Machine learning-assisted treatment selection for smoking cessation

Gaylen E Fronk[aff-1] and John J. Curtin[aff-1]

[aff-1]Department of Psychology, University of Wisconsin-Madison

Abstract

Precision mental health seeks to select the right treatment for a patient given personal characteristics. The purpose of this project was to build a machine learning model that could select among first-line medication treatments for cigarette smoking. We used data from a previously completed comparative effectiveness trial in which participants were richly characterized at baseline before being randomly assigned to varenicline, combination nicotine replacement therapy, or nicotine patch. We built a model predicting treatment success (abstinent vs. smoking) using baseline characteristics and their interactions with treatment. Models were fit, selected, and evaluated using nested cross-validation and the performance metric area under the receiving operator characteristic curve (auROC). Our best models had a median auROC of 0.69 in held-out test sets. We used this model to calculcate probabilities of smoking cessation success for each participant on each of the three treatments to identify their model-predicted best treatment. Individuals who received their model-predicted best treatment during the original trial were more likely to quit successfully than individuals who did not (OR = 1.851, p = 0.004). This project produces a clinically implementable treatment selection model to assist people quitting cigarette smoking.

Keywords: Substance use disordersPrecision mental healthCigarette smoking-Machine learningTreatment selection

Introduction

Precision mental health seeks to guide treatment selection for mental health conditions using individual difference characteristics that are likely to predict treatment success for each patient (Bickman, 2020; Bickman et al., 2016; DeRubeis, 2019; Insel, 2014). Successful precision medicine would increase the likelihood of treatment success for each patient and improve treatment effectiveness rates across the population.

There is a critical need for improved mental health treatments. In the U.S. in 2022, 84 million individuals aged 12 or older had a substance use disorder, another mental health disorder, or both (Substance Abuse and Mental Health Services Administration, 2023). Mental health disorders are leading causes of disability and death (Centers for Disease Control and Prevention (CDC), n.d.; Whiteford et al., 2013) and account for enormous economic burden (Substance Abuse and Mental Health Services Administration (US) & Office of the Surgeon General (US), 2016; Trautmann et al., 2016). Additionally, mental healthcare is plagued by disparities related to race,

ethnicity, geographic region, and socioeconomic status: Vulnerable sub-populations are more likely to have higher rates of mental health and substance use disorders and more difficulty accessing treatment (Barksdale et al., 2022; Jacobson et al., 2022; Morales et al., 2020; Office of the Surgeon General (US) et al., 2001; Substance Abuse and Mental Health Services Administration, 2023). Treatments for mental health conditions are also usually no more than moderately effective, and many treatments for the same disorder can be quite comparable, making it difficult to select among them (Adjei & Ali, 2022; Lewis et al., 2020; Weisz et al., 2019).

Many researchers have pursued precision mental health (DeRubeis, 2019). Despite the abundance of research, much less progress has been made to personalize treatments for mental health disorders compared to treatments for medical disorders (Bickman, 2020; Bickman et al., 2016; Kessler & Luedtke, 2021; Kranzler et al., 2017; Oliver & McClernon, 2017).

One possible reason for this is that many factors influence heterogeneous, complex clinical phenomena like mental health diagnoses and treatment success (Feczko & Fair, 2020). Thus, any single feature (i.e., predictor variable) cannot account for more than a small portion of the variance in treatment success (Insel, 2014; Kessler & Luedtke, 2021). Precision mental health efforts so far, however, have largely focused on personalizing treatments using only a single factor (DeRubeis, 2019). It is perhaps unsurprising that models that consider only one or a small handful of features - which also limits considering concurrently features across categories - have failed to capture the real-world complexity underlying clinical phenomena like treatment success.

Additionally, because models are typically developed and evaluated in the same sample, the models may become very overfit to that sample (Jonathan et al., 2000). This problem is exacerbated in precision mental health because sample sizes in psychological research have remained relatively small despite recommendations to increase sample size (Marszalek et al., 2011). Consequently, precision mental health models do not generalize well to new patients.

To capture sufficient complexity to predict treatment success, we need to increase the total number of features in precision mental health models. Incorporating more features, however, makes overfitting the data more likely. Thus, successful precision mental health requires an analytic approach that can handle high-dimensional data without becoming too overfit to generalize to new patients.

Applying machine learning approaches

Machine learning may be able to advance precision mental health goals (Bickman et al., 2016; Dwyer et al., 2018; MacEachern & Forkert, 2021; Mooney & Pejaver, 2018). These models use statistical algorithms trained on high-dimensional arrays (hundreds or even thousands) of features (James et al., 2013; Kuhn & Johnson, 2018; Ng, 2018). Machine learning uses various techniques (e.g., regularization, hyperparameter tuning, simultaneous consideration of many model configurations) within resampling methods such as cross-validation to accommodate high-dimensional sets of features while reducing overfitting (James et al., 2013; Jonathan et al., 2000; Krstajic et al., 2014; Kuhn & Johnson, 2018; Mooney & Pejaver, 2018; Ng, 2018). Thus, we can build precision mental health models that capture complex clinical phenomena and generalize accurately to new data.

High-dimensional datasets and complex modeling procedures can make interpretation difficult (Cohen & DeRubeis, 2018; MacEachern & Forkert, 2021; Mooney & Pejaver, 2018). Fortunately, advances in interpretable machine learning (e.g., SHAP method for feature importance (Lundberg & Lee, 2017)) can help to counteract this concern. These techniques allow us to consider many features across categories while identifying which features contribute most to model performance.

Cigarette smoking as a critical precision mental health target

Cigarette smoking could benefit greatly from combining precision mental health and machine learning. Smoking remains an enormous public health burden. Tobacco is the leading cause of preventable death in the U.S., accounting for more than 480,000 deaths annually (Cornelius, 2020; National Center for Chronic Disease Prevention and Health Promotion (US) Office on Smoking and Health, 2014; Schlam & Baker, 2013). Individuals who smoke have two- to three-fold likelihood of death across causes and lose over a decade of life expectancy (Jha Prabhat et al., 2013).

Although rates of smoking have declined considerably, approximately 14% of U.S. adults continue to smoke daily or near-daily [(Cornelius, 2020); (Substance Abuse and Mental Health Services Administration, 2023). Additionally, cigarette smoking rates remain much higher in potentially vulnerable populations. This includes people with chronic or severe mental illness including other substance use disorders; Native American and non-Hispanic Black individuals; individuals who are economically and educationally disadvantaged; people in the criminal legal system; people experiencing homelessness; individuals who are uninsured or insured through Medicaid; and people who identify as lesbian, gay, or bisexual (Baggett et al., 2013; Baker & McCarthy, 2021; Cornelius, 2020; Cropsey et al., 2004; Harrison et al., 2020; Jamal et al., 2015; Kelly et al., 2012; Soar et al., 2020).

The best available smoking cessation treatments are modestly effective. The medications varenicline and combination nicotine replacement therapy (C-NRT) are the most effective options when combined with psychosocial counseling, yielding 6-month treatment success rates of 30-35% (Baker & McCarthy, 2021; Cahill et al., 2013; Fiore et al., 2008; Rigotti et al., 2022; Schlam & Baker, 2013). Varenicline and C-NRT appear to be equally effective (Baker et al., 2016; Cahill et al., 2013). Indeed, national clinical guidelines note that "there are no well-accepted algorithms to guide optimal selection" between any of the first-line medications ((Fiore et al., 2008), p. 44). These facts suggest a critical need for precision mental health approaches in the cigarette smoking domain.

Previous research

A large body of research has identified features that predict who will or will not be able to quit successfully (Etter et al., 2023; Issabakhsh et al., 2023; Kaufmann et al., 2015; Kaye et al., 2020; Lai et al., 2021; Megan E. Piper, Schlam, et al., 2017). A systematic review identified that features that predict treatment success span many categories: economic, environmental, sociodemographic, psychological, and physical health variables; engagement in treatment; biomarkers; neurocognitive factors; and smoking use, history, and severity characteristics (Bickel et al., 2023). Related research seeks to understand who will succeed using a single treatment (Coughlin et al., 2020; Massago et al., 2024).

This work is important for understanding the mechanisms that underlie treatment success. However, models that predict only overall treatment success do not offer an actionable way forward to select among treatment options. Even research that informs us as to who might succeed within a specific treatment has limited utility for treatment selection: what do we do for a patient who is not predicted to succeed using that treatment?

Some research has begun to build models that consider multiple treatments simultaneously to find factors that can allow selection among treatments. One line of work has investigated whether genetic factors and biomarkers may permit treatment selection. Perhaps most promising seem to be the nicotine metabolite ratio and specific variants in one gene (cholinergic receptor nicotinic alpha 5 subunit) (Chen et al., 2018; 2020; Chenoweth et al., 2016; Glatard et al., 2017; Lerman et al., 2015; Schnoll et al., 2009; Shahab et al., 2019; Siegel et al., 2020). Other work has investi-

gated whether non-biological features, such as those that rely only on self-report data, can guide treatment selection (Kaye et al., 2020; Piper et al., 2016; Megan E. Piper, Cook, et al., 2017). For example, one study found that psychiatric history moderates treatment success for some treatments but not others (Megan E. Piper, Schlam, et al., 2017).

The evidence remains somewhat limited that any of these biological, genetic, or behavioral factors can be used to guide treatment selection. This is perhaps unsurprising given that these studies examined only a single factor or examined each factor in a separate model, which is unlikely to explain sufficient variance in complex clinical outcomes. Using biological and genetic factors comes with additional downsides when considering accessibility and implementation, as the use of these factors is likely to favor privileged individuals and exacerbate existing disparities in mental healthcare (Jacobson et al., 2022; MacEachern & Forkert, 2021; Siegel et al., 2020)

Opportunities for treatment selection

Although there is not yet much research selecting among smoking cessation treatments, there may be reason to expect that some treatments may work better than others for a specific individual. First, there is enormous heterogeneity among people who smoke cigarettes (Oliver & McClernon, 2017; Wang et al., 2009; Zheng et al., 2013). This heterogeneity is typically neglected when selecting among treatments, but precision mental health approaches would instead take advantage of it.

Second, smoking cessation medications have distinct pharmacological mechanisms of action at nicotinic acetylcholine receptors (nAChRs), which may affect how helpful they are for different people. Nicotine replacement therapy (NRT) provides nicotine, a full agonist at nAChRs. Different NRTs provide nicotine differently. C-NRT consists of a nicotine patch and ad libitum nicotine lozenge use. The patch offers transdermal administration of a low, steady dose of nicotine to replace nicotine from cigarettes. Lozenges provide oral nicotine with more rapid onset, which could provide a quick boost during craving. Other individuals who smoke may benefit from a medication like varenicline, a partial agonist at nAChRs (Cahill et al., 2016). Partial agonists have a pharmacological action that depends on the level of surrounding neurotransmitter. In the absence of a full agonist, partial agonists can act as a functional agonist with lower activity than a full agonist. In the presence of a full agonist (e.g., a cigarette), they act as functional antagonists because their binding to the receptor limits the amount of binding from the full agonist and consequently reduces that response (Jordan & Xi, 2018; Lieberman, 2004). Thus, varenicline may be more pharmacologically flexible than NRT medications: When an individual is not smoking, it can produce milder, nicotine-like effects; if an individual begins smoking again, it could block or reduce full agonist (nicotine from cigarettes) activity at the receptor. This would be expected to reduce the pharmacological effect of nicotine, likely reducing the behavioral pleasure of smoking (Cahill et al., 2016).

Third, features across several behavioral or environmental domains may also guide treatment selection, alone or in combination with medication mechanisms of action. Research has shown there are *many* factors that predict smoking cessation overall, and these factors span *many* clinical and behavioral domains (Bickel et al., 2023). It is possible that some of these factors may guide treatmment selection as well, but this has not yet been tested. Moreover, these many factors have not been considered simultaneously in a model as may be needed to unpack complex clinical phenomena like differential treatment success.

Finally, there is some evidence that individuals respond differently to different treatments. Individuals who switch medications are more likely to quit than individuals who are re-treated with the same medication (Fiore et al., 2008; Gonzales et al., 2014; Heckman et al., 2017; Tønnesen et al., 1993). Treatment adherence improves when individuals are given the opportunity to sample various NRT medications pre-quit (Cropsey et al., 2017), and meta-analyses show only smokers who are highly dependent may benefit from 4 mg (vs. 2 mg) nicotine gum (Lindson et al., 2019). These data

suggest differential preferences and even differential effectiveness despite a shared pharmacological mechanism of action across NRT medications. Additionally, clinical research has demonstrated differential treatment success on an individual basis for other mental health disorders (e.g., antipsychotic medications for schizophrenia (Roussidis et al., 2013); psychosocial interventions for depression (Cohen & DeRubeis, 2018)), suggesting it is worth investigating whether the same is true in smoking cessation.

