

# **Factors associated with opioid cessation: a machine learning approach**

Jiayi W. Cox<sup>1, 7</sup>, Richard M. Sherva<sup>1</sup>, Kathryn L. Lunetta<sup>2</sup>, Richard Saitz<sup>3</sup>, Mark Kon<sup>4</sup>,  
Henry R. Kranzler<sup>5</sup>, Joel E. Gelernter<sup>6</sup>, Lindsay A. Farrer<sup>1</sup>

<sup>1</sup> Section of Biomedical Genetics, Boston University School of Medicine, Boston, MA, USA

<sup>2</sup> Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA

<sup>3</sup> Department of Community Health Sciences, Boston University School of Public Health, Boston, MA, USA

<sup>4</sup> Department of Mathematics and Statistics, Boston University College of Arts & Sciences, Boston, MA, USA

<sup>5</sup> Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

<sup>6</sup> Department of Psychiatry, Yale School of Medicine, West Haven, CT, USA

<sup>7</sup> Department of Genetics and Genomics, Boston University School of Medicine, Boston, MA, USA

## Abstract

**Background and Aims** People with opioid use disorder (OUD) can stop using opioids on their own, with help from groups and with treatment, but there is limited research on the factors that influence opioid cessation. **Methods** We employed multiple machine learning prediction algorithms (LASSO, random forest, deep neural network, and support vector machine) to assess factors associated with ceasing opioid use in a sample comprised of African Americans (AAs) and European Americans (EAs) who met DSM-5 criteria for mild to severe OUD. Values for several thousand demographic, alcohol and other drug use, general health, and behavioral variables, as well as diagnoses for other psychiatric disorders, were obtained for each participant from a detailed semi-structured interview. **Results** Support vector machine models performed marginally better on average than those derived using other machine learning methods with maximum prediction accuracies of 75.4% in AAs and 79.4% in EAs. Subsequent stepwise regression analyses that considered the 83 most highly ranked variables across all methods and models identified less recent cocaine use ( $p < 5 \times 10^{-8}$ ), a shorter duration of opioid use ( $p < 5 \times 10^{-6}$ ), and older age ( $p < 5 \times 10^{-9}$ ) as the strongest independent predictors of opioid cessation. Factors related to drug use comprised about half of the significant independent predictors, with other predictors related to non-drug use behaviors, psychiatric disorders, overall health, and demographics. **Conclusions** These proof-of-concept findings provide information that can help develop strategies for improving OUD management and the methods we applied provide a framework for personalizing OUD treatment.

**Key words:** opioid use disorder, opioid cessation, opioid abstinence, machine learning, AI, diagnostic questionnaire, feature selection, outcome prediction

## Introduction

Misuse of prescription opioids and illicit opioids are a significant global problem that affects the health and economic welfare of individuals, families, and society. The opioid overdose rate has quadrupled since 1991(1). In 2017, more than 47,000 Americans died of an opioid overdose, and 36% of these deaths involved prescription opioids(2). A major goal of addressing opioid use disorder (OUD) is to achieve abstinence, or cessation, from illicit use and misuse. There is not a single, clinically accepted definition of cessation with regard to length of time without illicit use or prescribed opioids misuse. (3, 4). Previous studies identified age, employment status, and age at first drug use to be associated with successful completion of a treatment program (3). Other factors are likely to influence successful cessation. Based on studies of OUD severity (5-7), pain experiences, general health, use of antidepressants may also influence cessation. Delineation of these factors could inform efforts to help patients or be useful for patients to practice self-care. However, studies thus far have tended to focus on a small number of clinically relevant factors such as medication dose, duration and formulation (8-10). There are currently much larger epidemiological studies of OUD (11-13) with data on thousands of potential predictors that would enable a systemic hypothesis-free query to identify factors predicting opioid cessation.

Many statistical methods are limited in their ability to sort through large numbers of predictors(14). Data mining using machine learning, which is particularly well suited for identifying predictive factors among thousands of variables (15, 16), has successfully identified predictor variables for a diverse set of outcomes(17-21). Here, we applied multiple machine learning techniques to evaluate a large set of clinical, demographic, general health, and behavioral variables in a large, racially mixed cohort who were recruited for genetic studies of substance dependence and not necessarily treated for

OULD to identify factors that influence opioid cessation (defined as self-reported last illicit opioid use and/or prescription opioids misuse > 1 year before the interview date).

Identifying additional factors associated with cessation could support a personalized approach to improve the outcome of cessation attempts.

