

Predicting Substance Use Disorder Treatment Completion

Jared Collina, Kara McGaughey & Valerie Sydnor

05/05/2019

Contents

1. Introduction	1
1.1. Scope of the Problem	1
1.2. Project Goals	2
2. Methods and Model Results	2
2.1. The Treatment Episode Data Set—Discharges (TEDS-D) Data	2
Data Overview	2
Data Cleaning: Selection of Predictor Variables and Discharge Entries	2
Treatment Success Predictors	3
2.2. Study Sample Characteristics	4
2.3. Statistical Models	5
Training, Testing, and Validation Data Split	5
Statistical Models	6
3. Results & Model Comparisons	7
3.1 Regression-based Approaches	7
3.2 Random Forest	7
3.3 Boosting	8
3.4 Deep learning	8
4. Final Model	9
Variable Importance Plots	9
Partial Dependence Plots	10
5. Conclusions, Implications & Limitations	12
5.1 Conclusions & Implications	12
5.2 Limitations	13

1. Introduction

1.1. Scope of the Problem

A substance use disorder (SUD) is a medical, brain-based illness characterized by clinically significant impairments in health, social function, and voluntary control over substance use. SUDs range in severity, duration, and complexity. In addition to this inherent heterogeneity in symptomology and symptom severity, SUDs are incredibly wide-spread. In 2015, 20.8 million people aged 12 or older met official diagnostic criteria for a substance use disorder (Medina, 2016). Of these people, only about 1 in 10 received any type of specialty treatment. On top of this, approximately one-third of people who enroll in substance use treatment facilities drop out or are prematurely terminated by the facility (SAMHSA, 2012).

Critically, while treatment completion is associated with higher rates of drug use termination, non-completion is intimately linked with criminal involvement, relapse/readmission, illness chronicity, and even death. **As such, not only is successful treatment completion is a clinical utile outcome measure, but accounting for and acting on treatment completion rates should be a public health priority.**

Successful treatment completion rates can also be used to assess the functionality and efficacy of national and state-level healthcare systems. As such, identifying key predictors of successful treatment (or treatment completion) can serve not only to maximize positive outcomes based on individual patient characteristics, but also to discover disparities in

treatment access and – ultimately – inform policy to increase treatment success and reduce unmet treatment needs. Specifically, if we are able to a priori identify patients unlikely to complete traditional treatment trajectories, we will be better able to enroll these individuals in evidence-based interventions to diversify treatment opportunities and increase retention rates among this would-be treatment refractory population.

In order to make this leap towards individualized, precision medicine and matching patients to the most effective treatments based on their characteristics, we must first use these characteristics to predict who will succeed at a given treatment.

1.2. Project Goals

As such, the overarching aim of this project is to leverage advanced statistical techniques – logistic regression, penalized regression, random forest, boosting, and deep learning (i.e., neural networks) – to identify factors most predictive of successful substance use disorder treatment outcomes.

To accomplish this, we make use of a large, publicly-available dataset commonly used in the SUD field to evaluate which method works best at predicting successful treatment using real life patient information.

2. Methods and Model Results

2.1. The Treatment Episode Data Set—Discharges (TEDS-D) Data

Data Overview

The Treatment Episode Data Set—Discharges (TEDS-D) is a public dataset made available by the US Department of Health and Human Services that contains patient-level information for individuals treated for substance use disorders in the United States during the year 2006. The TEDS-D data includes variables related to patient demographics (e.g., age, sex, race), geographics (e.g. census division), and psychographics (e.g. employment); patient drug use (e.g. primary substance abused, age of first use, frequency of use); SUD treatment program (e.g. rehabilitation center type, length of treatment stay); and SUD risk factors (e.g. comorbid psychiatric diagnoses, veteran status, living situation).

The binary outcome variable of interest included in the TEDS-D dataset and used as the dependent variable in all statistical models developed here is **Treatment Completed** (yes=1 or no=0). A **completed treatment** indicates that all parts of the treatment plan were completed prior to patient discharge from the treatment program. On the other hand, **treatment non-completion** denotes that the patient left or was discharged prior to treatment completion. The most frequent causes of treatment non-completion included a patient leaving the program early against professional advice, program termination by the treatment facility due to patient non-compliance, or patient transfer to another facility.

Data Cleaning: Selection of Predictor Variables and Discharge Entries

The original TEDS-D dataset included 65 variables and 1,048,575 observations. However, each observation included in this original dataset represented *one hospital discharge*, rather than *one unique patient*. In order to avoid introducing bias into our predictive models via the presence of multiple entries for the same patient, we removed all observations for which the number of prior treatment episodes in a drug or alcohol program was greater than 0. This ensured that each discharge entry and all relevant variables were indeed obtained from a unique patient. We additionally removed all entries wherein the patient was discharged from a 24-hour detoxification service, as detoxification and treatment programs are quite distinctive with regards to program settings and goals. Consequently, we focused our analyses on short (< 30 days) and long term (> 30 days) treatment programs that were inpatient, partial hospitalization, or intensive outpatient programs.

Prior to developing statistical models aimed at predicting treatment completion versus non-completion, we additionally removed predictor variables that were redundant or non-variable across entries.

- Examples of redundant variables excluded from the dataset include state code, metropolitan area, and census region (redundant with regards to census division) and secondary and tertiary route of drug administration.

- Predictors that were non-variable across discharge entries and removed from the dataset included year of discharge and number of prior SUD treatment programs participated in.

Next, we looked at the number of missing values per each predictor variable and removed those with a large number of missing values (>40 % of entries missing data). The following variables were removed for missing data: number of arrests in 30 days prior to admission, DSM diagnosis, detailed not in labor force status, pregnant at time of admission (yes/no), detailed criminal justice referral, health insurance type, and primary source of treatment program payment.

After finalizing the set of predictor variables, we removed any discharge entries that still had missing values for one or more of the predictors. The final cleaned dataset included 42 predictor variables and 70,009 observations. The table below summarizes relevant treatment program information for these remaining observations:

	Did Not Complete	Completed	Total	% Completed
All Subjects	37153	32856	70009	47%
Service Setting	<i>X-squared = 1364.4, df = 3, p-value < 2.2e-16</i>			
Short-Term Rehab	3396 (10.34%)	3514 (9.46%)	6910	51%
Long-Term Rehab	2421 (7.37%)	2288 (6.16%)	4709	49%
Intensive Outpatient	3037 (9.24%)	7080 (19.06%)	10117	70%
Intensive Inpatient	24002 (73.05%)	24271 (65.33%)	48273	50%
Days Before Treatment	<i>X-squared = 59.422, df = 4, p-value = 3.837e-12</i>			
0->100 Days	36474 (98.17%)	32062 (97.58%)	68536	47%
101->200 Days	312 (0.84%)	305 (0.93%)	617	49%
201->300 Days	120 (0.32%)	103 (0.31%)	223	46%
301->900 Days	89 (0.24%)	100 (0.3%)	189	53%
901->1000 Days	158 (0.43%)	286 (0.87%)	444	64%
Length of Stay	<i>X-squared = 3619.9, df = 3, p-value < 2.2e-16</i>			
0->10 Days	5595 (15.06%)	1647 (5.01%)	7242	23%
11->20 Days	3908 (10.52%)	1492 (4.54%)	5400	28%
21->30 Days	6137 (16.52%)	4321 (13.15%)	10458	41%
31->40 Days	21513 (57.9%)	25396 (77.29%)	46909	54%