Purpose

The goal of this project was to produce and evaluate a treatment selection model to select among medications for smoking cessation. Specifically, we built a model to predict treatment success using rich baseline characteristics and treatment assignments from a previously completed comparative effectiveness trial of varenicline, C-NRT, and nicotine patch (Baker et al., 2016). Using our model, we identified the best treatment for each person and evaluated whether individuals who received their best treatment were more likely to be abstinent in the original trial at 4, 12, and 26 weeks post-quit. Throughout the project, we used rigorous resampling techniques to ensure that we evaluated our model's capacity to predict smoking cessation outcomes and select treatments in *new* patients. Additionally, we incorporated easy-to-collect self-report data as model inputs, and we employed a statistical algorithm that can reduce assessment requirements. These choices position our model for accessible clinical implementation.

Methods

Transparency & openness

We adhere to research transparency principles that are crucial for robust and replicable science. We reported how we determined the sample size, all data exclusions, all manipulations, and all study measures. We provide a transparency report in the supplement. Finally, our data and other study materials are publicly available on our OSF page, and our annotated analysis scripts and results are publicly available on our study website.

Data

The data for this project came from a completed randomized controlled trial conducted by the University of Wisconsin (UW) Center for Tobacco Research and Intervention (CTRI) (Baker et al., 2016). This trial compared the effectiveness of three cigarette smoking cessation treatments (varenicline, combination nicotine replacement therapy [C-NRT], and nicotine patch). Briefly, 1086 daily cigarette smokers looking to quit smoking were enrolled in Madison, WI, USA and Milwaukee, WI, USA. Exclusion criteria included contraindicated medical (e.g., severe hypertension) or psychiatric (e.g., severe and persistent mental illness) conditions; current use of contraindicated medications; and pregnancy or unwillingness to use appropriate methods of contraception while taking a study medication. Participants set a quit date with study staff and were enrolled for several weeks prior to the target quit date through at least 6 months following quitting smoking.

Treatment conditions

Participants were randomly assigned to one of three medication conditions: varenicline, C-NRT, or nicotine patch. Each medication treatment lasted 12 weeks. For varenicline, participants began medication use prior to their quit attempt, starting with 0.5 mg once daily for 3 days, followed by 0.5 mg twice daily for 4 days, and 1 mg twice daily for 3 days. They continued use of 1 mg twice daily for 11 weeks following their quit date except in response to adverse effects. For C-NRT or

nicotine patch, participants began using the patch on their quit date, starting with 21 mg for 8 weeks, followed by 14 mg for 2 weeks, and 7 mg for 2 weeks. All individuals who received C-NRT were also instructed to use 5 lozenges per day (2 or 4 mg nicotine lozenges determined by time to first daily cigarette) for the full 12 weeks except in the case of adverse effects.

All participants also received 6 sessions of motivational and skill-training counseling per clinical guidelines (Fiore et al., 2008).

Individual difference characteristics

Participants were comprehensively assessed for individual difference characteristics prior to treatment randomization. These characteristics fall into several domains expected to relate to cigarette smoking cessation: tobacco-related (e.g., cigarettes per day), psychological (e.g., psychiatric diagnoses, distress tolerance), physical health (e.g., vital signs), social/environmental (e.g., living with another person who smokes), and demographic (e.g., age, sex). A detailed list of all available individual differences variables appears in Table 1.

Table 1: Individual Differences Characteristics Available for Model Features

Demographic Characteristics	Feature Name	Type	# Items
	Gender	Categorical (unordered)	1
	Age	Numeric	1
	Race	Categorical (unordered)	1
	Marital Status	Categorical (unordered)	1
	Income	Categorical (ordered)	1
	Ethnicity	Categorical (unordered)	1
	Employment	Categorical (unordered)	1

Table 1: Individual Differences (Continued)

Smoking Use/History	Feature Name	Type	# Items
	Baseline Carbon Monoxide	Numeric	1
	Carbon Monoxide Exposure	Categorical (unordered)	1
	Age of 1st Cigarette	Numeric	1
	Age Became Daily Smoker	Numeric	1
	Years Smoking	Numeric	1
	Cigarettes Per Day (Heaviest)	Numeric	1
	Use of Other Tobacco Products	Categorical (unordered)	5
	Number of Previous Quit Attempts	Numeric	1
	Last Recent Quit Attempt	Categorical (ordered)	1
	Longest Quit Attempt	Categorical (ordered)	1
	Previous Quit Methods Used	Categorical (unordered)	6
	Cigarettes Per Day (Current)	Numeric	1
	Motivation to Quit	Categorical (ordered)	1
	Self-Efficacy for Quitting in Next 30 Days	Categorical (ordered)	1
	Confidence to Quit	Categorical (ordered)	1
	Importance to Quit	Categorical (ordered)	1
	DSM5 Tobacco Use Disorder (American Psychiatric Association, 2013)	Categorical (unordered)	13
	Fagerstrom Test of Nicotine Dependence (Heatherton et al., 1991)	Categorical (unordered); Categorical (ordered)	6
	Wisconsin Inventory of Smoking Dependence Motives-37 (Smith et al., 2010)	Categorical (ordered)	37
	Smoke Menthol Cigarettes	Categorical (unordered)	1
	Wisconsin Smoking Withdrawal Scale-2 (Smith et al., 2021)	Categorical (ordered)	38

Table 1: Individual Differences (Continued)

Social & Environmental Characteristics	Feature Name	Type	# Items
	Spouse Smokes	Categorical (unordered)	1
	Live with Another Smoker	Categorical (unordered)	1
	People Close to You Who Smoke	Categorical (unordered)	5
	Ban on Smoking at Home	Categorical (unordered)	1
	Ban on Smoking at Work	Categorical (unordered)	1
	Time Around Other Smokers	Categorical (ordered)	2

Mental Health & Psychological Traits	Feature Name	Type	# Items
	Frequency of Drinking Alcohol	Categorical (ordered)	1
	Quantity of Alcohol	Categorical (ordered)	1
	Binge Drinking	Categorical (ordered)	1
	Short Inventory of Problems-2 (Revised) (Kiluk et al., 2013)	Categorical (ordered)	15
	Life Satisfaction	Categorical (ordered)	1
	Life Enjoyment	Categorical (ordered)	1
	Psychological Disorder Diagnoses	Categorical (unordered)	7
	Positive and Negative Affect Schedule (Crawford & Henry, 2004)	Categorical (ordered)	6
	Snaith-Hamilton Pleasure Scale (Snaith et al., 1995)	Categorical (ordered)	14
	Anxiety Sensitivity Index-3 (Taylor et al., 2007)	Categorical (ordered)	18
	Distress Tolerance Scale (Simons & Gaher, 2005)	Categorical (ordered)	15
	Patient History Questionnaire-9 (Kroenke et al., 2001)	Categorical (ordered)	9

Table 1: Individual Differences (Continued)

Medical & Physical Health	Feature Name	Type	# Items
	Diabetes Diagnosis	Categorical (unordered)	1
	Multidimensional Fatigue Inventory (Smets et al., 1995)	Categorical (ordered)	20
	Berlin Sleep Questionnaire (Netzer et al., 1999)	Categorical (ordered), Categorical (unordered)	3
	Body Mass Index	Numeric	1
	Health-related Quality of Life Scale (Taylor & National Center for Chronic Disease Prevention and Health Promotion (U.S.). Division of Adult and Community Health, 2000)	Categorical (ordered), Numeric	4
	Healthy Days Symptoms Module (Moriarty, 1996)	Numeric	5

Miscellaneous Features	Feature Name	Type	# Items
	Treatment	Categorical (unordered)	1
	Incarcerated	Categorical (unordered)	1
	Works Third Shift	Categorical (unordered)	1

Treatment success outcome

Throughout study participation, treatment success was assessed via biologically confirmed, 7-day point-prevalence abstinence. Point-prevalence abstinence assesses for any smoking (here, in the previous 7 days) and yields a single, dichotomous outcome of "abstinent" or "smoking." Participants self-reported whether they had smoked over the past 7 days, and their report was biologically confirmed via exhaled carbon monoxide (CO). Participants were labeled as "abstinent" if their CO level was less than 6 parts per million (ppm; (Baker et al., 2016)). If participants self-reported any smoking in the past 7 days, their CO level contradicted their self-report (i.e., CO level $>$ 6 ppm), or biological confirmation could not be confirmed, participants were labeled as "smoking."

Assessments of treatment success were used in two separate ways. During model building, treatment success at 4 weeks post-quit served as the *prediction outcome for our models* (i.e., predicting if individuals were labeled "abstinent" or "smoking" at 4 weeks). We selected this prediction outcome to evaluate the differential effects of treatment while participants were actively using the medications.[1]

[1]We fit additional models predicting treatment success at 12 and 26 weeks, and these models had significantly worse performance than our 4 week prediction model, supporting our choice to use the 4 week model (See Supplement).

Treatment success also served as the outcome for *clinical benefit analyses*. We evaluated whether using our treatment selection model yielded higher treatment success (i.e., higher rates of abstinence) at 4, 12, and 26 weeks post-quit.

Model building

Data preprocessing, modeling, and Bayesian analyses were done in R using the tidymodels ecosystem (Kuhn & Wickham, 2020). Models were trained and evaluated using high-throughput computing resources provided by the University of Wisconsin Center for High Throughput Computing (Center for High Throughput Computing, 2006). Additional details about model building and evaluation appear in the Supplement.

Feature engineering

Generic feature engineering steps included: 1) imputing missing data (median imputation for numeric data, mode imputation for categorical [ordered and unordered] data); 2) removing zero-variance features; 3) using a Yeo-Johnson transformation of all numeric features to normalize distributions; 4) one-hot-coding of unordered categorical data; and 5) standardizing all features to have a mean of 0 and standard deviation of 1 (required by statisical algorithm). Medians/modes for missing data imputation, identification of zero variance features, and means/standard deviations for normalization and standardization were derived from held-in (training) data and applied to held-out (validation and test) data (see Cross-validation section below).

Treatment was also one-hot-coded such that there were three features, one corresponding to each treatment (varenicline, C-NRT, nicotine patch). The three treatment features were allowed to interact with all other features to permit differential prediction. A sample feature engineering script (i.e., tidymodels recipe) containing all feature engineering steps is available on our OSF study page (https://osf.io/qad4n/).

Model configurations

All model configurations used the statistical algorithm Elastic Net Logistic Regression (GLMNet) and differed across sensible values for the associated hyperparameters, alpha and lambda. This algorithm aligned with our application goals: GLMNet allows including interation terms explicitly (i.e., interactions between treatment and all other features), and it reduces dimensionality by penalizing model complexity. Model configurations differed by 1) whether feature sets included individual items within a self-report measure or total scale/sub-scale scores derived from items, and 2) whether ordered categorical data (e.g., Likert scale items) were ordinal scored or one-hot coded.

Cross-validation

We used nested k-fold cross-validation for model training, selection, and evaluation (Krstajic et al., 2014). Nested cross-validation uses two nested loops for dividing and holding out folds: an outer loop, where held-out folds serve as *test sets* for model evaluation; and inner loops, where held-out folds serve as *validation sets* for model selection. We used 1 repeat of 10-fold cross-validation (i.e., $k = 10$) for the inner loops and 3 repeats of 10-fold cross-validation for the outer loop.

Models were selected and evaluated using auROC, which indexes the probability that the model will predict a higher score for a randomly selected positive case (abstinent) relative to a randomly selected negative case (smoking) (Kuhn & Johnson, 2018; Youngstrom, 2014). Because it was important that our model retained interaction features, we defined a "satisficing metric" of 50 or more treatment interaction features retained (Ng, 2018). Best model configurations were selected using median auROC among model configurations that retained a median of 50 treatment

interaction features across the 10 *validation sets*. Final performance evaluation of those best model configurations used median auROC across the 30 *test sets*.

We also used leave-one-out cross-validation (LOOCV) for feature importance and clinical benefit evaluation (see below). In LOOCV, the value of k is set to N (sample size) such that each test set consists of a single held-out participant. We fit $N = 1086$ models, where each participant served as the test set once for a model fit with the other 1085 participants.