## **Material and Methods**

*Participants and Assessments.* Participants for this study were selected from a cohort of 6,188 African Americans (AAs) and 6,835 persons of European Americans (EAs) who were recruited for genetic studies of opioid, cocaine, or alcohol dependence between 2000 and 2017 through advertisements and treatment clinics at the University of Connecticut Health Center, Yale University School of Medicine, the Medical University of South Carolina, the University of Pennsylvania, and the McLean Hospital(22, 23). This cohort included affected sibling pairs and additional family members, as well as unrelated cases and controls. Persons with schizophrenia or schizoaffective disorder were excluded(22, 23). Information about use of various substances, demographics, general health, behavior, and other psychiatric illnesses was obtained by interview using the Semi-Structured Assessment for Drug Dependence and Alcoholism (SSADDA) (11, 24). Substance use disorder (SUD) and psychiatric disorders were established according to DSM-IV criteria. Institutional review boards from each recruitment site and Boston University approved this study, and written informed consent was obtained from all participants.

*Opioid Cessation Definition.* Participants who were eligible for this analysis met at least two DSM-5 criteria for OUD, corresponding to a lifetime diagnosis of mild to severe OUD. Current opioid cessation was determined by the response to the question, “When was the last time you used an opioid drug (including illicit methadone).” This question was asked as part of a series of items asked about illicit or non-prescribed use of

opioids. Individuals who last used an opioid >1 year before the date of interview were considered to have achieved cessation and those whose last use of an opioid was <6 months before the interview date were classified as non-cessation. Persons who used opioids between 6 months and 1 year before the interview date were excluded from further analysis. Filtering steps yielding a sample of 1,192 AAs and 2,557 EAs for analysis is shown in Figure S1.

*Phenotype Data Processing.* Preprocessing of 3,956 SSADDA variables was performed prior to machine learning analyses. Variables with narrative or invariable responses, containing redundant information (e.g., specific date of different episodes, drug names), and with a response rate <90% were removed. Missing values for binary and categorical variables were recoded as indicator variables to accommodate missing responses. Missing values for continuous variables were imputed to the population group mean value. Missing values for ordinal variables related to recency of drug use were coded as dummy variables to account for current or never use. Z-score normalization (mean of 0 and variance of 1) was applied to continuous variables within each population to minimize scaling issues. The number of variables remaining after these steps was 3,315 in AAs and 3,738 in EAs.

*Machine Learning Analyses.* AAs and EAs were analyzed separately because they had different SSADDA completion rates and combining the two groups might lead to response bias. Variables were grouped into three nested sets defining three analytical models. Model 1 contained all variables except those related to recency of opioid use which are strongly correlated with cessation, thus with variable totals of 3,093 in AAs and 3,503 in EAs. Model 2 further excluded all opioid-related questions, leaving 2,863 variables in AAs and 3,252 in EAs. Model 3 further excluded all drug use variables, leaving only demographic, non-SUD diagnoses, behavioral, and general health variables

(1,656 in AAs and 1,907 in EAs). Models were evaluated using four machine learning methods described in Supplemental Materials to identify variables predicting opioid cessation. We modeled different types of intra-variable relationships between predictors and the outcome using linear (LASSO(25) and linear support vector machine with recursive feature elimination [SVM](26)) and non-linear (random forest with recursive feature elimination [RF](27) and deep neural network with feature selection [DNN] (28)) techniques. These four methods were applied to capture predictive variables under different model assumptions and allow for different outcome-predictor relationships. Variables from each model that were associated with the highest accuracy reflected by either F1 score or area under the curve (AUC) and generated by each machine learning method were retained. The F1 score is a harmonic measure of precision (true positive / [true positive + false negative]) and recall ( true positive / [true positive + true negative] ), defined by  $2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$  at a given case/control split, and AUC is an overall evaluation of model performance that accounts for the true positive and false positive rates for all possible diagnostic splits(29, 30). Both measurements were considered because of their popularity in the clinical setting(31). The F1 score was used to assess accuracy due to limitations of the AUC, which includes bias when performed on imbalanced datasets as well as impractical and uninterpretable split points for evaluation(29, 32).

### Statistical Methods for Testing the Association of Opioid Cessation with Phenotypic

Variables. To determine which variables selected by the machine learning methods independently predict cessation, we applied different cutoffs for the importance measurement of each machine learning method, namely the odds ratio (OR) for LASSO, coefficient(33) denoted by weight for SVM, feature importance(33) for RF, and activation potential(28) for DNN. For LASSO, we chose variables that yielded ORs >1.05 or <0.95.