Treatment Success Predictors

The following variables were considered potential predictors of treatment completion versus non-completion for all statistical models:

- Demographic information: Age, sex, race, ethnicity, marital status, years of education, employment status, primary source of income, living arrangement, veteran status, census division,
- Treatment-related variables: Service setting at discharge, number of days patient waited before entering treatment program, length of treatment stay, principal source of treatment program referral,
- Substance-related variables: Primary SUD diagnosis, primary route of drug administration, frequency of drug use, age of first drug use, number of substances reported at admission, IV drug use reported at admission, prescribed a pharmacological opioid therapy at admission, alcohol/drug substance use type (alcohol only, drugs only, both), and comorbid psychiatric diagnosis (yes/no).

The cleaned data additionally included 18 variables pertaining to whether 18 different substances were reported as a substance of use/abuse at admission (i.e. “drug flags” at admission).

In order to reduce the number of statistical predictors, increase model degrees of freedom, and better capture variance across individuals in the sample, a logistic principal components analysis (PCA) was performed on the 18 drug flags. This data reduction strategy allowed for the identification of a small number of principal components that maximally captured variance in drug use at admission.

Logistic PCA solutions with 1-7 principal components were tested, and a 6 principal component solution was chosen. The 6 principal components identified explained 94% of drug flag variable variance. Principal component scores for the 6 components were generated for each observation, and these 6 scores were used as potential model predictors in

addition to the 24 variables described above. The R function `logisticPCA` from package `logisticPCA` was used for this analysis.

2.2. Study Sample Characteristics

70,009 unique patients (46,986 male and 23,023 female patients) who were discharged from a substance use disorder treatment program in 2006 were included in this report. Patient ages ranged from 12-55+, and most patients were African American or Caucasian. Comprehensive sample demographics are shown below:

	Did Not Complete	Completed	Total	% Completed
All Subjects	37153	32856	70009	47%
Age	<i>X-squared = 177.42, df = 10, p-value < 2.2e-16</i>			
12->14	953 (2.6%)	831 (2.5%)	1784	47%
15->17	4978 (13.4%)	4495 (13.7%)	9473	47%
18->20	3686 (9.9%)	3265 (9.9%)	6951	47%
21->24	5410 (14.6%)	4635 (14.1%)	10045	46%
25->29	5573 (15%)	4595 (14%)	10168	45%
30->34	4058 (10.9%)	3305 (10.1%)	7363	45%
35->39	3724 (10%)	3137 (9.5%)	6861	46%
40->44	3625 (9.8%)	3148 (9.6%)	6773	46%
45->49	2725 (7.3%)	2560 (7.8%)	5285	48%
50->54	1441 (3.9%)	1533 (4.7%)	2974	52%
55+	980 (2.6%)	1352 (4.1%)	2332	58%
Gender	<i>X-squared = 200.52, df = 1, p-value < 2.2e-16</i>			
Male	24056 (64.8%)	22930 (69.8%)	46986	49%
Female	13097 (35.3%)	9926 (30.2%)	23023	43%
Race	<i>X-squared = 684.94, df = 8, p-value < 2.2e-16</i>			
White	25218 (67.9%)	24777 (75.4%)	49995	50%
Alaskan Native	22 (0.1%)	6 (0%)	28	21%
American Indian	475 (1.3%)	414 (1.3%)	889	47%
Pacific Islander	20 (0.1%)	33 (0.1%)	53	62%
Black	10176 (27.4%)	6404 (19.5%)	16580	39%
Asian	86 (0.2%)	100 (0.3%)	186	54%
Other Single Race	926 (2.5%)	742 (2.3%)	1668	44%
Two+ Races	198 (0.5%)	356 (1.1%)	554	64%
Native Hawaiian	32 (0.1%)	24 (0.1%)	56	43%
Ethnic	<i>X-squared = 94.322, df = 5, p-value < 2.2e-16</i>			
Not Hispanic	34884 (93.9%)	30402 (92.5%)	65286	47%
Puerto Rican	414 (1.1%)	403 (1.2%)	817	49%
Mexican	1066 (2.9%)	989 (3.0%)	2055	48%
Cuban	170 (0.5%)	261 (0.8%)	431	61%
Hispanic, Other	540 (1.5%)	727 (2.2%)	1267	57%
Hispanic, Unspecified	79 (0.2%)	74 (0.2%)	153	48%
Marriage Status	<i>X-squared = 170.58, df = 3, p-value < 2.2e-16</i>			
Never Married	24515 (66%)	20262 (61.7%)	44777	45%
Married	5787 (15.6%)	6188 (18.8%)	11975	52%
Separated	1779 (4.8%)	1607 (4.9%)	3386	47%
Divorced/Widowed	5072 (13.7%)	4799 (14.6%)	9871	49%
Education	<i>X-squared = 534.28, df = 4, p-value < 2.2e-16</i>			
8 Years or Less	3948 (10.6%)	3231 (9.8%)	7179	45%
9 To 11 Years	13521 (36.4%)	9871 (30%)	23392	42%
12 Years	13109 (35.3%)	12305 (37.5%)	25414	48%
13 To 15 Years	5442 (14.6%)	5725 (17.4%)	11167	51%
16+ Years	1133 (3%)	1724 (5.2%)	2857	60%
Employment	<i>X-squared = 2087.9, df = 3, p-value < 2.2e-16</i>			
Full Time	7963 (21.4%)	11566 (35.2%)	19529	59%
Part Time	3321 (8.9%)	3294 (10%)	6615	50%
Unemployed	13917 (37.5%)	8234 (25.1%)	22151	37%
Not in Labor Force	11952 (32.2%)	9762 (29.7%)	21714	45%

With regards to substance use in the current sample, 42% of the sample had a primary alcohol use disorder, 30% were primarily cannabis users, and 15% of the sample reported a primary cocaine use disorder. About 3% of the sample reported primary use of heroin, and another 3% reported primary use of opioids. Of note, 52% of individuals reported regularly using more than 1 substance at treatment admission. The majority of the study sample reported that their age of first drug use of the drug listed as “primary” was 17 or younger (7% at age 11 or under, 23% at ages 12-14, and 33% at ages 15-17).