Model evaluation

We used a Bayesian hierarchical generalized linear model to estimate the posterior probability distributions and 95% Bayesian credible intervals (CIs) for test set auROC. We estimated the posterior probability distribution around model performance following recommendations from the tidymodels team (Kuhn, 2022). The median posterior probability for auROC represents our best estimate for the magnitude of the auROC parameter for each model. If the credible intervals do not contain .5 (chance performance), this suggests our model is capturing signal in the data.

Fairness benchmarks

We aim for the responsible and transparent reporting of model performance and clinical utility, and we acknowledge that models that work for only a subset of people, if implemented, could widen existing treatment disparities. Consequently, we calculated median posterior probability and 95% Bayesian CIs for auROC separately for three dichotomized demographic groups with potential disparities in access to substance use treatment - race and ethnicity (non-Hispanic White vs. non-White) (Kilaru et al., 2020; Pinedo, 2019), income (above poverty vs. below poverty) (Olfson et al., 2022), and sex at birth (male vs. female) (Greenfield et al., 2007; Kilaru et al., 2020). We conducted Bayesian model comparisons to assess the likelihood that the model performs differently by group. We report median differences in posterior probabilities and the 95% Bayesian CIs around these differences.

Model interpretation

Interpretability is important when using machine learning for clinical applications to identify important features and encourage implementation (Cohen & DeRubeis, 2018; MacEachern & Forkert, 2021; Mooney & Pejaver, 2018). We reviewed the retained features and their parameter estimates as a metric of feature importance because GLMNet only retains features whose contribution outweighs the additional complexity.

We also computed Shapley Values (Lundberg & Lee, 2017) to provide a consistent, objective explanation of feature importance. We calculated Shapley values using LOOCV with the final, best selected model configuration. We used the DALEX and DALEXtra packages (Biecek, 2018) in R, which provide local Shapley values (i.e., for each observation) in log-odds units for binary classification models. To index global feature importance, we averaged the absolute value of the local Shapley values of each feature across observations. Shapley values are additive, which allowed us to create two feature categories, Main Effects and Interactions. We calculated global importance of each feature category by averaging the absolute value of the Shapley values of all features in the category across observations. These global importance scores allowed us to contextualize relative feature importance for each feature and feature category.

Clinical benefit evaluation

Select final model configuration

To get a single, final model, we replicated our inner loop resampling (1 repeat of 10-fold cross-validation) on the full dataset. The best model configuration was selected using median auROC across the 10 held-out folds. A final model was fit on the full dataset using this best selected model configuration to obtain parameter estimates for interpretation.

Identify model-predicted best treatment

Using LOOCV, we made predictions for each held-out participant to match real-world implementation (i.e., making predictions for a new patient). We calculated three predicted probabilities for each participant by substituting each treatment into the model inputs. This produced one prediction per person per treatment. The treatment that yielded the highest model-predicted probability of abstinence was identified as that participant's "best" treatment.

For example, an individual received varenicline in the original trial. We calculated their probability of abstinence using their data (i.e., with varenicline as the "treatment" feature), and we calculated two additional probabilities by substituting C-NRT and nicotine patch for varenicline. Their probability of abstinence was highest when substituting C-NRT, meaning C-NRT was identified as their best treatment.

Categorize treatment matching

Some participants' best treatment matched what they were randomly assigned in the original trial. Other participants may have received a sub-optimal treatment (i.e., what the model identified as their second-best or worst treatment based on calculated probabilities). Thus, participants could be categorized by whether they received their best treatment or another treatment in the original trial.

For example, the individual described above received varenicline in the original trial, but their model-predicted probability of abstinence was highest when substituting C-NRT for treatment. This participant's best treatment did not match their trial treatment, so they were labeled as matched to "other treatment."

Evaluate clinical benefit

Our primary analysis evaluated the clinical benefit of our treatment selection model by comparing the observed outcomes (i.e., abstinence vs. smoking from the original trial) for people who were matched to their best treatment or to another treatment (between-subjects; 0.5 ["best treatment"] vs. −0.5 ["other treatment"]).

We examined this effect over time at 4, 12, and 26 weeks. Calculating and interpreting interactions in logistic models is not straightforward because significance can differ based on the link function used (Collins, 2018; Karaca-Mandic et al., 2012). Consequently, we conducted three, separate generalized logistic regression models where we regressed the outcome (abstinent vs. smoking) on treatment matching. We conducted these regressions using the lme4 package in R (Bates et al., 2015). We also conducted sensitivity analyses using only participants who reported starting their medication (i.e., any medication use reported during the first 4 weeks post-quit; N = 988).

Results

Sample characteristics

All 1086 participants who were randomized to treatment in the comparative effectiveness trial (Baker et al., 2016) were included in our analysis sample.

Demographic characteristics of this sample appear in Table 2. Our sample was 52.12% (N = 566) female. We had good representation of White (67.03%, N = 728) and Black (28.45%, N = 309) individuals but poor representation among Asian (0.28%, N = 3), Multiracial (2.03%, N = 22), Native American/Alaska Native (0.55%, N = 6), and Other individuals (1.66%, N = 18). This was also a primarily non-Hispanic sample (97.51%, N = 1059). There was wide variety with respect to marital status, employment status, and income, including over one third of the sample reporting annual income lower than $25,000 (national poverty line at the time of data collection).

Table 2: Demographic Characteristics

Characteristic	N (%)	Mean (SD)
Age		48.13 (11.6)
Gender		
Female	566 (52.12%)	
Male	520 (47.88%)	
Race		
Asian	3 (0.28%)	
Black/African American	309 (28.45%)	
Multiracial	22 (2.03%)	
Native American/Alaska Native	6 (0.55%)	
Other	18 (1.66%)	
White	728 (67.03%)	
Ethnicity		
Hispanic or Latino/a	27 (2.49%)	
Non-Hispanic	1059 (97.51%)	
Marital Status		
Divorced	224 (20.63%)	
Living with a domestic partner	87 (8.01%)	
Married	384 (35.36%)	
Never married	299 (27.53%)	
Separated	53 (4.88%)	
Widowed	34 (3.13%)	
Did not respond	5 (0.46%)	
Employment		
Employed (full-time)	474 (43.65%)	
Employed (part-time)	283 (26.06%)	
Unemployed	329 (30.29%)	

Table 2: Demographic Characteristics (Continued)

Characteristic	N (%)	Mean (SD)
Income		
< $10,000	210 (19.34%)	
$10,000 - $19,999	135 (12.43%)	
$20,000 - $24,999	83 (7.64%)	
$25,000 - $34,999	129 (11.88%)	
$35,000 - $49,999	149 (13.72%)	
$50,000 - $74,999	160 (14.73%)	
$75,000+	167 (15.38%)	
Did not respond	53 (4.88%)	

Smoking-related characteristics of this sample appear in Table 3. On average, individuals had been smoking almost 30 years and were currently smoking 17 cigarettes per day. Over three quarters of the sample reported smoking their first cigarette of the day within 30 minutes of waking up.

Table 3: Smoking Use & History Characteristics

Characteristic	Mean (SD)	N (%)
Age of first cigarette	14.6 (3.42)	
Age became daily smoker	17.56 (4.62)	
Years smoking	28.65 (12.03)	
Cigarettes per day (heaviest)	22.74 (10.67)	
Number of previous quit attempts	3.91 (6.03)	
Cigarettes per day (current)	17.03 (8.31)	
Number of DSM5 tobacco use disorder symptoms	4.57 (2.17)	
WISDM37 Total Score	3.99 (1.17)	
WSWS Total Score	48.82 (19.91)	
Time to first cigarette upon waking		
After 60 minutes		82 (7.55%)
31 - 60 minutes		157 (14.46%)
6 - 30 minutes		477 (43.92%)
Within 5 minutes		366 (33.7%)
Did not respond		4 (0.37%)

Model performance

The median auROC across the 30 test sets for the 4-week model was 0.695 (IQR = 0.667 - 0.718, range = 0.592 - 0.788). Figure 1, Panel A shows the ROC curve for held-out test set performance (concatenated across 30 held-out folds).

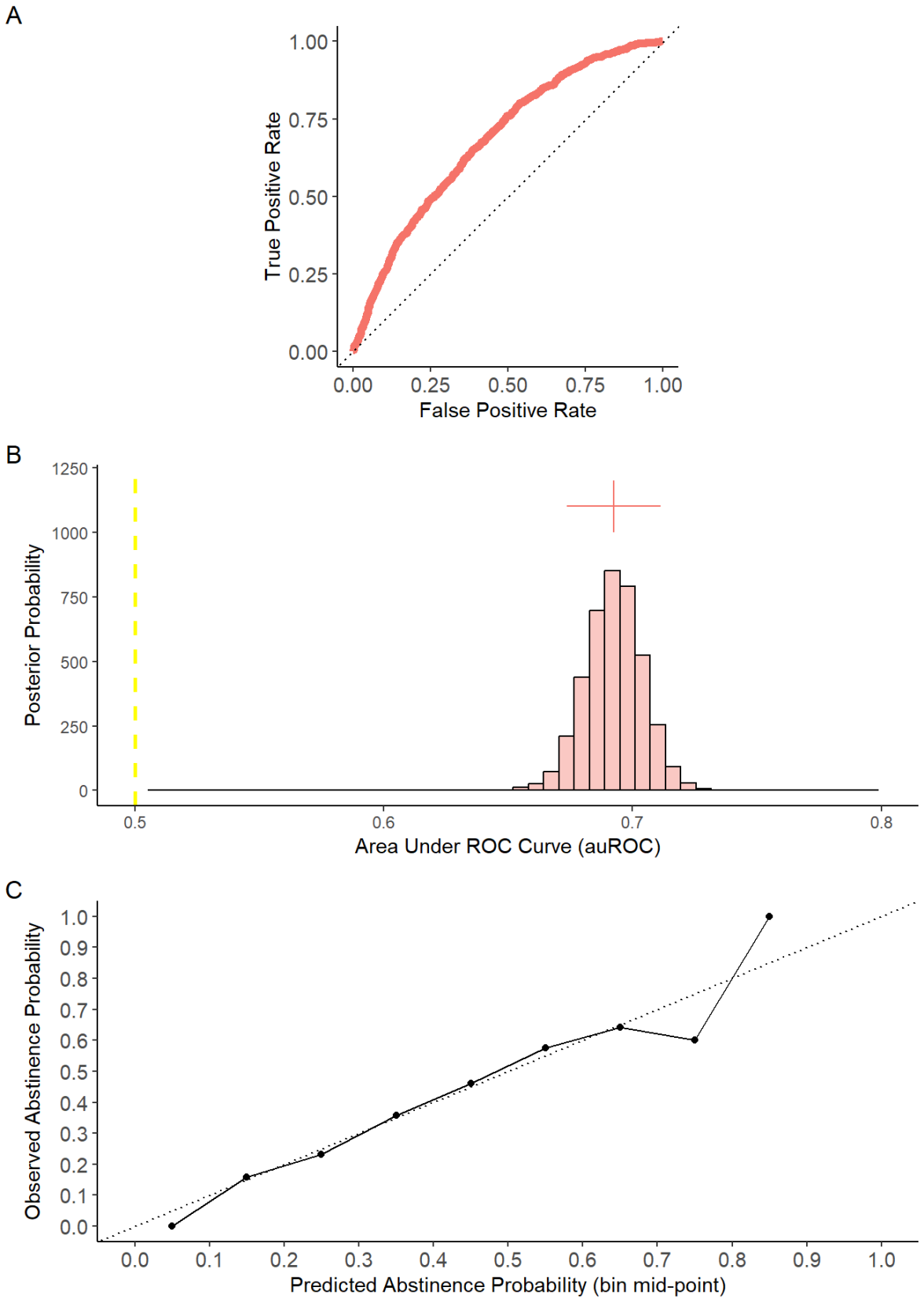

Figure 1: Model performance for prediction model. A) ROC curve plotted across all values of sensitivity (true positive rate) and specificity (1 - false positive rate). Dotted line indicates chance performance. B) Posterior probability distribution for the median auROC in test sets. Histogram represents posterior probability distribution. Horizontal line displays 95% Bayesian credible interval. C) Model calibration between predicted probabilities and observed values. Predicted probabil-

We used the 30 test set auROCs to estimate the posterior probability distribution for the auROC of these models. The median auROC from the posterior distribution was 0.693. This value represents our best estimate for the magnitude of the auROC parameter. The 95% Bayesian CI for the auROC was relatively narrow [0.674 - 0.711] and did not contain 0.5 (chance performance), suggesting this model has predictive signal. Figure 1, Panel B displays the posterior probability distribution for the auROC.