We applied the following criteria for selecting variables from SVM and RF analyses depending on the number of variables (n) selected for each model: if  $n > 200$ , the top 30% of variables measured by absolute weight in SVM or importance in RF were designated as high impact; if  $100 < n < 200$ , the top 50% were selected; if  $n < 100$ , all variables were designated as high impact. For DNN, all selected variables were designated “high impact”. Joint association tests were performed using bi-directional stepwise logistic regression that included 83 “high-impact” variables that were culled from three models across four machine learning methods in the AA and EA datasets. Variables that yielded the highest Akaike information criterion (AIC) with  $p < 0.05$  from bi-directional stepwise logistic regression were grouped into “drug related”, “behavioral”, “general health”, and “demographic” categories.

## Results

Characteristics of the study samples are shown in Table 1. The sample included 1,069 unrelated AAs and 2,252 unrelated EAs, as well as 123 AAs and 305 EAs participants who were members of families containing a pair of siblings both with either opioid or cocaine dependence. There is a higher proportion of females among individuals who ceased opioids in both AAs ( $OR=1.35$ ,  $P=6.7 \times 10^{-3}$ ) and EAs ( $OR=1.31$ ,  $P=1.1 \times 10^{-3}$ ) compared to those who did not cease. Furthermore, participants who ceased opioid use were older by an average of 3.18 years in the AA group ( $P=1.0 \times 10^{-10}$ ) and 6.1 years in the EA group ( $P=2.2 \times 10^{-16}$ ) than those who did not cease. The mean number of lifetime DSM-5 OUD criteria met among those who ceased opioid use was not significantly different from those who did not cease.

### Feature selection

The F1 score was generally higher across models in both AAs and EAs using SVM than

the other machine learning algorithms (Figure S2), although the differences in F1 score across methods were generally small, especially for models 1 and 2. A detailed discussion of the performance of each method for the three models is provided in Supplemental Materials and Figure S2.

Figure S3 shows the overlap of high impact variables chosen by the four machine learning methods. LASSO “high impact” variables almost entirely overlap with those from the other methods, while DNN-selected variables overlap the least with other method-selected variables. The majority of variables selected by non-LASSO methods are unique to those methods, however, there was high overlap in “high impact” variables selected by SVM and RF. Age was among the five top-ranked variables consistently identified by each method for each model in both AAs and EAs (Table S1).

### *Factors Associated with Opioid Cessation*

Stepwise regression analysis that considered independent predictors among 83 “high impact” variables culled from all models and machine learning methods (Table 2) revealed that age was one of the most significant predictors of opioid cessation in both AAs (OR=2.44 per standard deviation (SD) change,  $P=1.41 \times 10^{-12}$ ) and EAs (OR=2.00 per SD change,  $P=5.74 \times 10^{-9}$ ). Variables related to drug use comprised over 50% of the nominally significant predictors of opioid cessation in AAs (29 of 41) and EAs (27 of 50). Recency of last cocaine injection in AAs (OR=2.30 per level change,  $P = 9.11 \times 10^{-6}$ ) or use in EAs (OR=1.91 per level change,  $P=3.30 \times 10^{-15}$ ) were also among the most significant predictors. The number of years using heroin in AAs (OR=0.55 per SD change,  $P=5.78 \times 10^{-6}$ ) and age at first heavy opioid use in EAs (OR=0.56 per SD change,  $P=2.67 \times 10^{-12}$ ) were also significant predictors. Greater odds of cessation in AAs were associated with variables related to use of other drugs including “less recent of having last alcohol symptoms” (OR=1.45 per level change,  $P=2.84 \times 10^{-3}$ ) and “had 2



marijuana symptoms lasting >1 month” in AAs (OR=2.13,  $P=4.83 \times 10^{-3}$ ). Drug use-related variables associated with cessation in EAs included less recent of last having alcohol symptoms (OR=1.34 per level change,  $P=4.13 \times 10^{-5}$ ), “able to cut down smoking” (OR=1.28,  $P=3.69 \times 10^{-2}$ ), and “marijuana interfered with work or home activities” (OR<sub>EA</sub>=1.67,  $P_{EA}=1.61 \times 10^{-3}$ ). The Fagerstrom Test for Nicotine Dependence (FTND) item “smoked more frequently after waking up” (OR=0.57,  $P=7.76 \times 10^{-3}$ ) and having attended a self-help group for OUD (OR=0.57,  $P=1.41 \times 10^{-2}$ ) were associated with decreased odds of cessation in AAs, whereas in EAs, starting attendance at an OUD self-help group sooner (OR=1.28 per level change,  $P=2.4 \times 10^{-3}$ ) and recent attendance at a cocaine self-help group meeting (OR=0.76 per level change,  $P=3.27 \times 10^{-4}$ ) increased the odds of cessation.