	Did Not Complete	Completed	Total	% Completed
All Subjects	37153	32856	70009	47%
Primary Substance	<i>X-squared = 2398.5, df = 17, p-value < 2.2e-16</i>			
Alcohol	12936 (34.82%)	16400 (49.91%)	29336	56%
Cocaine	6486 (17.46%)	3725 (11.34%)	10211	36%
Marijuana	11581 (31.17%)	9317 (28.36%)	20898	45%
Heroin	2041 (5.49%)	491 (1.49%)	2532	19%
Methadone	104 (0.28%)	59 (0.18%)	163	36%
Other Opiates	1528 (4.11%)	858 (2.61%)	2386	36%
PCP	8 (0.02%)	3 (0.01%)	11	27%
Hallucinogens	24 (0.06%)	22 (0.07%)	46	48%
Meth	1609 (4.33%)	1397 (4.25%)	3006	46%
Amphetamines	166 (0.45%)	133 (0.4%)	299	44%
Stimulants	12 (0.03%)	13 (0.04%)	25	52%
Benzos	143 (0.38%)	150 (0.46%)	293	51%
Tranquilizers	12 (0.03%)	11 (0.03%)	23	48%
Barbituates	29 (0.08%)	27 (0.08%)	56	48%
Other Sedatives	94 (0.25%)	64 (0.19%)	158	41%
Inhalants	32 (0.09%)	29 (0.09%)	61	48%
OTC	25 (0.07%)	20 (0.06%)	45	44%
Other	323 (0.87%)	137 (0.42%)	460	30%
Age of First Use	<i>X-squared = 94.322, df = 5, p-value < 2.2e-16</i>			
11 and Under	2938 (7.92%)	1895 (5.77%)	4833	39%
12->14	9198 (24.79%)	7091 (21.61%)	16289	44%
15->17	11466 (30.9%)	11691 (35.62%)	23157	50%
18->20	5782 (15.58%)	6221 (18.96%)	12003	52%
21->24	2931 (7.9%)	2629 (8.01%)	5560	47%
25->29	2006 (5.41%)	1422 (4.33%)	3428	41%
30->34	1223 (3.3%)	799 (2.43%)	2022	40%
35->39	773 (2.08%)	500 (1.52%)	1273	39%
40->44	483 (1.3%)	327 (1%)	810	40%
45->49	235 (0.63%)	175 (0.53%)	410	43%
50->54	69 (0.19%)	69 (0.21%)	138	50%
55+	49 (0.13%)	37 (0.11%)	86	43%
Alcohol Use	<i>X-squared = 2591.7, df = 2, p-value < 2.2e-16</i>			
Alcohol Only	6522 (17.55%)	11274 (34.31%)	17796	63%
Other Drugs Only	14608 (39.32%)	10010 (30.47%)	24618	41%
Both	16023 (43.13%)	11572 (35.22%)	27595	42%

Comorbid drug and alcohol use disorders were quite prevalent, with 40% of the sample reporting combined alcohol and drug SUDs (35% reported only a drug SUD, and 25% only an alcohol SUD). Interestingly, drug and alcohol disorder comorbidity was higher in the group of individuals that did not complete treatment (43% of this subsample had comorbid drug and alcohol SUDs) as compared to those that completed treatment (35% of this subsample had comorbid SUDs), as was the number of substances reported at admission (59% of treatment non-completers were using 2+ drugs, whereas only 45% of treatment completers were). Overall, substance use in this population of individuals seeking treatment appears to be chronic, complex, and highly comorbid, three factors that increase risk of treatment termination or relapse and that make predicting treatment completion critical.

2.3. Statistical Models

Training, Testing, and Validation Data Split

In order to test the prediction accuracy of each of the statistical models developed in an unbiased manner (i.e. on data that was *not* used to train the models), the 70,009 observations included in the cleaned TEDS-D dataset were split into:

- Training data cohort (N= 45,000)
- Validation data cohort (N= 5,009)
- Testing data cohort (N= 20,000)

Each model was trained using the same training data observations, and model classification accuracy was tested on the left out testing data. The smaller validation cohort was used to tune neural network parameters and to report an additional, unbiased testing error for the final model.

Statistical Models

Regression-based Approaches

Three types of regression-based approaches were applied in this work. For each regression type, training data were used to fit the model. Misclassification error (MCE), area under the curve (AUC), and Akaike Information Criterion (AIC) were calculated using testing data to assess model performance.

Full Model Logistic Regression: A logistic regression model was fit with all possible predictors (R function `glm`) for the prediction of treatment completion versus non-completion. Beta coefficients were estimated, and the probability of treatment completion was calculated for each patient using the formula:

$$P(Y = 1|X) = \frac{e^{\beta_0 + \beta_1 * X_1 + \dots + \beta_N * X_N}}{1 + e^{\beta_0 + \beta_1 * X_1 + \dots + \beta_N * X_N}}$$

Individual probabilities were used to classify patients into completers (1) and non-completers (0) using a binary cut-off threshold probability=0.5.

Logistic Regression with Backwards Selection In an effort to create a more parsimonious and more interpretable model by removing predictors, `stepAIC` – an alternative to the function `bestglm` for model selection – was employed to perform iterative backwards selection, stopping with a local minimum AIC is reached. The final model included 26 variables.

Penalized Regression: Finally, we used elastic net regression, which, by adjusting the alpha, combines the shrinkage and sparsity of LASSO with the uniqueness and tolerance of collinearity of ridge regression. In other words, the elastic net penalized regression (`cv.glmnet` function) provides some balance between LASSO and ridge regressions. Our elastic net model was fit with coefficients extracted from both `lambda.min` and `lambda.1se`. Of these, the model with the lowest MCE and highest AUC was selected. This final model included 28 variables.

Random Forest

In considering the limitations of logistic regression, we investigated the potential benefit of data-driven classification with the R function `randomForest`.

Random forest provides **many layers of randomness** such that each tree is produced from a random sample of cases and split at a random sample of predictors. As such, this partitioning process not only avoids snooping/polluting the data, but also enables a sort of model selection even in instances where p is large. Perhaps more importantly, because different sets of predictors are evaluated for different splits a wide variety of mean functions are evaluated, each potentially constructed from rather different basis functions. This means we get a better fit, a **less restricted approximation of the true response surface**, and, resultantly, a smaller bias. Random forest also incorporates bagging – the idea that averaging across trees increases stability and that aggregating fitting values across trees makes them more **independent** while simultaneously reducing testing error.

Specifically, we used the training dataset to tune both `mtry` and `ntree` parameters, settling on an `mtry` of 5 (approximately equal to the suggested \sqrt{p} for classification trees) and `ntree` of 400. As such, the final random forest model included 400 trees; for which each node was split via one of 5 randomly selected variables. To assess the model's predictive accuracy, not only was OOB error assessed, but probabilities and responses were fitted using the testing data as to most appropriately report MCE and AUC. Sensitivity, specificity, as well as positive and negative prediction error were also assessed.