In Figure 1, panel C, we display our model's calibration. This figure plots the predicted probabilities for each individual for their original trial-assigned treatment against the observed trial treatment success. Predicted probabilities are binned (bin width = 10%) and plotted against the observed probability of abstinence for observations in that bin. If probabilities were perfectly calibrated, all bin means would fall on the dotted line (e.g., bin from 0 - 10 with an observed mean probability of 0.05, bin from 10 - 20 with an observed mean probability of 0.15). Probabilities are well calibrated and ordinal in their relationship with the true probability of abstinence.

Fairness benchmarks

All group comparisons for race/ethnicity, sex at birth, and income were significant (probability > 0.95). There was a median increase in auROC of 0.048 (95% CI 0.013 - 0.084) for non-White participants (vs. non-Hispanic White participants), indicating better performance for non-White participants. There was a median increase in auROC of 0.033 (95% CI 0.004 - 0.062) for participants below (vs. above) the national poverty line, indicating better performance for participants with lower income.[2] There was a median increase in auROC of 0.056 (95% CI 0.026 - 0.087) for males (vs. females), indicating better performance for male participants.

Model interpretation

Retained features

Our final model fit with the full dataset retained 155 features overall including 74 treatment interactions. Retained treatment interaction features appear in Table 4, grouped by feature category from Table 1.

[2]Fifty-three participants (4.9%) did not provide income data and were excluded from this analyses.

Table 4: Retained Interaction Features. Effect direction indicates whether increasing values of that feature (or coded positive for that level for one-hot-coded categorical features) increased or decreased the probability of treatment success when using that specific treatment (vs. the other two treatments).

Category	Feature	Effect Direction (Varenicline)	Effect Direction (C-NRT)	Effect Direction (Patch)
Demographic				
	Divorced		-	
	Greater income	+		
	Has never been married			+
	Identifies as Black or African American		-	
	Identifies as White		+	+
	Identifies as female			-
	Identifies as male			+
	Married		+	
Medical				
	Berlin: Feels tired, fatigued, or not up to par	-		
	Berlin: No sleep apnea diagnosis		+	
	HDSM: More days in past month feeling worried, tense, or anxious		-	
	MFI: Does not feel it takes a lot of effort to concentrate		+	
	MFI: Does not have a lot of plans			+
	MFI: Does not think they do a lot in a day	+		
Psychological				
	ASI-3: Worries they are going		-	
Social/Environ	Use/	+	+	+

To perform treatment selection, only interactive features need to be assessed, as features that increase or decrease probabilities equally across treatments do not help with differential prediction. Implementing this model for treatment selection requires assessing only 52 unique items (e.g., multiple dummy variables are from a single item, the same feature interacts with more than one treatment).

Feature importance via Shapley values

Global feature importance (mean |Shapley value|) from our model appears in Figure 2, both overall (Panel A) and for treatment interaction features specifically (Panel B). Shapley values describe the relative importance of these individual features for making predictions. Six of the top 25 most globally important features were treatment interactions.

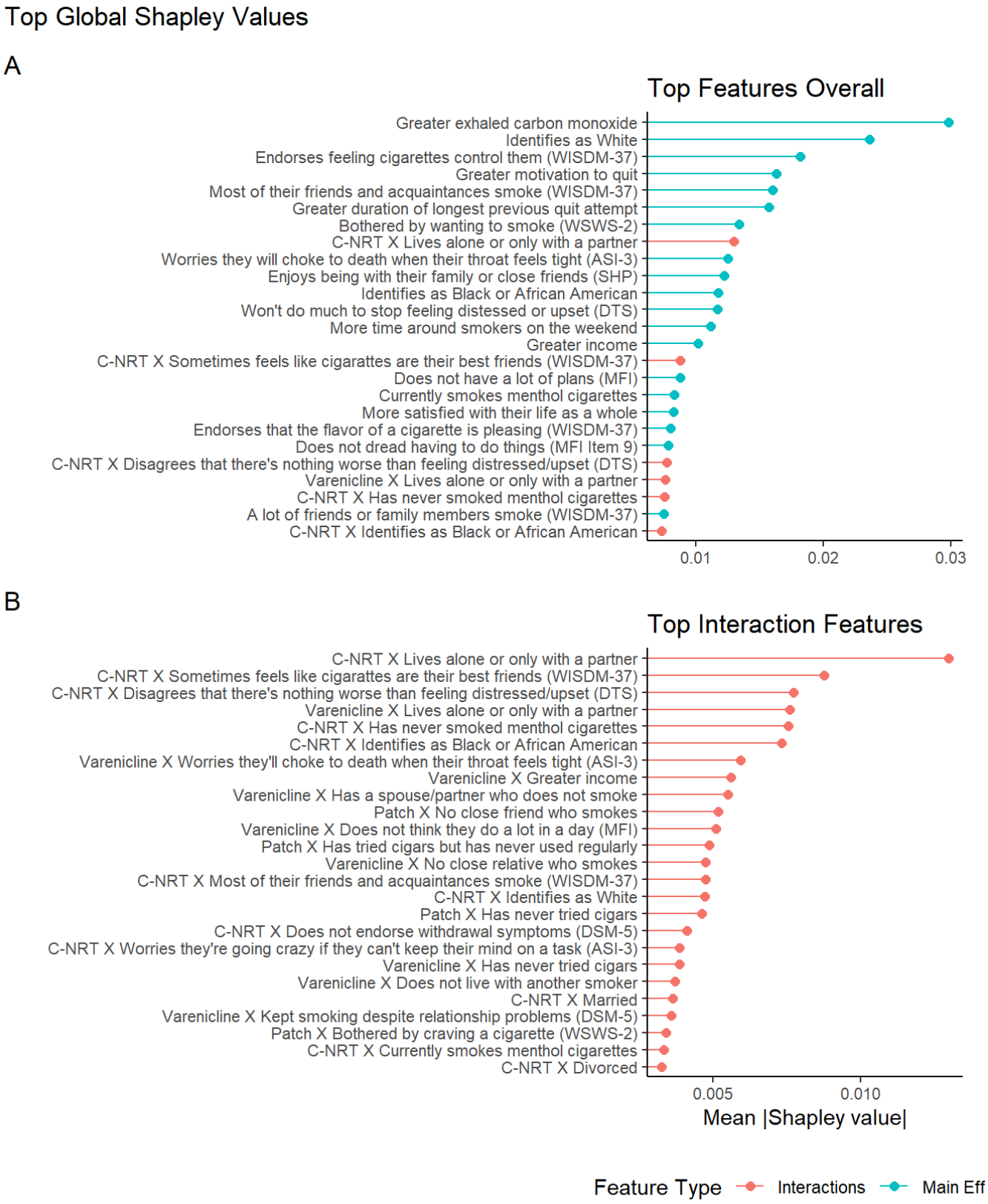

Figure 2: Global feature importance via Shapley values. Bar represents magnitude of global feature importance, which is calculated from the mean of the absolute value across all observations for that feature. A) Global feature importance for top 25 features overall. B) Global feature importance for top 25 treatment interaction features.

Global feature importance grouped by feature type (main effect or interaction) appear in Figure 3. Main effect features were relatively more important than treatment interaction features for predicting overall treatment success.

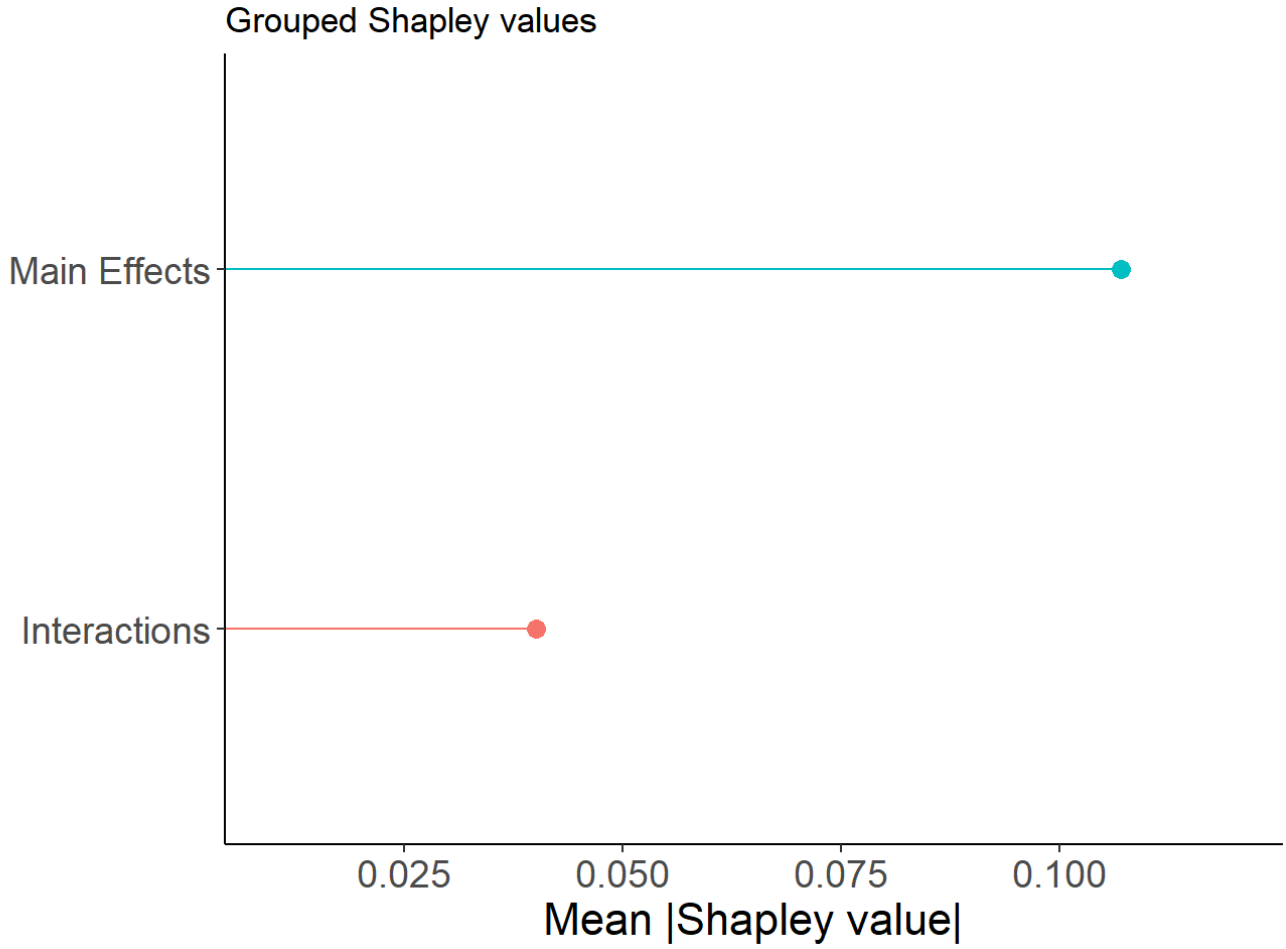

Figure 3: Global feature importance via Shapley values grouped by feature type (main effect or interaction feature)

Clinical benefit

There was a significant fixed effect of treatment matching on abstinence at 4 weeks (OR = 1.382, $z = 2.452$, $p = 0.014$) such that individuals who received their model-predicted best treatment were more likely to be abstinent. The effect of treatment matching was no longer significant at 12 weeks ($p = 0.232$) or at the 26-week follow-up assessment ($p = 0.943$). Figure 4 shows the mean abstinence rate by treatment matching at each time point. The pattern of results was identical when including only individuals who reported any medication use in the analysis sample.