Among behavior-related variables, suicidal ideation or self harm was significantly associated with higher odds of cessation in both AAs (OR=1.39,  $P=1.96 \times 10^{-2}$ ) and EAs (OR=1.2,  $P=8.08 \times 10^{-3}$ ). In EAs, recovering from an event causing PTSD assessed by the question “No fear in most disturbing event” (OR=1.93,  $P=1.66 \times 10^{-6}$ ) and no recent antisocial behavior episodes (OR=1.35 per SD change in age,  $P=1.03 \times 10^{-4}$ ) were also associated with increased odds of cessation.

Several other health-related variables including past depressive episode lasting >1 week in AAs (OR=1.31,  $P=1.66 \times 10^{-2}$ ), having drug-use associated depression in EAs (OR=1.64,  $P=2.55 \times 10^{-3}$ ), body mass index in EAs (OR=1.32 per SD change,  $P=2.58 \times 10^{-6}$ ), and being HIV positive in AAs (OR=2.47,  $P=1.39 \times 10^{-3}$ ) were significantly predictive of cessation, whereas asthma (OR=0.68,  $P=1.22 \times 10^{-2}$ ) significantly lowered the odds of cessation in EAs.

Other significant predictors of opioid cessation that were identified by this analysis include being employed (OR=1.84,  $P=1.82 \times 10^{-2}$ ) and not having been raised by a single

parent (OR=1.58,  $P=1.3 \times 10^{-2}$ ) in AAs, and lower household gross income (OR=0.87 per level change,  $P=1.3 \times 10^{-3}$ ) and being an atheist (OR=1.45,  $P=1.34 \times 10^{-2}$ ) in EAs.

## Discussion

We employed both regression and non-regression based machine learning approaches to evaluate the association of more than 3,000 SUD, other psychiatric disorder, other health, and demographic variables with opioid cessation among EAs and AAs assessed in a cohort study of opioid, cocaine, or alcohol dependence. We observed moderate to high predictive accuracy across all methods, noting that SVM on average marginally outperformed the other methods. Although the specific set of predictive variables differed in EAs and AAs, a common profile emerged. People who ceased opioid tended to be older, initiated drug use later in life, had used opioids for a shorter period, experienced fewer problems related to cocaine or alcohol use, were currently employed, and had recovered from other psychiatric disorders including depression and PTSD.

Previous research using machine learning for addiction outcomes focused mainly on predictive accuracy, although a few studies attempted to identify and interpret specific variables that were associated with the outcomes(3, 4, 34, 35). Acion et al.(3) reported that ensemble super learning was superior to other machine learning methods, including penalized regression, SVM, and neural networks for predicting SUD treatment success indicated by treatment discharge status in a Hispanic cohort. In that study, less than 10% of participants had problems with cocaine or illicit opioids and fewer than 30 potential predictors were assessed. In contrast, we evaluated several thousand predictors, including detailed measures of drug-use activities and psychiatric disorders, and ranked the importance of the top-ranked variables with four distinct machine learning algorithms. Gowin et al.(3) identified regional brain activity changes predicting relapse from imaging data on fewer than 70 methamphetamine-dependent patients

without including any lifestyle factors. Che et al.(4) applied deep learning to electronic health record data to predict people who used opioids short term, long-term, or with dependence. Similar to our study, they identified comorbid substance use and anxiety disorders as predictors(4). Several other studies used only regression-based methods to predict opioid and stimulant dependence(36), cocaine dependence(37) and alcohol dependence(38), which might not capture other relationships among variables. Several of the non-regression based methods we employed have also been applied in other studies, which focused mainly on MRI brain images as predictors of substance use disorder diagnoses(18, 19, 21, 34).