Boosting

Boosting is an extension of random forest, but rather than building a forest of *independent* trees using a set number of randomly chosen predictors per tree, each tree built during the boosting process attempts to *improve the prediction performance of the previously generated tree* (i.e. correct previous tree errors). As such, boosted trees are *non-independent*.

Boosting models improve probability and classification predictions by using an ensemble of decision trees that were generated (and iteratively improved upon) by modifying the weights of individual observations in the training data. More specifically, with boosting, trees fitted initially from equally or randomly weighted observations are improved upon for observations that were misclassified by increasing the weight of the misclassified observations. Over many iterations,

weak learning trees with small depth can thus be converted into strong learners by changing input observation weights and optimizing a specified metric (here, accuracy/AUC). Of note, whereas in random forest all generated trees are weighted equally when generating final predictions, trees built during boosting are given different weights in the final predictive model.

Here, the functions `trainControl`, `train`, and `gbm` from packages `caret` and `gbm` were used to develop an optimized boosting tree model.

- Boosting parameters were chosen by testing interaction depths of 2, 4, 6, 8, 10, 12, 14, and 16 and tree numbers from 100 to 2,000 (in increments of 100 trees), using a set shrinkage of 0.01 and a minimum of 20 observations per node.
- Parameters were tested using training data and cross-validation with five folds.
- The finalized parameters were: an interaction depth of 16, 2,000 trees, a shrinkage of 0.01 and a minimum of 20 observations per node.

Deep Learning (Neural Net)

To create a more sophisticated, more complex model allowing for multiple levels of non-linearity, we built a deep learning network (also known as a neural network model).

In deep learning, these layered representations are (almost always) learned via models called *neural networks*. Each layer in the network consists of a number of individual units, called “neurons”, and each neuron has the ability to perform a simple function on its inputs. In a dense layer, all of the input neurons connect to all of the output neurons with certain weights. During the course of training this network model, various parameters are adjusted, including the weights of the connections and the biases (effectively the intercept terms for the linear combinations).

Here, we use the package ‘keras’ to develop the deep learning model.

We chose to use a network structure that contained layers of varying size and the “relu” activation function, which transforms negative values into zeros. In order to avoid over-fitting the model, we added dropout layers between the “relu” layers. For the final layer, we chose to use a sigmoid activation function, which transforms the input information into a prediction from 0 to 1. In order to train the model, we chose to use ‘rmsprop’ as our optimizer logarithm, binary cross-entropy as our loss function (which the model will minimize), and the fraction of correctly classified patients as our metric of success. We first tuned the hyperparameters of the model (number of patients trained upon before changing the weights and total number of cycles through the data) by running our model on the training data and “validating” it on the validation data after every parameter change. In terms of the batch size, we wanted to train on a large enough set each time that the parameters would be tuned in the right direction. At the same time, having a very large batch size misses the point of dividing the data into batches, which is to avoid getting stuck in local maxima. To solve this optimization problem, we decided on a batch size of 200.

3. Results & Model Comparisons

3.1 Regression-based Approaches

In considering the balance between MCE and AUC, a logistic regression fit following automated stepwise, backwards tuning of AIC (`stepAIC`) was our best regression-based approach (MCE = 0.308; AUC = 0.756). However, it provided little benefit over regular `glm` or penalized regression (i.e., elastic net). Moreover, 20 of the 26 variables included in the regression-based model were significant at the $p = 0.001$ level. While this means that the TEDS-D dataset includes many helpful and informative predictors of treatment success, it also means that the final model is not particularly parsimonious and, therefore, not particularly relevant for informing targeted policy or intervention. Furthermore, while the regression-based models are easily implemented and easily interpreted, the assumptions that need to be met for regressions are strong and – since the true data generating model is unknown – these assumptions are usually violated.

3.2 Random Forest

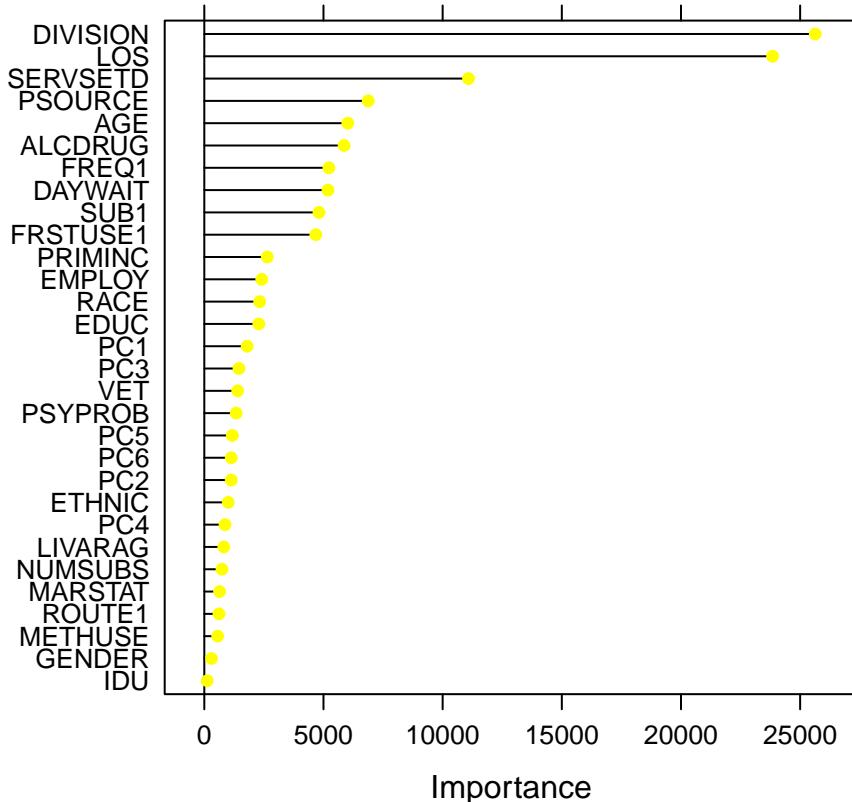
Prediction performance was substantially increased with respect to `randomForest`. Not only were we able to reduce MCE from 0.308 to 0.266, but we were also able to increase AUC from 0.756 to 0.812. Of note, the final random forest

model was also particularly balanced:

- Sensitivity: 77%; Specificity: 71%; Positive prediction error: 70%; Negative prediction error: 78%

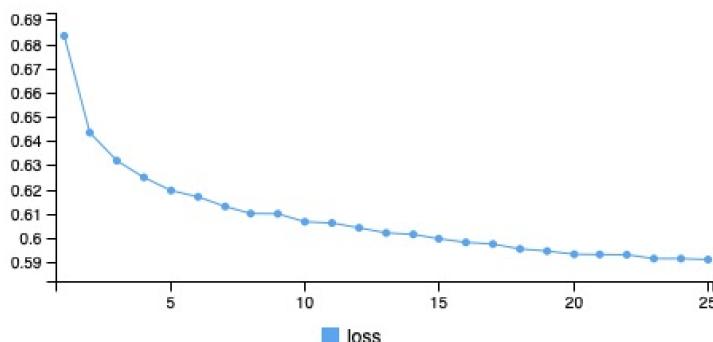
3.3 Boosting

For our boosting model, despite the expected advantage due to attempts by the boosting process to improve the prediction performance of the previously generated tree, the AUC (0.807) was slightly lower than that obtained with `randomForest`. Of note, both `randomForest` and boosting identified the variables `DIVISION` and `LOS` as most predictive of treatment outcomes.