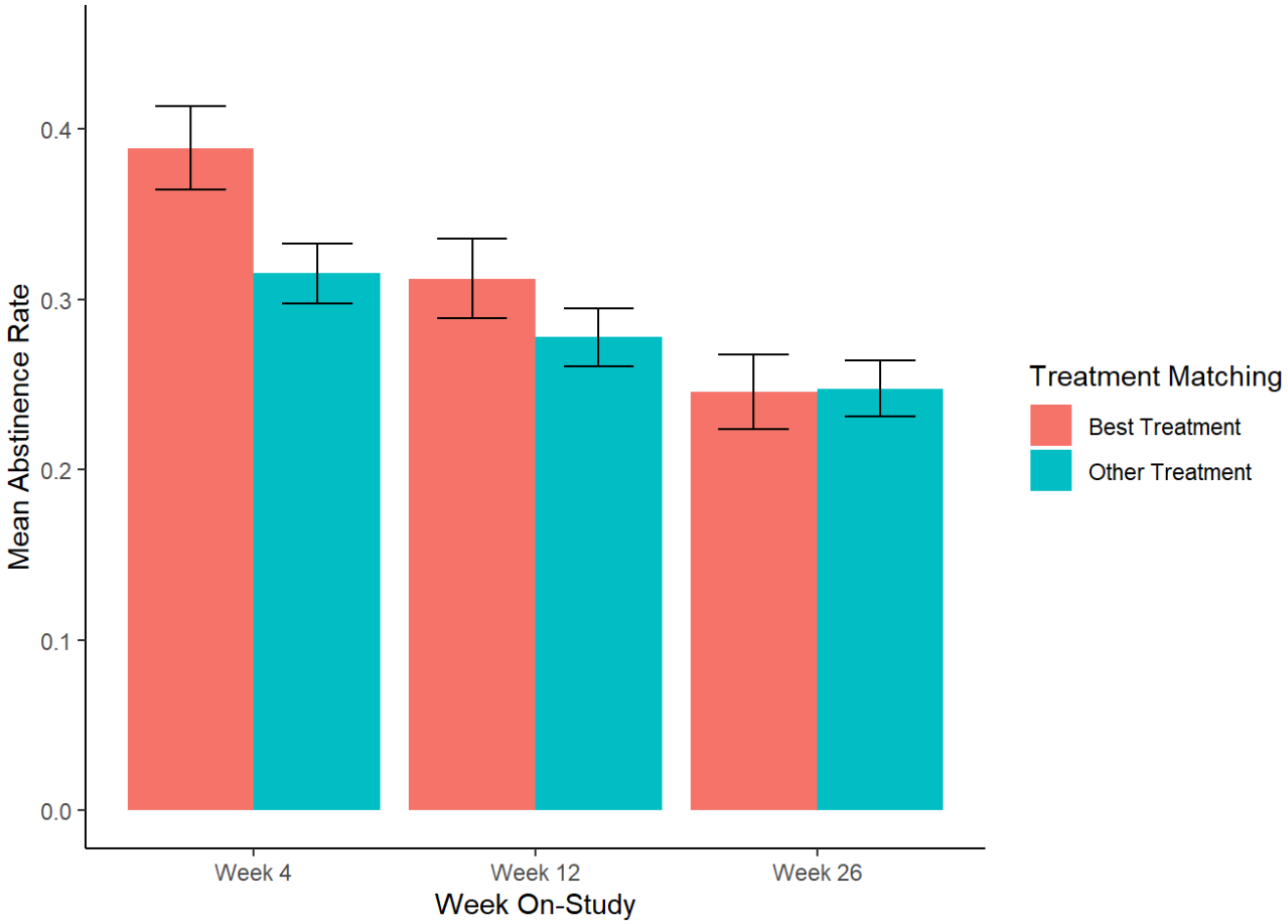

Figure 4: Benefit of treatment matching. Bars represent mean observed abstinence (from original trial) for individuals who did and did not receive their model-predicted best treatment, over time. Error bars indicate standard errors.

Discussion

In this project, we produced a treatment selection model that can offer immediate benefit to individuals looking to quit smoking using several first-line medications. Individuals who received their model-predicted best treatment in the original trial had a mean abstinence rate that was 7.4% higher than individuals who did not (38.9% vs. 31.5%) at 4 weeks. Although this absolute difference may seem somewhat small, this corresponds to a 23.5% *relative* improvement - simply by allocating treatments to the right person.

We feel confident in this effect for several reasons. First, we made predictions for each individual to identify their best treatment while they were held-out from model fitting to match how this model will be used in clinical practice (i.e., to make predictions and select a treatment for new patients). Second, our model is capturing predictive signal, as supported by the Bayesian CI around our model's auROC, adding to our trust in the model's prediction outputs. Third, these predictions were well-calibrated such that we can trust their ordinal ranking. This is critical because what is used in our treatment selection process is the relative *order* (i.e., rank) of predicted probabilities for each person rather than the values themselves.

We can achieve this benefit using an accessible, low-burden assessment. Implementing this treatment selection model would require assessing approximately 50 multiple choice and yes/no questions, which survey research suggests would take 11-12 minutes to complete (Lenzner et al., 2010). Additionally, because all items are self-report questions, this assessment can be completed remotely (e.g., administered online) and can be made available to people without access to in-person medical care. This assessment tool is particularly valuable in this context because two treatments in the model (C-NRT, nicotine patch) are widely available over-the-counter, offering scalable implementation when healthcare access is limited.

This focus on accessibility is espeically important given disparities in mental healthcare. Access to treatment is a known barrier in mental healthcare and a contributing factor driving healthcare disparities (Jacobson et al., 2022). Cigarette smoking rates remain higher in many marginalized populations (Baggett et al., 2013; Baker & McCarthy, 2021; Cornelius, 2020; Cropsey et al., 2004; Harrison et al., 2020; Jamal et al., 2015; Kelly et al., 2012; Soar et al., 2020). Prioritizing accessibility in implementation is a critical first step towards mitigating rather than exacerbating health disparities (MacEachern & Forkert, 2021).

We aimed to further address this goal with our fairness benchmark analyses. We found that our prediction model performed worse for females than men; however, it performed better for non-White individuals and individuals below the poverty line. These findings suggest our model may be well-positioned to offer benefit to individuals usually disadvantaged by research efforts and mental healthcare.

Long-term outcomes

Challenges of predicting distal treatment success

Alongside these exciting findings, this study also demonstrates that it is difficult to predict complex outcomes like treatment success using distal predictors. Our best estimate of median auROC in held-out data (from the posterior probability distribution) was 0.69, indicating that our model correctly assigns a higher probability to a positive (abstinent) case than a negative (smoking) case 69% of the time. An auROC of 0.7 to 0.8 is considered "acceptable" (Mandrekar, 2010). Our model was right on the cusp of this range. Nevertheless, there is room for improvement in our model's predictive performance.

It is important to note, however, that our primary goal was *not* overall prediction of treatment success. Capturing some signal with our model was necessary to yield credible predictions and establish the relevance of features, and so our model's modest predictive performance may have impacted our ability to select treatments. However, overall prediction was insufficient for our purpose. A model that predicted treatment success perfectly (i.e., auROC of 1.0) but included no interactions would be useless for treatment selection. Indeed, this fact motivated our use of a satisficing metric for a minimum number of interaction terms retained in our models, even though it is possible that we sacrificed some level of predictive power by imposing that additional criterion.

Improving long-term benefit of treatment selection

The clinical benefit that our treatment selection model offers is short-lived. There is no longer statistically significant benefit of treatment matching at 12 weeks, though there is a numeric difference (31.2% abstinence vs. 27.8% abstinence), and there is no statistical or numeric difference at all by 6 months.

Initial treatment success is critical, especially for cigarette smoking where even reducing smoking or quitting for some period of time can improve health outcomes and life expectancy (Jha Prabhat et al., 2013). Additionally, smoking early in a quit attempt can have strong negative con-

sequences: decreased self-efficacy, reduced treatment adherence, and premature treatment cessation (Schlam & Baker, 2013). Findings such as these highlight that selecting a treatment that increases treatment success in early recovery (i.e., at 4 weeks) is necessary - though not sufficient - for long-term success.

Nevertheless, we were of course hopeful that treatment matching benefits would endure. However, it is perhaps unsurprising that they do not. We believe there are several possibilities for the lack of benefit at later assessment points.

First, it is possible that we could have improved prediction overall and treatment selection if we incorporated biological markers or genetic features (Chen et al., 2018). However, extant literature has primarily used single candidate genes or biomarkers, has also not found long-term benefits, and has not translated well into real-world settings (Chen et al., 2020; Chenoweth et al., 2016; Glatard et al., 2017; Lerman et al., 2015; Schnoll et al., 2009; Shahab et al., 2019; Siegel et al., 2020). Additionally, the potential improvement that could come from including biological or genetic features carries an associated cost to implementation given the relative inaccessibility of genetic and biological testing (MacEachern & Forkert, 2021).

A second possibility is that we failed to include non-biological/non-genetic features that are critical for predicting treatment success later on. However, our data come from a large comparative effectiveness trial conducted by a nationally recognized center and designed by foremost experts in the field (Baker et al., 2016). The baseline assessment was quite comprehensive and was based on domain expertise and decades of research. Thus, it seems unlikely that we could be missing enough important features to bridge the gap in our benefit of treatment selection.

We believe the most likely possibility is that our model did not capture dynamic changes over time. It may be that the same features predict treatment success across time. However, because these characteristics can change dynamically *within an individual*, what was the right treatment based on pre-quit characteristics is no longer the right treatment by 12 weeks, 6 months, or beyond. Many of the features used for prediction in this model were baseline measurements of *current* states - withdrawal, dependence, confidence/motivation to quit, time around other smokers, distress tolerance, depression symptoms, among others. Even features that feel more "static" like employment or marital status can be subject to change.

Dynamic change is the rule rather than the exception when it comes to chronic diseases like substance use disorders. Tobacco and other substance use disorders are dynamic and relapsing; both risk for use and the factors driving that risk fluctuate over time (Brandon et al., 2007). This change over time has been identified as a key barrier to overcome for precision mental health goals in addiction (Oliver & McClernon, 2017).

Accounting for dynamic changes will require ongoing assessment of key features that predict treatment success. Such monitoring is now feasible given developments in personal sensing (i.e., in situ data collection via sensors embedded in individuals' daily lives) (Bae et al., 2018; Chih et al., 2014; Epstein et al., 2020; Soyster et al., 2022; Wyant et al., 2024) and is acceptable to individuals with substance use disorders (Wyant et al., 2023). Sensing via ecological momentary assessment is well-positioned to capture the self-report items used as features in this treatment selection model. Although frequently answering a 50-item survey is likely not feasible, it may be that more proximal features have greater predictive power, thus requiring fewer features for successful treatment selection. Indeed, in previous work in our laboratory, we have been able to predict hour-by-hour probabilities of alcohol lapses quite accurately (auROCs > 0.9) when using self-report features measured much closer to the outcome (Wyant et al., 2024).

Using this dynamic monitoring, we can select and adapt treatments and supports over time. There are certainly opportunities to adapt medications – for example, titrating doses or identifying moments when someone could benefit most from a lozenge. In general, medications are more static

treatments that are less well-suited to dynamic changes; however, they do not need to be used on their own. We observed that matching people to the right medication improves treatment success at 4 weeks. Medications could help create early treatment success that positions people to engage with additional supports that might be able to adapt more dynamically. Alternative treatments and supports via web-based interventions and mobile health apps may offer platforms for sustainable, scalable, ongoing support. We can recommend specific modules and individual tools that map onto currently important features affecting treatment success. This mapping between risk factors and supports is likely to be quite complex and will require considerable future research. But if we hope to advance precision mental health for smoking, addiction, and even mental health broadly, we must consider the dynamic nature of risk and recovery inherent in these conditions.

Interpreting our treatment selection model

Several recent reviews have noted that the low interpretability of "black box" machine learning models may impede their utility for clinical and public health goals (Cohen & DeRubeis, 2018; MacEachern & Forkert, 2021; Mooney & Pejaver, 2018). Consequently, we aimed to make our model as interpretable as possible. We identified features that predict differential treatment success (i.e., features that interact with treatment to help us select among treatments). Our best selected model configuration retained 74 interaction terms that spanned 52 unique items. Each feature's associated global Shapley value, which indicates overall magnitude of feature importance, was relatively small. This finding supports what has long been suspected in precision mental health: There is no one feature that explains sufficient variance to make differential predictions by treatment on its own. Rather, each feature offers only a small contribution, but these features together can guide treatment selection.

We also identified features that predict treatment success overall (i.e., "main effects"). These features contribute to the smoking cessation literature and support the conclusion from a recent review that predictors of treatment success span many categories (Bickel et al., 2023). We found similar breadth in important features in this model: economic (e.g., income), environmental (e.g., living with another smoker), sociodemographic (e.g., marital status), psychological (e.g., depression diagnosis), physical health (e.g., pain interfering with daily activities), and smoking use/history (e.g., longest previous quit attempt) characteristics all contributed to predicting treatment success.