We identified several predictors of opioid cessation that were previously associated with other SUDs, including co-morbid drug use, antisocial behavior, suicidal thoughts, HIV infection, and asthma. Our finding that the majority of people who ceased opioids (60% in AAs and 66% in EAs) also ceased cocaine is consistent with evidence of high rates of co-occurrence of OUD and cocaine use disorder (CUD)(39, 40), which supports the development of treatment strategies to target both disorders (39, 41), and suggests that ceasing one substance might influence, or reflect, the ability to cease the other. Our findings are also consistent with observations that failure to address tobacco use lowers the efficacy of opioid cessation treatment(42) and that a behavioral intervention in patients with antisocial personality disorder reduces substance use(43). Unlike problems that are associated with other drug use, which predicts lower odds of opioid cessation, we found having cannabis use related problems predicts higher odds of opioid cessation (i.e had two marijuana symptoms lasting a month; marijuana interferes with work). This finding is puzzling and not immediately explainable. Previous findings of the co-occurrence of drug addiction, suicide attempts, depression, family conflicts, and PTSD, which may suggest bi-directional casual relationships,(44-49) are consistent with

our observation that better management of comorbid psychiatric problems (fewer recent suicide attempts and psychiatric symptoms) increases the likelihood of opioid cessation.

One explanation for our findings of significant associations of ceasing opioids with a self-reported HIV diagnosis in AAs and asthma in EAs is that OUD patients with severe or life-threatening illnesses are more likely to seek or adhere to treatment,(50) an idea supported by evidence that HIV-infected OUD patients have better treatment outcomes for OUD.(51-53) Alternatively, poorer general health may lead to reduced drug use(54) (the so-called “sick quitter”).

We and Acion et al.(3) identified age, employment status, and age at first drug use as factors for treatment success. The protective effect of older age may be due to ascertainment bias because persons who survived severe dependence are more likely to have stopped using opioids. Further, our ascertainment strategy likely led to a higher proportion of older individuals and females who are in recovery from substance use disorders. Full-time employment likely reduces the time or urge for persons dependent on opioids to seek and use the drug. In addition, drug screening associated with some jobs may reduce the likelihood of current opioid use(55). The inverse correlation of age at first drug use and opioid cessation may reflect the increased difficulty to reverse the effect of long-term opioid exposure on the brain reward system(56), or increased severity associated with earlier onset. Other novel predictors of successful opioid cessation identified in this study among EAs include lower household atheism and income. The explanations underlying these associations are unclear, although a previous study showed that loss of religiosity between childhood and adulthood was associated with increased substance use while recent religiosity increased the odds of illicit drug use in the past year(57). In addition, lower household income might affect one’s ability to get opioids.

Implications can be drawn in terms of the overall accuracy from machine learning methods and individual factors identified. (A detailed discussion of the performance of the machine learning methods is provided in Supplemental Materials). Unlike standards established by a cardiologist reading an EKG or a radiologist reading a lung nodule, the accuracy as high as ~80% from our model has few comparable standards made by a physician specialized in substance misuse treatment. However, our accuracy result could serve as a reference for developing ways to address prescription opioid misuse and illicit opioid use, since we identified a solid list of individual variables to consider. While some of the factors identified were plausible and consistent with prior studies, factors like atheism and racing cars were less interpretable. Given that our research is atheoretical, results should be interpreted with caution and be validated.

The current study has several strengths. First, because the input dataset contains thousands of variables related to drug use activities, psychiatric disorders, medical history, and demographics obtained from several thousand individuals meeting DSM-5 criteria for mild to severe OUD, we were able to explore many factors in addition to those included in other studies. Second, both linear and non-linear machine learning methods were employed to best approach the true underlying model, which likely increased the robustness of our association findings. Third, we evaluated three models for each machine learning method to better understand the contribution of opioid and other drug use information. Finally, we considered only independent predictors in the association analyses to prevent over-representation of correlated factors.

Limitations of this work should be noted. First, given the cross-sectional nature of our data and the over 90% relapse rate(58) for OUD, many individuals classified as not using opioids may have subsequently relapsed to opioid use. However, it has been shown that prior abstinence is predictive of future abstinence.(59) Second, the machine

learning analyses were based on samples that may have been underpowered to detect associations with some variables compared to other studies that included tens of thousands of individuals(60). Third, most persons in our cohort were evaluated prior to the current opioid epidemic and may not reflect recent secular trends in the prevalence and associated features of OUD. Fourth, associations of cessation with some variables and overall prediction accuracy may have been inflated because our analysis did not fully account for familial correlations. Fifth, in spite of the large number of variables that were included in the machine learning analyses, potentially important variables such as the reason for first use and details of treatment and support programs were unavailable. Finally, the rate of responses to many interview questions was substantially higher in EA participants, which may account for some of the observed racial differences in predictive models. Given these limitations, our findings require external validation in larger samples before they can be incorporated in prediction models for clinical purposes.