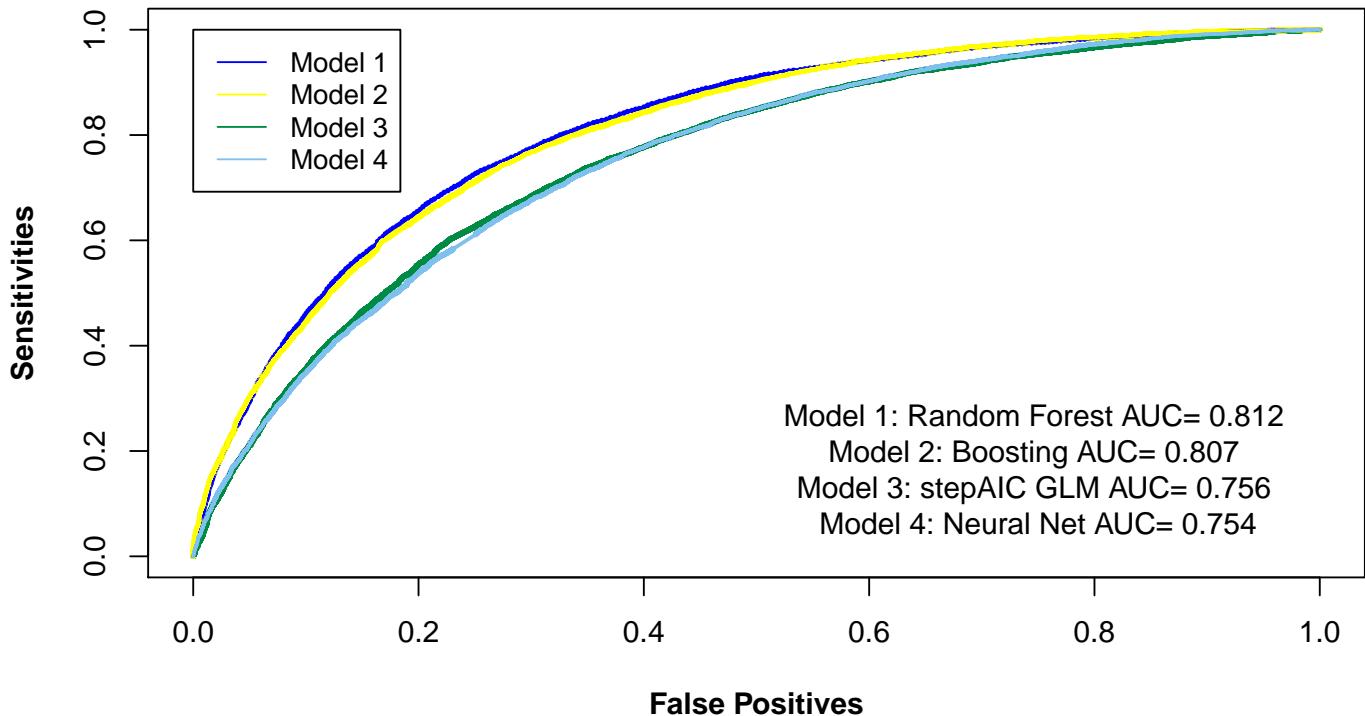


3.4 Deep learning

As seen in the plot below, our deep learning model revealed a decreasing loss (and ACC) over the selected 25 epochs. The model was able to correctly classify 68.5% of the data- significantly above average, but certainly not remarkable. MCE was 0.315 and AUC of 0.75 – falling below those of both random forest and boosting.



Model Comparisons: ROC Curves



4. Final Model

Using MCE and AUC as primary model selection criteria (plotted above), we chose random forest as our final model.

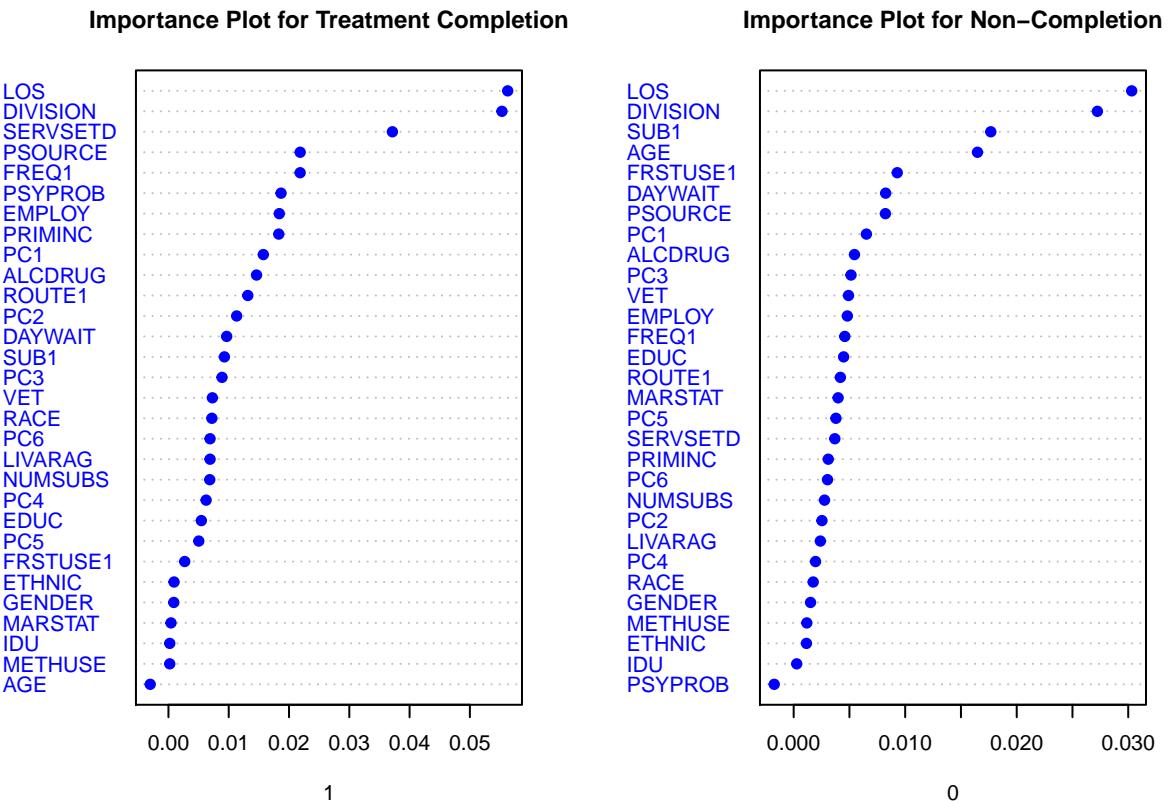
In order to further assess performance of this final model and to get a second, honest/unbiased estimate of testing error, we fit the chosen `mtry` and `ntree` parameters with the reserved validation data. While the MCE for this validation-based fit was comparable to the original model, the AUC actually increased to 0.814 – confirming the generalizability of the model.

