47.

Fronk GE, Sant'Ana SJ, Kaye JT, Curtin JJ. Stress Allostasis in Substance Use Disorders: Promise, Progress, and Emerging Priorities in Clinical Research. *Annual Review of Clinical Psychology* 2020;16(1):401–430. PMID:32040338

48.

Alvarez MJ, Richards DK, Oviedo Ramirez S, Field CA. Social network heavy drinking moderates the effects of a brief motivational intervention for alcohol use among injured patients. *Addictive Behaviors* 2021 Jan;112:106594. [doi: ghqvdc]

49.

Gordon AJ, Zrull M. Social networks and recovery: One year after inpatient treatment. *Journal of Substance Abuse Treatment* 1991 Jan;8(3):143–152. [doi: dg32sj]

50.

Mohr CD, Avera S, Kenny DA, Del Boca FK. "Getting by (or getting high) with a little help from my friends": An examination of adult alcoholics' friendships. *Journal of Studies on Alcohol* 2001 Jul;62(5):637–645. [doi: ghqvc3]

51.

Havassy BE, Hall SM, Wasserman DA. Social support and relapse: Commonalities among alcoholics, opiate users, and cigarette smokers. *Addictive Behaviors* 1991;16(5):235–246. PMID:1663695

52.

Hidalgo CA, Rodriguez-Sickert C. The dynamics of a mobile phone network. *Physica A: Statistical Mechanics and its Applications* 2008 May;387(12):3017–3024. [doi: bwdsbj]

53.

Sekara V, Stopczynski A, Lehmann S. Fundamental structures of dynamic social networks. *Proceedings of the National Academy of Sciences* 2016 Sep;113(36):9977–9982. [doi: f84d8f]

54.

Saramäki J, Moro E. From seconds to months: An overview of multi-scale dynamics of mobile telephone calls. *The European Physical Journal B* 2015 Jun;88(6):164. [doi: ghqvc5]

55.

Kossinets G. Empirical Analysis of an Evolving Social Network. *Science* 2006 Jan;311(5757):88–90. [doi: bwvwz8]

56.

Schönbrodt FD, Maier M, Heene M, Zehetleitner M. Voluntary commitment to research transparency. <http://www.researchtransparency.org>; 2015.

57.

Simmons JP, Nelson LD, Simonsohn U. A 21 Word Solution. <http://dx.doi.org/10.2139/ssrn.2160588>; 2012.

58.

R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2021.

59.

Team Rs. RStudio: Integrated Development for R. RStudio, PBC; 2020.

60.

Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Golemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H. Welcome to the Tidyverse. *Journal of Open Source Software* 2019 Nov;4(43):1686. [doi: 10.21105/joss.01686]

61.

Kuhn M, Wickham H. *Tidymodels: A collection of packages for modeling and machine learning using tidyverse principles*. 2020.

62.

Derogatis LR, Lipman RS, Covi L. SCL-90: An outpatient psychiatric rating scale—preliminary report. *Psychopharmacology Bulletin* 1973 Jan;9(1):13–28. PMID:4682398

63.

Freeston MH. Why do people worry? *Personal Individual Differences* 1994;17(6):791–802.

64.

Taylor S, Zvolensky MJ, Cox BJ, Deacon B, Heimberg RG, Ledley DR, Abramowitz JS, Holaway RM, Sandin B, Stewart SH, Coles M, Eng W, Daly ES, Arrindell WA, Bouvard M, Cardenas SJ. Robust dimensions of anxiety sensitivity: Development and initial validation of the anxiety sensitivity index-3. *Psychological Assessment* 2007;19(2):176–188.

65.

Simons JS, Gaher RM. The distress tolerance scale: Development and validation of a self-report measure. *Motivation and Emotion* 2005;29(2):83–102. [doi: 10.1007/s11031-005-7955-3]

66.

Miller IW, Epstein NB, Bishop DS, Keitner GI. The McMaster Family Assessment Device: Reliability and validity. *Journal of Marital and Family Therapy* 1985;11(4):345–356.

67.

Patrick CJ, Curtin JJ, Tellegen A. Development and validation of a brief form of the Multidimensional Personality Questionnaire. *Psychological Assessment* 2002 Jun;14(2):150–163. PMID:12056077

68.

Benavoli A, Corani G, Demsar J, Zaffalon M. Time for a Change: A Tutorial for Comparing Multiple Classifiers Through Bayesian Analysis. *Journal of Machine Learning Research* 2017;18:1–36.

69.

Klasnja P, Hekler EB, Shiffman S, Boruvka A, Almirall D, Tewari A, Murphy SA. Micro-Randomized Trials: An Experimental Design for Developing Just-in-Time Adaptive Interventions. *Health psychology : official journal of the Division of Health Psychology, American Psychological Association* 2015 Dec;34(0):1220–1228. PMID:26651463

Table 1

Demographics

	<i>n</i>	<i>%</i>	<i>M</i>	<i>SD</i>
Age			41	11.9
Sex				
Female	74	49.0		
Male	77	51.0		
Race				
American Indian/Alaska Native	3	2.0		
Asian	2	1.3		
Black/African American	8	5.3		
White/Caucasian	131	86.8		
Other/Multiracial	7	4.6		
Hispanic, Latino, or Spanish Origin				
Yes	4	2.6		
No	147	97.4		
Education				
Less than high school or GED degree	1	0.7		
High school or GED	14	9.3		
Some college	41	27.2		
2-Year degree	14	9.3		
College degree	58	38.4		
Advanced degree	23	15.2		
Employment				
Full-time	72	47.7		
Part-time	26	17.2		