Additionally, these prognostic factors may yet help to advance precision mental health goals. These factors may represent mechanisms underlying smoking cessation success and thus offer targeted areas for future treatment development. They also may be used to tailor existing treatments to increase success across individuals. Main effect features in our model may also interact with other treatments not included in this study. For example, buproprion is another first-line smoking cessation medication (Cahill et al., 2013). It may be that some features in this model do not differentiate among C-NRT, nicotine patch, or varenicline but would differentiate between one of these treatments and buproprion.

The role of demographic features

Several demographic features emerged as important: race, gender, income, and marital status. Ethnicity did not emerge as an important feature, and race-based features were specifically related to identifying as a Black or White individual. However, the limited representation of Hispanic, Latino/a, Asian, Multiracial, and Native American/Alaska Native individuals in this sample warrants caution in drawing conclusions about the predictive utility of ethnicity- and race-based features.

In some contexts, it would be problematic to use features that tap into constructs delineating marginalized identities such as race or socioeconomic status. For example, making decisions about who gets insurance (or doesn't) and who gets released earlier from prison (or doesn't) based

on race is discriminatory (e.g., (Farayola et al., 2023)). However, in the precision mental health landscape, we are not deciding *who* gets treatment. Rather, we are deciding *which* treatment to give a specific patient. Thus, we can take advantage of experiential or symptomatological differences as a result of characteristics such as race, ethnicity, sex, income, or comorbid health conditions to improve treatment outcomes across vulnerable subpopulations.

The downside of transparency

When patients and clinicians can easily see non-intuitive features, they may be dissuaded from using or trusting this treatment selection model. For example, an interaction feature with the ASI-3 item "When my throat is tight, I worry I will choke to death" was retained. Because it seems to indicate a rather severe level of anxiety about physical conditions, perhaps it offered more predictive power than other correlated features and was thus retained by GLMNet. Regardless of the reason, the reality is that seeing questions like this on a treatment selection assessment may make patients or clinicians doubt the trustworthiness of the model.

Future directions

We made a concerted effort in this project to evaluate how our treatment selection model would perform for new patients. Regardless, the ultimate test of this model's clinical benefit will be in a prospective trial. This trial will offer two tests. First, it will assess whether using this model is feasible and acceptable to patients and clinicians in clinical practice. Second, we can evaluate the benefit of our treatment selection model in an entirely new sample. Individuals who receive their model-predicted best treatment could be compared to any one of several possible comparison groups. Individuals in the comparison group could receive a random treatment assignment (mimicking clinical trials). Alternatively, they could receive clinician-assigned treatment to mirror traditional treatment selection (and best clinical practice). Another option is that we could compare to a simpler model. For example, there is evidence that treatment success is lower when people are re-treated with the same treatment (Fiore et al., 2008; Gonzales et al., 2014; Heckman et al., 2017; Tønnesen et al., 1993) and that treatment adherence is higher when people choose their preferred treatment (Cropsey et al., 2017). Thus, treatments in the comparison group could be assigned based on patient preference and re-treatment status. Each comparison offers different advantages and disadvantages that should be considered thoughtfully when designing a prospective trial.

Conclusion

Overall, this study has potential for immediate benefit to individuals looking to quit smoking. Our treatment selection model can improve the probability of abstinence during early recovery by a statistically significant and clinically meaningful margin. Moreover, it can do so with a relatively low-burden assessment that uses widely accessible features. This treatment selection model may serve as an initial tool embedded in a continuing care landscape where treatments are adapted dynamically over time. The ultimate test of our model will be in a prospective trial that assesses its feasibility, acceptability, and effectiveness in clinical practice. We are optimistic about the promise our model holds to improve the public health burden of cigarette smoking.

Bibliography

Adjei, K., & Ali, A. A. (2022). CO110 Comparative Effectiveness of Sertraline, Fluoxetine Vs Escitalopram Among Adults with Depression in the United States. *Value in Health*, *25*(7), S324-S325. https://doi.org/10.1016/j.jval.2022.04.206

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders* (Fifth Edition). American Psychiatric Association. https://doi.org/10.1176/appi.books.9780890425596

Bae, S., Chung, T., Ferreira, D., Dey, A. K., & Suffoletto, B. (2018). Mobile Phone Sensors and Supervised Machine Learning to Identify Alcohol Use Events in Young Adults: Implications for Just-in-Time Adaptive Interventions. *Addictive Behaviors*, *83*, 42–47. https://doi.org/10.1016/j.addbeh.2017.11.039

Baggett, T. P., Tobey Matthew L., & Rigotti Nancy A. (2013). Tobacco Use among Homeless People — Addressing the Neglected Addiction. *New England Journal of Medicine*, *369*(3), 201–204. https://doi.org/10.1056/NEJMp1301935

Baker, T. B., & McCarthy, D. E. (2021). Smoking Treatment: A Report Card on Progress and Challenges. *Annual Review of Clinical Psychology*, *17*(Volume17, 2021), 1–30. https://doi.org/10.1146/annurev-clinpsy-081219-090343

Baker, T. B., Piper, M. E., Stein, J. H., Smith, S. S., Bolt, D. M., Fraser, D. L., & Fiore, M. C. (2016). Effects of Nicotine Patch vs Varenicline vs Combination Nicotine Replacement Therapy on Smoking Cessation at 26 Weeks: A Randomized Clinical Trial. *JAMA*, *315*(4), 371–379. https://doi.org/10.1001/jama.2015.19284

Barksdale, C. L., Pérez-Stable, E., & Gordon, J. (2022). Innovative Directions to Advance Mental Health Disparities Research. *American Journal of Psychiatry*, *179*(6), 397–401. https://doi.org/10.1176/appi.ajp.21100972

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., Dai, B., & Grothendieck, G. (2015). *Lme4: Linear Mixed-Effects Models Using 'Eigen' and S4*.

Bickel, W. K., Tomlinson, D. C., Craft, W. H., Ma, M., Dwyer, C. L., Yeh, Y.-H., Tegge, A. N., Freitas-Lemos, R., & Athamneh, L. N. (2023). Predictors of Smoking Cessation Outcomes Identified by Machine Learning: A Systematic Review. *Addiction Neuroscience*, *6*, 100068. https://doi.org/10.1016/j.addicn.2023.100068

Bickman, L. (2020). Improving Mental Health Services: A 50-Year Journey from Randomized Experiments to Artificial Intelligence and Precision Mental Health. *Administration and Policy in Mental Health*, 1–49. https://doi.org/10.1007/s10488-020-01065-8

Bickman, L., Lyon, A. R., & Wolpert, M. (2016). Achieving Precision Mental Health through Effective Assessment, Monitoring, and Feedback Processes. *Administration and Policy in Mental Health and Mental Health Services Research*, *43*(3), 271–276. https://doi.org/10.1007/s10488-016-0718-5

Biecek, P. (2018). DALEX: Explainers for Complex Predictive Models in R. *Journal of Machine Learning Research*, *19*(84), 1–5.

Brandon, T. H., Vidrine, J. I., & Litvin, E. B. (2007). Relapse and Relapse Prevention. *Annual Review of Clinical Psychology*, *3*(1), 257–284. https://doi.org/10.1146/annurev.clinpsy.3.022806.091455

Cahill, K., Lindson-Hawley, N., Thomas, K. H., Fanshawe, T. R., & Lancaster, T. (2016). Nicotine Receptor Partial Agonists for Smoking Cessation. *The Cochrane Database of Systematic Reviews*, *2016*(5). https://doi.org/10.1002/14651858.CD006103.pub7

Cahill, K., Stevens, S., Perera, R., & Lancaster, T. (2013). Pharmacological Interventions for Smoking Cessation: An Overview and Network Meta-Analysis. *The Cochrane Database of Systematic Reviews*, *5*, CD9329. https://doi.org/10.1002/14651858.CD009329.pub2

Center for High Throughput Computing. (2006). *Center for High Throughput Computing*. Center for High Throughput Computing. https://doi.org/10.21231/GNT1-HW21

Centers for Disease Control and Prevention (CDC). *Annual Average for United States 2011–2015 Alcohol-Attributable Deaths Due to Excessive Alcohol Use, All Ages*.

Chen, L.-S., Baker, T. B., Miller, J. P., Bray, M., Smock, N., Chen, J., Stoneking, F., Culverhouse, R. C., Saccone, N. L., Amos, C. I., Carney, R. M., Jorenby, D. E., & Bierut, L. J. (2020). Genetic Variant in CHRNA5 and Response to Varenicline and Combination Nicotine Replacement in a Randomized Placebo-Controlled Trial. *Clinical Pharmacology & Therapeutics*, *108*(6), 1315–1325. https://doi.org/10.1002/cpt.1971

Chen, L.-S., Horton, A., & Bierut, L. (2018). Pathways to Precision Medicine in Smoking Cessation Treatments. *Neuroscience Letters*, *669*, 83–92. https://doi.org/10.1016/j.neulet.2016.05.033

Chenoweth, M. J., Schnoll, R. A., Novalen, M., Hawk, L. W., George, T. P., Cinciripini, P. M., Lerman, C., & Tyndale, R. F. (2016). The Nicotine Metabolite Ratio Is Associated With Early Smoking Abstinence Even After Controlling for Factors That Influence the Nicotine Metabolite Ratio. *Nicotine & Tobacco Research*, *18*(4), 491–495. https://doi.org/10.1093/ntr/ntv125

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

Cohen, Z. D., & DeRubeis, R. J. (2018). Treatment Selection in Depression. *Annual Review of Clinical Psychology*, *14*(1), 209–236. https://doi.org/10.1146/annurev-clinpsy-050817-084746

Collins, L. M. (2018). *Optimization of Behavioral, Biobehavioral, and Biomedical Interventions: The Multiphase Optimization Strategy (MOST)*. Springer International Publishing. https://doi.org/10.1007/978-3-319-72206-1

Cornelius, M. E. (2020). Tobacco Product Use Among Adults — United States, 2019. *MMWR. Morbidity and Mortality Weekly Report*, *69*. https://doi.org/10.15585/mmwr.mm6946a4

Coughlin, L. N., Tegge, A. N., Sheffer, C. E., & Bickel, W. K. (2020). A Machine-Learning Approach to Predicting Smoking Cessation Treatment Outcomes. *Nicotine & Tobacco Research*, *22*(3), 415–422. https://doi.org/10.1093/ntr/nty259

Crawford, J. R., & Henry, J. D. (2004). The Positive and Negative Affect Schedule (PANAS): Construct Validity, Measurement Properties and Normative Data in a Large Non-Clinical Sample. *British Journal of Clinical Psychology*, *43*(3), 245–265. https://doi.org/10.1348/0144665031752934

Cropsey, K. L., Hendricks, P. S., Schiavon, S., Sellers, A., Froelich, M., Shelton, R. C., & Carpenter, M. J. (2017). A Pilot Trial of In Vivo NRT Sampling to Increase Medication

Adherence in Community Corrections Smokers. *Addictive Behaviors*, *67*, 92–99. https://doi.org/10.1016/j.addbeh.2016.12.011

Cropsey, K., Eldridge, G. D., & Ladner, T. (2004). Smoking among Female Prisoners: An Ignored Public Health Epidemic. *Addictive Behaviors*, *29*(2), 425–431. https://doi.org/10.1016/j.addbeh.2003.08.014

DeRubeis, R. J. (2019). The History, Current Status, and Possible Future of Precision Mental Health. *Behaviour Research and Therapy*, *123*, 103506. https://doi.org/10.1016/j.brat.2019.103506

Dwyer, D. B., Falkai, P., & Koutsouleris, N. (2018). Machine Learning Approaches for Clinical Psychology and Psychiatry. *Annual Review of Clinical Psychology*, *14*(1), 91–118. https://doi.org/10.1146/annurev-clinpsy-032816-045037

Epstein, D. H., Tyburski, M., Kowalczyk, W. J., Burgess-Hull, A. J., Phillips, K. A., Curtis, B. L., & Preston, K. L. (2020). Prediction of Stress and Drug Craving Ninety Minutes in the Future with Passively Collected GPS Data. *Npj Digital Medicine*, *3*(1), 26. https://doi.org/10/ghqvcw

Etter, J.-F., Vera Cruz, G., & Khazaal, Y. (2023). Predicting Smoking Cessation, Reduction and Relapse Six Months after Using the Stop-Tabac App for Smartphones: A Machine Learning Analysis. *BMC Public Health*, *23*(1), 1076. https://doi.org/10.1186/s12889-023-15859-6

Farayola, M. M., Tal, I., Connolly, R., Saber, T., & Bendechache, M. (2023). Ethics and Trustworthiness of AI for Predicting the Risk of Recidivism: A Systematic Literature Review. *Information*, *14*(8), 426. https://doi.org/10.3390/info14080426

Feczko, E., & Fair, D. A. (2020). Methods and Challenges for Assessing Heterogeneity. *Biological Psychiatry*, *88*(1), 9–17. https://doi.org/10.1016/j.biopsych.2020.02.015

Fiore, M. C., Jaen, C. R., Baker, T. B., Bailey, W. C., Benowitz, N. L., Curry, S. J., Dorfman, S. F., Froelicher, E. S., Goldstein, M. G., Healton, C. G., Henderson, P. N., Heyman, R. B., Koh, H. K., Kottke, T. E., Lando, H. A., Mecklenburg, R. E., Mermelstein, R. J., Mullen, P. D., Orleans, C. T., … Wewers, M. E. (2008). *A Clinical Practice Guideline for Treating Tobacco Use and Dependence: 2008 Update. A U.S. Public Health Service Report.* Rockville, MD: U.S. Department of Health and Human Services, U.S. Public Health Service.