After selecting the desired model we turn to interpreting its results. There are two main tools we will use to understand the model's output: variable importance plots and partial dependency plots.

Variable Importance Plots

One approach to using `randomForest` to get measures of predictor importance is to record the decrease in the fitting measure (e.g. Gini index; mean squared error) each time a given variable is used to define a split. In this sense, the sum of these reductions for a given tree becomes a measure of the importance of that variable when that tree is grown. In the context of random forest, this measure of variable importance can be averaged over the set of trees.

In the plots below, reduction in prediction accuracy is shown on the horizontal axis. As such, we can see reductions in prediction accuracy for successful treatment completion (left) and non-successful treatment completion (right) when each predictor is in turn randomly shuffled.



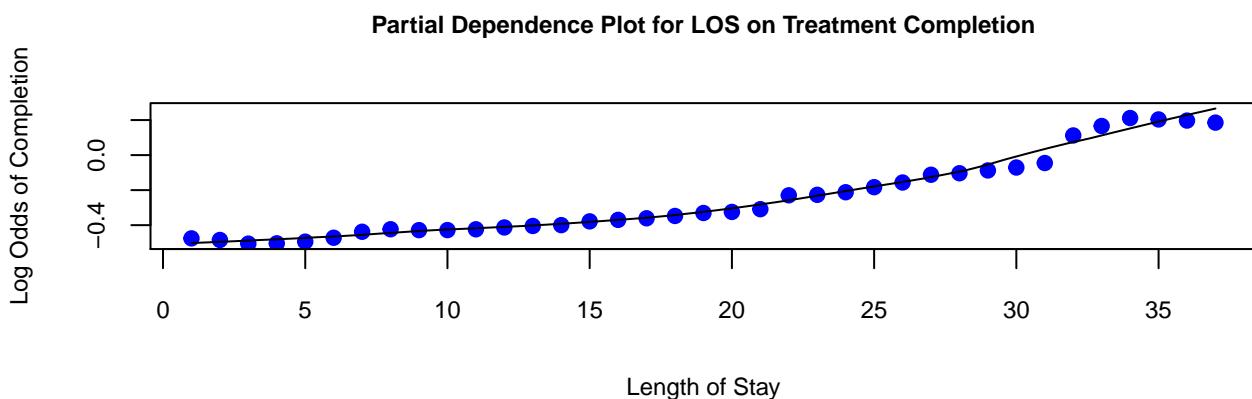
Specifically, we note that two variables **LOS** (length of stay) and **DIVISION** (census-based geographic location) are most predictive of both treatment completion and treatment non-completion.

Factors like **SERVSETD** (service setting at discharge), **PSOURCE** (principal source of referral), **FREQ1** (frequency of use), **PSYCPROB** (psychiatric problems/symptoms/diagnoses), **EMPLOY** (employment status at time of admission), and **PRIMINC** (source of income/support) are most predictive of successful treatment completion. On the other hand, factors such as **SUB1** (substance problem code), **AGE** (age at time of admission), **FRSTUSE1** (age at first substance use), and **DAYWAIT** (days spent waiting for admission) were most predictive of cases where treatment was not completed.

Partial Dependence Plots

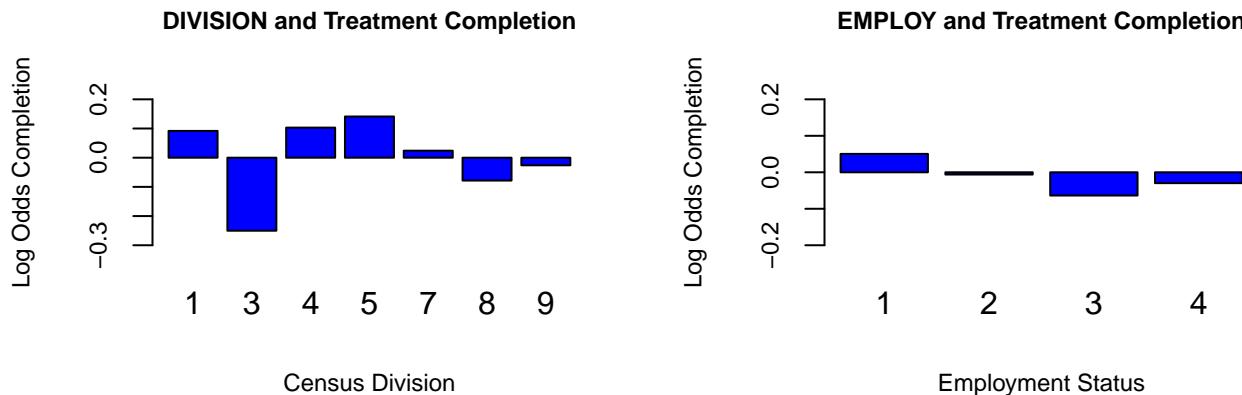
For tree-based approaches, these dependence plots show the average relationship between a given input and the response within the fixed, joint distribution of the other inputs. In other words, these plots show how what the proportion for a particular classification outcome would be for different values of the variable of interest.

Taking this approach for **LOS**, we see that the longer a patient stays in the treatment program, the higher their log odds of completing it becomes.