Full-time student	7	4.6		
Homemaker	1	0.7		
Disabled	7	4.6		
Retired	8	5.3		
Unemployed	18	11.9		
Temporarily laid off, sick leave, or maternity leave	3	2.0		
Other, not otherwise specified	9	6.0		
Income			34298	31807
Marital Status				
Never Married	67	44.4		
Married	32	21.2		
Divorced	45	29.8		
Separated	5	3.3		
Widowed	2	1.3		

Note:

N = 151

Table 2

History of Lifetime Drug and Alcohol Use

	<i>n</i>	<i>%</i>	<i>M</i>	<i>SD</i>
AUD Milestones				
Age of first drink			14.6	2.9
Age of regular drinking			19.5	6.6
Age at which drinking became problematic			27.8	9.6
Age of first quit attempt			31.5	10.4
Number of Quit Attempts			8.5	30.7
Days (per week) Drinking 6 Mos Before Quit Date			5.2	1.8
Types of Treatment (Can choose more than 1)				
Long-term residential (6+ mos.)	8	5.3		
Short-term residential (< 6 mos.)	49	32.5		
Outpatient	74	49.0		
Individual counseling	97	64.2		
Group counseling	62	41.1		
Alcoholics Anonymous/Narcotics Anonymous	93	61.6		
Other	40	26.5		
Received Medication for AUD				
Yes	59	39.1		
No	92	60.9		
AUD DSM-5 Symptom Count			8.9	1.9
Lifetime Drug Use				
Tobacco products (cigarettes, chewing tobacco, cigars, etc.)	120	79.5		
Cannabis (marijuana, pot, grass, hash, etc.)	129	85.4		
Cocaine (coke, crack, etc.)	84	55.6		

Amphetamine type stimulants (speed, diet pills, ecstasy, etc.)	79	52.3		
Inhalants (nitrous, glue, petrol, paint thinner, etc.)	35	23.2		
Sedatives or sleeping pills (Valium, Serepax, Rohypnol, etc.)	70	46.4		
Hallucinogens (LSD, acid, mushrooms, PCP, Special K, etc.)	86	57.0		
Opioids (heroin, morphine, methadone, codeine, etc.)	64	42.4		
Lifetime Drugs Endorsed			4.4	2.5
Lifetime Drug Injection				
Yes	12	7.9		
No	139	92.1		

Note:

N = 151

Table 3

Mental Health Characterization

	<i>M</i>	<i>SD</i>
Symptom Checklist-90		
Total Score	0.0	0.7
Somatization Subscale	-0.1	0.7
Obsessive-compulsive Subscale	0.2	0.8
Interpersonal Sensibility Subscale	0.0	0.8
Depression Subscale	0.5	0.9
Anxiety Subscale	-0.1	0.8
Anger-hostility Subscale	-0.3	0.6
Phobic-anxiety Subscale	-0.5	0.8
Paranoid Ideation Subscale	-0.2	0.7
Psychoticism Subscale	-0.4	0.6
Intolerance of Uncertainty	70.2	22.0
Anxiety Sensitivity Index-3		
Total Score	7.7	16.2
Physical Concerns Subscale	1.3	6.3
Cognitive Concerns Subscale	1.3	6.8
Social Concerns Subscale	5.0	5.9
Distress Tolerance Scale	2.7	1.0
McMaster Family Assessment Device		
Problem Solving Subscale	16.5	3.2
Communications Subscale	24.3	4.7
Roles Subscale	30.8	5.0
Affective Responsiveness Subscale	16.3	3.5

Affective Involvement Subscale	19.3	4.0
Behavior Control Subscale	27.2	4.2
General Functioning Subscale	34.5	7.3
Multidimensional Personality Questionnaire Short Form		
Wellbeing Subscale	7.6	3.6
Social Potency Subscale	6.2	3.4
Achievement Subscale	7.1	3.1
Social Closeness Subscale	6.7	3.4
Stress Reaction Subscale	7.6	3.5
Alienation Subscale	3.5	3.4
Aggression Subscale	2.0	2.4
Control Subscale	7.2	2.9
Harm Avoidance Subscale	7.8	2.9
Traditionalism Subscale	5.3	2.9
Absorption Subscale	7.1	2.9
Unlikely Virtues Subscale	2.6	2.2

Note:

N = 151

Table 4

Balanced Accuracy for Best Configurations by Statistical Algorithm

algorithm	feature set	feature type(s)	class resample	balanced accuracy
random forest	passive	relative, raw	down	0.61
glmnet	active	raw	up	0.60
knn	passive	raw	down	0.58

Note. Table reports the best model configuration for each statistical algorithm in separate rows. Details about the model configuration are provided in subsequent columns. Balanced accuracy was estimated by grouped 10-fold cross validation during model selection.

Table 5

Confusion Matrix for Best Model Configuration

Prediction	Truth	
	no	yes
no	143196	5386
yes	52314	4904

Note. The best model configuration was a random forest algorithm using passive features and down-sampling. Results based on grouped 10x10-fold cross validation.

Table 6

*Performance Metrics for Best Model
Configuration*

metric	estimate
balanced accuracy	0.60
accuracy	0.72
sensitivity	0.46
specificity	0.73
ppv	0.09
npv	0.96
Area under the ROC Curve	0.64

Note. The best model configuration was a random forest algorithm using passive features and down-sampling.

Performance estimated by grouped 10x10-fold cross validation.