Glatard, A., Dobrinas, M., Gholamrezaee, M., Lubomirov, R., Cornuz, J., Csajka, C., & Eap, C. B. (2017). Association of Nicotine Metabolism and Sex with Relapse Following Varenicline and Nicotine Replacement Therapy. *Experimental and Clinical Psychopharmacology*, *25*(5), 353–362. https://doi.org/10.1037/pha0000141

Gonzales, D., Hajek, P., Pliamm, L., Nackaerts, K., Tseng, L.-J., McRae, T. D., & Treadow, J. (2014). Retreatment with Varenicline for Smoking Cessation in Smokers Who Have Previously Taken Varenicline: A Randomized, Placebo-Controlled Trial. *Clinical Pharmacology and Therapeutics*, *96*(3), 390–396. https://doi.org/10.1038/clpt.2014.124

Greenfield, S. F., Brooks, A. J., Gordon, S. M., Green, C. A., Kropp, F., McHugh, R. K., Lincoln, M., Hien, D., & Miele, G. M. (2007). Substance Abuse Treatment Entry, Retention, and Outcome in Women: A Review of the Literature. *Drug and Alcohol Dependence*, *86*(1), 1–21. https://doi.org/10.1016/j.drugalcdep.2006.05.012

Harrison, A., Ramo, D., Hall, S. M., Estrada-Gonzalez, V., & Tolou-Shams, M. (2020). Cigarette Smoking, Mental Health, and Other Substance Use among Court-Involved Youth. *Substance Use & Misuse*, *55*(4), 572–581. https://doi.org/10.1080/10826084.2019.1691593

Heatherton, T. F., Kozlowski, L. T., Frecker, R. C., & Fagerström, K. O. (1991). The Fagerström Test for Nicotine Dependence: A Revision of the Fagerström Tolerance Questionnaire. *British Journal of Addiction*, *86*(9), 1119–1127.

Heckman, B. W., Cummings, K. M., Kasza, K. A., Borland, R., Burris, J. L., Fong, G. T., McNeill, A., & Carpenter, M. J. (2017). Effectiveness of Switching Smoking-Cessation Medications Following Relapse. *American Journal of Preventive Medicine*, *53*(2), e63-e70. https://doi.org/10.1016/j.amepre.2017.01.038

Insel, T. R. (2014). The NIMH Research Domain Criteria (RDoC) Project: Precision Medicine for Psychiatry. *American Journal of Psychiatry*, *171*(4), 395–397. https://doi.org/10.1176/appi.ajp.2014.14020138

Issabakhsh, M., Sánchez-Romero, L. M., Le, T. T. T., Liber, A. C., Tan, J., Li, Y., Meza, R., Mendez, D., & Levy, D. T. (2023). Machine Learning Application for Predicting Smoking Cessation among US Adults: An Analysis of Waves 1-3 of the PATH Study. *Plos One*, *18*(6), e286883. https://doi.org/10.1371/journal.pone.0286883

Jacobson, N. C., Kowatsch, T., & Marsch, L. A. (Eds.). (2022). *Digital Therapeutics for Mental Health and Addiction: The State of the Science and Vision for the Future* (1st edition). Academic Press.

Jamal, A., Homa, D. M., O'Connor, E., Babb, S. D., Caraballo, R. S., Singh, T., Hu, S. S., & King, B. A. (2015). Current Cigarette Smoking among Adults - United States, 2005-2014. *MMWR. Morbidity and Mortality Weekly Report*, *64*(44), 1233–1240. https://doi.org/10.15585/mmwr.mm6444a2

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning: With Applications in R* (7th ed.). Springer-Verlag. https://doi.org/10.1007/978-1-4614-7138-7

Jha Prabhat, Ramasundarahettige Chinthanie, Landsman Victoria, Rostron Brian, Thun Michael, Anderson Robert N., McAfee Tim, & Peto Richard. (2013). 21st-Century Hazards of Smoking and Benefits of Cessation in the United States. *New England Journal of Medicine*, *368*(4), 341–350. https://doi.org/10.1056/NEJMsa1211128

Jonathan, P., Krzanowski, W. J., & McCarthy, W. V. (2000). On the Use of Cross-Validation to Assess Performance in Multivariate Prediction. *Statistics and Computing*, *10*(3), 209–229. https://doi.org/10.1023/A:1008987426876

Jordan, C. J., & Xi, Z.-X. (2018). Discovery and Development of Varenicline for Smoking Cessation. *Expert Opinion on Drug Discovery*, *13*(7), 671–683. https://doi.org/10.1080/17460441.2018.1458090

Karaca-Mandic, P., Norton, E. C., & Dowd, B. (2012). Interaction Terms in Nonlinear Models. *Health Services Research*, *47*(1pt1), 255–274. https://doi.org/10.1111/j.1475-6773.2011.01314.x

Kaufmann, A., Hitsman, B., Goelz, P. M., Veluz-Wilkins, A., Blazekovic, S., Powers, L., Leone, F. T., Gariti, P., Tyndale, R. F., & Schnoll, R. A. (2015). Rate of Nicotine Metabolism

and Smoking Cessation Outcomes in a Community-based Sample of Treatment-Seeking Smokers. *Addictive Behaviors*, *51*, 93–99. https://doi.org/10.1016/j.addbeh.2015.07.019

Kaye, J. T., Johnson, A. L., Baker, T. B., Piper, M. E., & Cook, J. W. (2020). Searching for Personalized Medicine for Binge Drinking Smokers: Smoking Cessation Using Varenicline, Nicotine Patch, or Combination Nicotine Replacement Therapy. *Journal of Studies on Alcohol and Drugs*, *81*(4), 426–435. https://doi.org/10.15288/jsad.2020.81.426

Kelly, P. J., Baker, A. L., Deane, F. P., Kay-Lambkin, F. J., Bonevski, B., & Tregarthen, J. (2012). Prevalence of Smoking and Other Health Risk Factors in People Attending Residential Substance Abuse Treatment. *Drug and Alcohol Review*, *31*(5), 638–644. https://doi.org/10.1111/j.1465-3362.2012.00465.x

Kessler, R. C., & Luedtke, A. (2021). Pragmatic Precision Psychiatry—A New Direction for Optimizing Treatment Selection. *JAMA Psychiatry*, *78*(12), 1384–1390. https://doi.org/10.1001/jamapsychiatry.2021.2500

Kilaru, A. S., Xiong, A., Lowenstein, M., Meisel, Z. F., Perrone, J., Khatri, U., Mitra, N., & Delgado, M. K. (2020). Incidence of Treatment for Opioid Use Disorder Following Nonfatal Overdose in Commercially Insured Patients. *JAMA Network Open*, *3*(5), e205852. https://doi.org/10.1001/jamanetworkopen.2020.5852

Kiluk, B. D., Dreifuss, J. A., Weiss, R. D., Morgenstern, J., & Carroll, K. M. (2013). The Short Inventory of Problems – Revised (SIP-R): Psychometric Properties within a Large, Diverse Sample of Substance Use Disorder Treatment Seekers. *Psychology of Addictive Behaviors : Journal of the Society of Psychologists in Addictive Behaviors*, *27*(1), 307–314. https://doi.org/10.1037/a0028445

Kranzler, H. R., Smith, R. V., Schnoll, R., Moustafa, A., & Greenstreet-Akman, E. (2017). Precision Medicine and Pharmacogenetics: What Does Oncology Have That Addiction Medicine Does Not?. *Addiction*, *112*(12), 2086–2094. https://doi.org/10.1111/add.13818

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: Validity of a Brief Depression Severity Measure. *Journal of General Internal Medicine*, *16*(9), 606–613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x

Krstajic, D., Buturovic, L. J., Leahy, D. E., & Thomas, S. (2014). Cross-Validation Pitfalls When Selecting and Assessing Regression and Classification Models. *Journal of Cheminformatics*, *6*. https://doi.org/10.1186/1758-2946-6-10

Kuhn, M. (2022). *Tidyposterior: Bayesian Analysis to Compare Models Using Resampling Statistics*.

Kuhn, M., & Johnson, K. (2018). *Applied Predictive Modeling* (1st ed. 2013, Corr. 2nd printing 2018 edition). Springer. https://doi.org/10.1007/978-1-4614-6849-3

Kuhn, M., & Wickham, H. (2020). *Tidymodels: A Collection of Packages for Modeling and Machine Learning Using Tidyverse Principles*.

Lai, C.-C., Huang, W.-H., Chang, B. C.-C., & Hwang, L.-C. (2021). Development of Machine Learning Models for Prediction of Smoking Cessation Outcome. *International Journal of Environmental Research and Public Health*, *18*(5), 2584. https://doi.org/10.3390/ijerph18052584

Lenzner, T., Kaczmirek, L., & Lenzner, A. (2010). Cognitive Burden of Survey Questions and Response Times: A Psycholinguistic Experiment. *Applied Cognitive Psychology*, *24*(7), 1003–1020. https://doi.org/10.1002/acp.1602

Lerman, C., Schnoll, R. A., Hawk, L. W., Cinciripini, P., George, T. P., Wileyto, E. P., Swan, G. E., Benowitz, N. L., Heitjan, D. F., Tyndale, R. F., & PGRN-PNAT Research Group. (2015). Use of the Nicotine Metabolite Ratio as a Genetically Informed Biomarker of Response to Nicotine Patch or Varenicline for Smoking Cessation: A Randomised, Double-Blind Placebo-Controlled Trial. *The Lancet. Respiratory Medicine*, *3*(2), 131–138. https://doi.org/10.1016/S2213-2600(14)70294-2

Lewis, C., Roberts, N. P., Andrew, M., Starling, E., & Bisson, J. I. (2020). Psychological Therapies for Post-Traumatic Stress Disorder in Adults: Systematic Review and Meta-Analysis. *European Journal of Psychotraumatology*, *11*(1), 1729633. https://doi.org/10.1080/20008198.2020.1729633

Lieberman, J. A. (2004). Dopamine Partial Agonists: A New Class of Antipsychotic. *CNS Drugs*, *18*(4), 251–267. https://doi.org/10.2165/00023210-200418040-00005

Lindson, N., Chepkin, S. C., Ye, W., Fanshawe, T. R., Bullen, C., & Hartmann-Boyce, J. (2019). Different Doses, Durations and Modes of Delivery of Nicotine Replacement Therapy for Smoking Cessation. *Cochrane Database of Systematic Reviews*, *4*. https://doi.org/10.1002/14651858.CD013308

Lundberg, S. M., & Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 4768–4777.