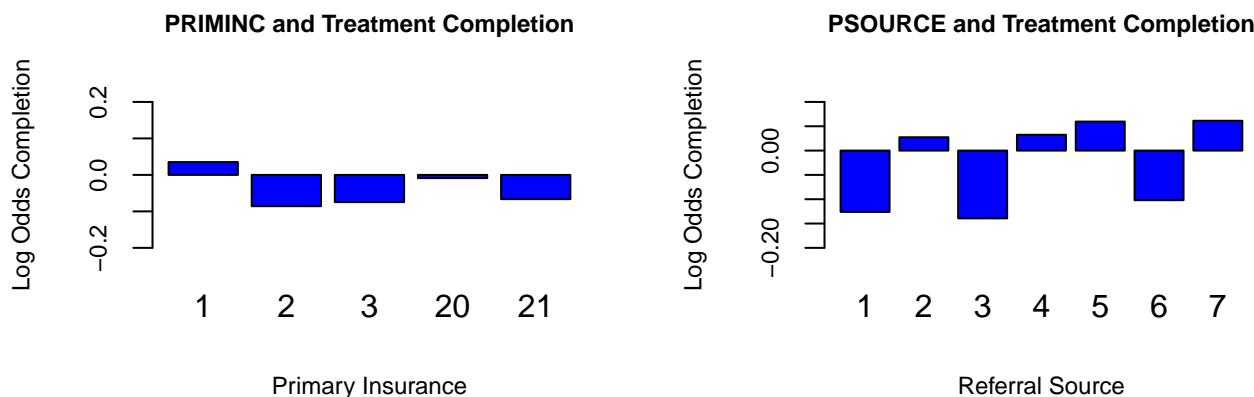
For DIVISION, we begin to see a clear effect of geographic location. Whereas being in treatment in New England (1), Central North West (4), or South Atlantic (5), seems to correspond to greater log odds of treatment completion for patients, being in treatment in – for example – the North East (3) or the Mountains (8) has the opposite effect. Intuitively, this makes sense as we consider factors like the prevalence and concentration of drugs in the North East counteracting treatment seeking and, perhaps, effecting relapse.

An interesting and intuitive relationship is present for EMPLOY, too. Having a full-time job (1) as you enter treatment is associated with increased log odds of completing the program whereas log odds of completion fall for patients who enter unemployed (3) or not in the labor force (4).



For PRIMINC (primary source of income), we also see the expected relationship – Patients earning wages/salary (1) having higher log odds of treatment completion whereas log odds are decreased for patients reliant on public assistance (2) or retirement/pension/disability (3).

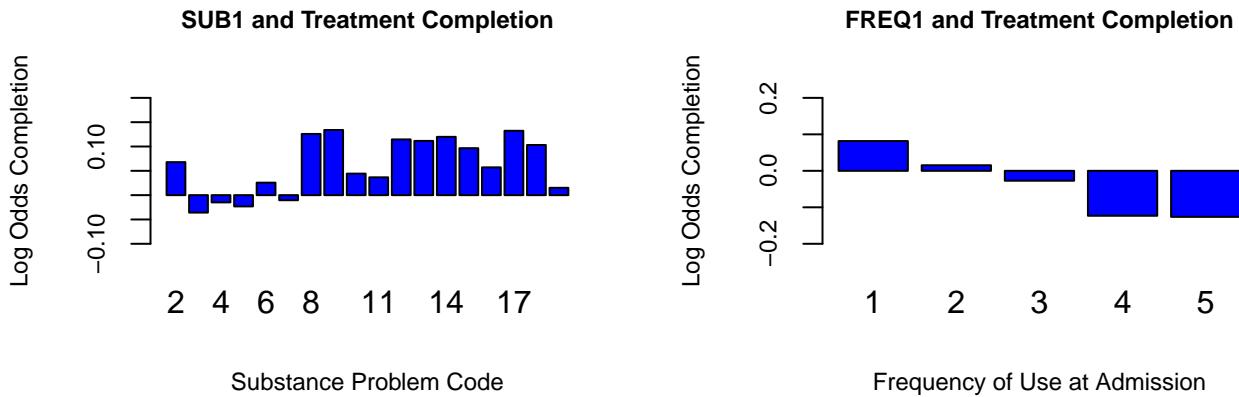
In addition to economic support, social networks appear to play a role in treatment completion. Here, we see the effects of PSOURCE (principal referral source). Interestingly, patients enrolled by themselves or by one other individual (1), healthcare provider (3) or member of the community (6) have lower log odds of completing. On the otherhand, referrals by drug/alcohol abuse mentors and providers (2), schools (4), employers (5), or the court/criminal justice system (6) have more positive effects.



Looking at predictors more directly related to substance use, we see that the drug of choice (SUB1) tends to have some effect on treatment outcomes. For example, individuals primarily using drugs like PCP (8), hallucinogens (9), methamphetamine/amphetamine (10/11), benzodiazepines (13), or inhalants (17) have higher log odds of treatment completion. Interestingly, drugs like cocaine (3), marijuana (4), and heroin (5) are associated with lower log odds of success. Perhaps this is due to the greater, more wide-spread availability of drugs like marijuana or cocaine such that participants see less value in sticking out the treatment plan if they know the drug will still be there when they return home.

Relatedly, the frequency of drug use (FREQ1) is also tightly related to success/failure of treatment. Specifically, individuals who had no use in the month prior to admission (1) displayed much higher log odds of completion than

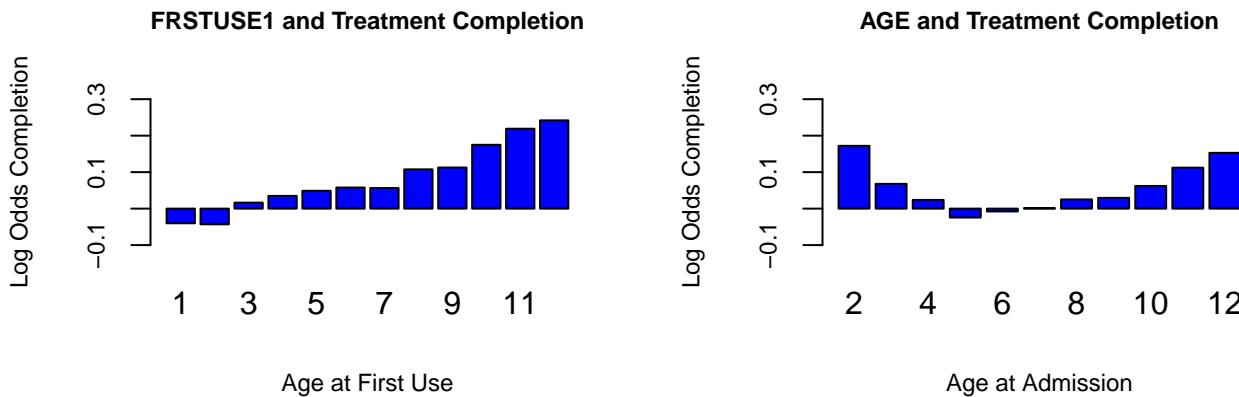
individuals who came into treatment using 3-6 times in the last week (4) or daily (5). Clinically, this checks out as the more rapidly and more frequently a patient is going through the binge/intoxication, withdrawal/negative affect, anticipation/craving, relapse cycle, the more difficult it become to break free from the chemical imbalanced induced by a hijacking of the brain's dopamine (neurotransmitter) system.



Finally, our model suggests clear age-based effects on treatment completion. We can see this on two fronts – age at first drug use (**FRSTUSE1**) and age at treatment admission (**AGE**).

We see an obvious relationship between **FRSTUSE1** and treatment success such that the younger an individual was when they began using drugs/alcohol (e.g. 1 → 11 years old vs. 12 → 55 and older), the less likely they are to complete the program. This relates back to the progression of addiction as a brain disease. The longer you suffer, the less likely you are to be able to return to self-sufficient, drug-free functioning.

However, the plot of **AGE** and log odds of treatment completion offers some hope. Here, we see that if a patient seeks treatment at these younger ages, their odds of finishing said treatment are positive. This suggests that it's less of game of age per-say and more dependent on the length of the time window between drug use initiation and treatment seeking.



5. Conclusions, Implications & Limitations

5.1 Conclusions & Implications

TEDS-D captures a significant share of all discharges from treatment facilities across the United States, especially those that reflect public spending. As such, the findings from our model and our analysis are – to an extent – generalizable. Given that every dollar spent on substance use disorder treatment saves 4 dollars in health care costs and 7 dollars in criminal justice costs, our analysis has far-reaching public policy and financial implications.

In considering the results discussed above, targeted intervention should manifest in a few ways.

First, as we saw through analysis of LOS and log odds of treatment completion, just getting people in the door increases their chances of finishing the treatment program. As such, increasing recruitment should be a top priority. Hiring social workers affiliated with the substance use treatment facilities who comb through the community or serve as supportive points of contact for struggling families is an excellent starting point. More importantly, seeing as the longer you spend in the program, the better your outcome, it is absolutely critical that policy targets major insurance companies to force funding of these treatment programs for longer than 2-4 weeks (as is current status).

We also noted a significant effect of geography on treatment completion. In light of diminishing completion rates in the North East where the magnitude of the drug problem might be greater, finances and resources should be concentrated around areas like Boston, New York, and Philadelphia. At the same time, funding and infrastructure should be delegated to mountainous regions of the country where treatment completion also suffers – perhaps because more isolated conditions are associated with poorer treatment facilities/programs and thus poorer completion rates.

On top of geography, we noted that employment and economic status are particularly predictive of treatment completion outcomes. Since it appears as though having a full-time position and a salary (and perhaps an associated personal incentive to discontinue using) sets one up for a greater chance of treatment completion and SUD recovery, it makes sense to establish funding initiatives or non-profits that work to establish short-term (or, ideally, longer-term) employment opportunities for patients who complete the program. In this sense, not only will the individual's time – which might have contributed to drug seeking – be occupied following discharge, but completing the program now has some sort of a financial incentive. One model that has seen tremendous success lately is North Carolina's Triangle Residential Options for Substance Abusers (TROSA). This organization contracts out its members to moving companies, thrift stores, etc. not only to subsidize the cost of their treatment, but also with the idea that – following program completion – they have a resume and a chance at future, stable employment opportunities. A model like this is certainly extendable to the North East and elsewhere. Additionally, since log odds were particularly low among individuals receiving disability/pensions, perhaps requiring a drug screening or a drug-based online course prior to receiving that money could be helpful in targeting that specific demographic.

Another critical aspect of the economic side to substance use and abuse is self-medication. It is well-known that there is substantial comorbidity between SUDs and other psychiatric disorders, and individuals with these disorders often choose to self-medicate in attempts to treat, reduce, or ignore psychiatric symptoms. Furthermore, individuals with psychiatric disorders that are associated with increased impulsivity, e.g. Bipolar Disorder, Psychotic-Spectrum Disorders, Borderline Personality Disorder, and Conduct Disorder, have a higher propensity to first try and to repeatedly use substances of abuse. These phenomena were present in our sample; having a psychiatric diagnosis was the 6th most predictive factor of treatment completion. Consequently, it is critical that SUD treatment programs integrate substance and psychiatric treatment, and that doctors in these programs make identifying, diagnosing, and properly treating comorbid psychiatric conditions a top priority. Integrating these two treatments will require a collaborative effort between psychiatrists, psychologists, social workers, treatment counselors, and/or case workers; such an effort is likely to increase patient retention by decreasing mental health issues that lead to drug seeking/using/relapse, by creating a more personalized, multifaceted treatment strategy, and by increasing patient interaction with staff and individuals of support – and thus hopefully increasing engagement and retention.

Last but not least, policy informed by our model and our analysis should continue to support and to target interventions through the avenues we know to be working: employers, the courts/criminal justice system, drug/alcohol abuse mentors, and – in particular – schools. Since we know patients in their teens and early twenties to have the lowest log odds of treatment completion, but, at the same time, to be among the most amenable to positive treatment outcomes if admitted early, publicizing and de-stigmatizing these programs is a necessity.

5.2 Limitations

Although the study utilized a large, geographically diverse administrative dataset of annual discharges, it has several limitations.

Facility-level characteristics, such as type of ownership (i.e. public or privatized), size of the facility, patient-to-provider ratio, service offerings, and other characteristics can play an important role in influencing treatment completion (Arndt et al., 2013). However, these facility-level variables were not available in the TEDS-D and, therefore, were not included in our analysis.

Error associated with self-report cannot be accounted for in the dataset. There is no real incentive for patients to be transparent or forthcoming with respect to the drugs/substances playing a role in their admission, the frequency of

their substance use, etc. In this sense, our analysis hinges on highly-personal information that, in many cases, cannot be verified.

All data included in this analysis came from 2006. The accuracy and fluctuation of substance use and abstinence data would be more appropriately and more fully captured by long-term/longitudinal reports extending far beyond the treatment milieu.

Critically, despite the flexible way in which our final model (`randomForest`) and associated **CART** algorithms can respond to data, substantial bias is still a real possibility. At the end of the day, the algorithms are trying in a single-minded manner to use associations in the data to maximize the homogeneity of data partitions. How those associations come to be represented have no foundation in subject matter understanding. While it is possible to peer into the black box, no matter how much information you can extract, `randomForest` remains a prediction tool and not a descriptive tool. Moreover, portions of our analysis rely on variable importance plots in which only one variable can be shuffled at a time. As a result, there is no consideration of joint importance over several predictors. This can be particularly detrimental when variables are correlated (i.e. there is a contribution to prediction accuracy that is uniquely linked to each predictor **and** joint contributions shared between two or more predictors).

Finally, in 2006 the main drugs of use/abuse were alcohol, cannabis, and cocaine. This is in stark contrast to where we stand today admits our nation's largest opioid epidemic. In recognizing that the number of patients entering substance use disorder treatment programs for opioid-related addiction, it is both possible and likely that predictors of treatment completion may be different among this population. As a result, it is difficult – if not impossible – to extend or predict the performance of our model relative to this change in substance use patterns.

Nonetheless, our final model enables data-driven discussions that have the potential to design interventions specific to improving treatment outcomes for patients struggling with SUD. It, combined with our overarching and well-supported analysis, paves the way for future research needed to inform specific interventions and policy designs that will be most effective in encouraging treatment completion for SUD.