MacEachern, S. J., & Forkert, N. D. (2021). Machine Learning for Precision Medicine. *Genome*, *64*(4), 416–425. https://doi.org/10.1139/gen-2020-0131

Mandrekar, J. N. (2010). Receiver Operating Characteristic Curve in Diagnostic Test Assessment. *Journal of Thoracic Oncology: Official Publication of the International Association for the Study of Lung Cancer*, *5*(9), 1315–1316. https://doi.org/10.1097/JTO.0b013e3181ec173d

Marszalek, J. M., Barber, C., Kohlhart, J., & Cooper, B. H. (2011). Sample Size in Psychological Research over the Past 30 Years. *Perceptual and Motor Skills*, *112*(2), 331–348. https://doi.org/10.2466/03.11.PMS.112.2.331-348

Massago, M., Massago, M., Iora, P. H., Tavares Gurgel, S. J., Conegero, C. I., Carolino, I. D. R., Mushi, M. M., Chaves Forato, G. A., de Souza, J. V. P., Hernandes Rocha, T. A., Bonfim, S., Staton, C. A., Nihei, O. K., Vissoci, J. R. N., & de Andrade, L. (2024). Applicability of Machine Learning Algorithm to Predict the Therapeutic Intervention Success in Brazilian Smokers. *Plos One*, *19*(3), e295970. https://doi.org/10.1371/journal.pone.0295970

Mooney, S. J., & Pejaver, V. (2018). Big Data in Public Health: Terminology, Machine Learning, and Privacy. *Annual Review of Public Health*, *39*, 95–112. https://doi.org/10.1146/annurev-publhealth-040617-014208

Morales, D. A., Barksdale, C. L., & Beckel-Mitchener, A. C. (2020). A Call to Action to Address Rural Mental Health Disparities. *Journal of Clinical and Translational Science*, *4*(5), 463–467. https://doi.org/10.1017/cts.2020.42

Moriarty, D. (1996). CDC Studies Community Quality of Life. *NACCHO News*, *10*, 13.

National Center for Chronic Disease Prevention and Health Promotion (US) Office on Smoking and Health. (2014). *The Health Consequences of Smoking—50 Years of Progress: A Report of the Surgeon General*. Centers for Disease Control and Prevention (US).

Netzer, N. C., Stoohs, R. A., Netzer, C. M., Clark, K., & Strohl, K. P. (1999). Using the Berlin Questionnaire to Identify Patients at Risk for the Sleep Apnea Syndrome. *Annals of Internal Medicine*, *131*(7), 485–491. https://doi.org/10.7326/0003-4819-131-7-199910050-00002

Ng, A. (2018). *Machine Learning Yearning*. 1–118.

Office of the Surgeon General (US), Center for Mental Health Services (US), & National Institute of Mental Health (US). (2001). *Mental Health: Culture, Race, and Ethnicity*. Substance Abuse and Mental Health Services Administration (US).

Olfson, M., Mauro, C., Wall, M. M., Choi, C. J., Barry, C. L., & Mojtabai, R. (2022). Healthcare Coverage and Service Access for Low-Income Adults with Substance Use Disorders. *Journal of Substance Abuse Treatment*, *137*, 108710. https://doi.org/10.1016/j.jsat.2021.108710

Oliver, J. A., & McClernon, F. J. (2017). Precision Medicine in Addiction Research: Where Has the Time Gone?. *Addiction*, *112*(12), 2096–2097. https://doi.org/10.1111/add.14023

Pinedo, M. (2019). A Current Re-Examination of Racial/Ethnic Disparities in the Use of Substance Abuse Treatment: Do Disparities Persist?. *Drug and Alcohol Dependence*, *202*, 162–167. https://doi.org/10.1016/j.drugalcdep.2019.05.017

Piper, Megan E., Cook, et al. (2017). Toward Precision Smoking Cessation Treatment II: Proximal Effects of Smoking Cessation Intervention Components on Putative Mechanisms of Action. *Drug and Alcohol Dependence*, *171*, 50–58. https://doi.org/10.1016/j.drugalcdep.2016.11.027

Piper, M. E., Fiore, M. C., Smith, S. S., Fraser, D., Bolt, D. M., Collins, L. M., Mermelstein, R., Schlam, T. R., Cook, J. W., Jorenby, D. E., Loh, W.-Y., & Baker, T. B. (2016). Identifying Effective Intervention Components for Smoking Cessation: A Factorial Screening Experiment. *Addiction (Abingdon, England)*, *111*(1), 129–141. https://doi.org/10.1111/add.13162

Piper, Megan E., Schlam, et al. (2017). Toward Precision Smoking Cessation Treatment I: Moderator Results from a Factorial Experiment. *Drug and Alcohol Dependence*, *171*, 59–65. https://doi.org/10.1016/j.drugalcdep.2016.11.025

Rigotti, N. A., Kruse, G. R., Livingstone-Banks, J., & Hartmann-Boyce, J. (2022). Treatment of Tobacco Smoking: A Review. *JAMA*, *327*(6), 566–577. https://doi.org/10.1001/jama.2022.0395

Roussidis, A., Kalkavoura, C., Dimelis, D., Theodorou, A., Ioannidou, I., Mellos, E., Mylonaki, T., Spyropoulou, A., & Yfantis, A. (2013). Reasons and Clinical Outcomes of Antipsychotic Treatment Switch in Outpatients with Schizophrenia in Real-Life Clinical Settings: The ETOS Observational Study. *Annals of General Psychiatry*, *12*, 42. https://doi.org/10.1186/1744-859X-12-42

Schlam, T. R., & Baker, T. B. (2013). Interventions for Tobacco Smoking. *Annual Review of Clinical Psychology*, *9*, 675–702. https://doi.org/10.1146/annurev-clinpsy-050212-185602

Schnoll, R. A., Patterson, F., Wileyto, E. P., Tyndale, R. F., Benowitz, N., & Lerman, C. (2009). Nicotine Metabolic Rate Predicts Successful Smoking Cessation with Transdermal Nicotine: A Validation Study. *Pharmacology Biochemistry and Behavior*, *92*(1), 6–11. https://doi.org/10.1016/j.pbb.2008.10.016

Shahab, L., Bauld, L., McNeill, A., & Tyndale, R. F. (2019). Does the Nicotine Metabolite Ratio Moderate Smoking Cessation Treatment Outcomes in Real-World Settings? A Prospective Study. *Addiction*, *114*(2), 304–314. https://doi.org/10.1111/add.14450

Siegel, S. D., Lerman, C., Flitter, A., & Schnoll, R. A. (2020). The Use of the Nicotine Metabolite Ratio as a Biomarker to Personalize Smoking Cessation Treatment: Current Evidence and Future Directions. *Cancer Prevention Research*, *13*(3), 261–272. https://doi.org/10.1158/1940-6207.CAPR-19-0259

Simons, J. S., & Gaher, R. M. (2005). The Distress Tolerance Scale: Development and Validation of a Self-Report Measure. *Motivation and Emotion*, *29*(2), 83–102. https://doi.org/10.1007/s11031-005-7955-3

Smets, E. M. A., Garssen, B., Bonke, B., & De Haes, J. C. J. M. (1995). The Multidimensional Fatigue Inventory (MFI) Psychometric Qualities of an Instrument to Assess Fatigue. *Journal of Psychosomatic Research*, *39*(3), 315–325. https://doi.org/10.1016/0022-3999(94)00125-O

Smith, S. S., Piper, M. E., Bolt, D. M., Fiore, M. C., Wetter, D. W., Cinciripini, P. M., & Baker, T. B. (2010). Development of the Brief Wisconsin Inventory of Smoking Dependence Motives. *Nicotine & Tobacco Research*, *12*(5), 489–499. https://doi.org/10.1093/ntr/ntq032

Smith, S. S., Piper, M. E., Bolt, D. M., Kaye, J. T., Fiore, M. C., & Baker, T. B. (2021). Revision of the Wisconsin Smoking Withdrawal Scale: Development of Brief and Long Forms. *Psychological Assessment*, *33*(3), 255–266. https://doi.org/10.1037/pas0000978

Snaith, R. P., Hamilton, M., Morley, S., Humayan, A., Hargreaves, D., & Trigwell, P. (1995). A Scale for the Assessment of Hedonic Tone the Snaith-Hamilton Pleasure Scale. *The British Journal of Psychiatry: The Journal of Mental Science*, *167*(1), 99–103. https://doi.org/10.1192/bjp.167.1.99

Soar, K., Dawkins, L., Robson, D., & Cox, S. (2020). Smoking amongst Adults Experiencing Homelessness: A Systematic Review of Prevalence Rates, Interventions and the Barriers and Facilitators to Quitting and Staying Quit. *Journal of Smoking Cessation*, *15*(2), 94–108. https://doi.org/10.1017/jsc.2020.11

Soyster, P. D., Ashlock, L., & Fisher, A. J. (2022). Pooled and Person-Specific Machine Learning Models for Predicting Future Alcohol Consumption, Craving, and Wanting to Drink: A Demonstration of Parallel Utility. *Psychology of Addictive Behaviors: Journal of the Society of Psychologists in Addictive Behaviors*, *36*(3), 296–306. https://doi.org/10.1037/adb0000666

Substance Abuse and Mental Health Services Administration. (2023). *Key Substance Use and Mental Health Indicators in the United States: Results from the 2022 National Survey on Drug Use and Health.*

Substance Abuse and Mental Health Services Administration (US), & Office of the Surgeon General (US). (2016). *Facing Addiction in America.* US Department of Health and Human Services.

Taylor, S., Zvolensky, M. J., Cox, B. J., Deacon, B., Heimberg, R. G., Ledley, D. R., Abramowitz, J. S., Holaway, R. M., Sandin, B., Stewart, S. H., Coles, M., Eng, W., Daly, E. S., Arrindell, W. A., Bouvard, M., & Cardenas, S. J. (2007). Robust Dimensions of Anxiety Sensitivity: Development and Initial Validation of the Anxiety Sensitivity Index-3. *Psychological Assessment*, *19*(2), 176–188. https://doi.org/10.1037/1040-3590.19.2.176

Taylor, V. R., & National Center for Chronic Disease Prevention and Health Promotion (U.S.). Division of Adult and Community Health. (2000). *Measuring Healthy Days: Population Assessment of Health-Related Quality of Life*.

Trautmann, S., Rehm, J., & Wittchen, H.-U. (2016). The Economic Costs of Mental Disorders. *EMBO Reports*, *17*(9), 1245–1249. https://doi.org/10.15252/embr.201642951

Tønnesen, P., Nørregaard, J., Säwe, U., & Simonsen, K. (1993). Recycling with Nicotine Patches in Smoking Cessation. *Addiction (Abingdon, England)*, *88*(4), 533–539. https://doi.org/10.1111/j.1360-0443.1993.tb02060.x

Wang, J., Simons-Morton, B. G., Farhat, T., Farhart, T., & Luk, J. W. (2009). Socio-Demographic Variability in Adolescent Substance Use: Mediation by Parents and Peers. *Prevention Science: The Official Journal of the Society for Prevention Research*, *10*(4), 387–396. https://doi.org/10.1007/s11121-009-0141-1

Weisz, J. R., Kuppens, S., Ng, M. Y., Vaughn-Coaxum, R. A., Ugueto, A. M., Eckshtain, D., & Corteselli, K. A. (2019). Are Psychotherapies for Young People Growing Stronger? Tracking Trends Over Time for Youth Anxiety, Depression, Attention-Deficit/Hyperactivity Disorder, and Conduct Problems. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *14*(2), 216–237. https://doi.org/10.1177/1745691618805436

Whiteford, H. A., Degenhardt, L., Rehm, J., Baxter, A. J., Ferrari, A. J., Erskine, H. E., Charlson, F. J., Norman, R. E., Flaxman, A. D., Johns, N., Burstein, R., Murray, C. J. L., & Vos, T. (2013). Global Burden of Disease Attributable to Mental and Substance Use Disorders: Findings from the Global Burden of Disease Study 2010. *Lancet (London, England)*, *382*(9904), 1575–1586. https://doi.org/10.1016/S0140-6736(13)61611-6

Wyant, K., Moshontz, H., Ward, S. B., Fronk, G. E., & Curtin, J. J. (2023). Acceptability of Personal Sensing Among People With Alcohol Use Disorder: Observational Study. *JMIR Mhealth and Uhealth*, *11*(1), e41833. https://doi.org/10.2196/41833

Wyant, K., Sant'Ana, S. J. K., Fronk, G., & Curtin, J. J. (2024). Machine Learning Models for Temporally Precise Lapse Prediction in Alcohol Use Disorder. *Psychopathology and Clinical Science*. https://doi.org/10.31234/osf.io/cgsf7

Youngstrom, E. A. (2014). A Primer on Receiver Operating Characteristic Analysis and Diagnostic Efficiency Statistics for Pediatric Psychology: We Are Ready to ROC. *Journal of Pediatric Psychology*, *39*(2), 204–221. https://doi.org/10.1093/jpepsy/jst062

Zheng, Y., Wiebe, R. P., Cleveland, H. H., Molenaar, P. C. M., & Harris, K. S. (2013). An Idiographic Examination of Day-to-Day Patterns of Substance Use Craving, Negative Affect, and Tobacco Use Among Young Adults in Recovery. *Multivariate Behavioral Research*, *48*(2), 241–266. https://doi.org/10.1080/00273171.2013.763012