## **Conclusions**

Using machine learning techniques with feature selection, we analyzed a large number of variables include demographic, behavioral, health and drug use activities and found variables in a wide range of domains that were associated with cessation. These included some that are consistent with prior literature, others that plausibly could be associated but have not been well-studied, and others that do not have readily apparent explanations for their associations. These findings suggest hypotheses for future studies and could inform how one might increase the likelihood of cessation with and without treatment. The findings support a number of widely known treatment strategies for OUD, such as treating psychiatric comorbidity, adding wraparound services like employment counseling, and simultaneously addressing polydrug use problems. Finally, in an era of increasing availability of digitized health-related records, our study provides a framework

for disease outcome prediction using high dimensional health-related records.

**Table 1.** Participant characteristics.

		Time since last use	
		≤ 6 month (not cease)	>1 year (ceased)
AAs (N=1192)	Total N (% female)	701 (33.5%)	491 (40.5%)
	Age (Mean ± SD)	42.6 ± 8.5	45.6 ± 8.3
	OUD Symptom Counts (Mean ± SD)	7.8 ± 2.4	7.6 ± 2.5
	# of families (N in families)	35 (76)	23 (47)
EAs (N=2557)	Total N (% female)	1714 (34.4%)	843 (40.6%)
	Age (Mean ± SD)	34.4 ± 10	40.5 ± 10.3
	OUD symptom Counts (Mean ± SD)	8.8 ± 1.9	8.4 ± 2.3
	# of families (N in families)	114 (241)	31 (64)

AAs: African Americans, EAs: European Americans, OUD: opioid use disorder



Table 2. Stepwise regression results using all significant variables from all machine learning methods and models in (A) African Americans and (B) European Americans.

A) African Americans

	Variable	OR	p-value
Drug related	Less recent since 1st opioid treatment	1.56	1.90E-04
	Older age at first opioid misuse <sup>§</sup>	1.4	2.45E-02
	Older age at first opioid symptoms <sup>§</sup>	0.46	2.23E-05
	Diarrhea after stopping opioid use	1.54	3.68E-02
	Attend opioid self-help group	0.58	1.41E-02
	Older age at first heavy use of opioids <sup>§</sup>	0.73	2.70E-02
	N years using heroin <sup>§</sup>	0.55	5.78E-06
	Depressed after reduced cocaine use	0.53	4.05E-03
	Less recent since last injected cocaine*	2.30	9.11E-06
	Less recent since last used cocaine*	1.82	9.19E-05
	Less recent since last stayed high in cocaine*	1.41	2.93E-03
	Used cocaine <11 times within year of interview	2.67	1.38E-03
	Treated in outpatient program for cocaine use	1.88	4.06E-03
	Less recent since of first cocaine craving*	0.71	1.59E-03
	Never injected cocaine	2.53	1.75E-03
	Often used marijuana more than intended to	0.40	5.67E-04
	Mixed alcohol with drug >3 times in 12 months	0.51	2.08E-03
	Less recent since last had alcohol symptoms lasting >1 month*	1.45	2.84E-03
	Drinking interfered with work/responsibility	0.60	2.71E-02
	Less recent since last attend alcohol self help group*	0.79	2.65E-02
	Being alcohol dependent	1.73	2.56E-02
	Smoked in dangerous situation >3 times	0.66	3.75E-02
	Smoked more frequently after waking up	0.57	7.76E-03
	Older age first had cigarette <sup>§</sup>	1.31	6.24E-03
	Give up social activities because of smoking	1.82	1.73E-02
	Less recent since first had sedatives*	1.3	1.06E-02
	Had 2 marijuana symptoms lasting a month	2.13	4.83E-03
	Disclosed drug problems to professionals	1.53	2.00E-02
	N years using sedatives <sup>§</sup>	1.29	1.33E-02
Behavioral	Gambling severity	0.8	3.32E-02
	Less recent since last hurt oneself on purpose*	1.39	1.96E-02
Other Health	Less recent since last had depression >1 week*	1.31	1.66E-02
	Ever treated with medication or ECT for depression	1.75	4.33E-02
	HIV positive	2.47	1.39E-03
	Hear delusional noises when awake	1.61	3.40E-02
	Anxious for >6 months	0.58	2.06E-02
	Health has always been better than now	0.62	9.64E-03
Demographic	Female sex	1.91	1.83E-03
	Not raised primarily by single parent	1.58	1.30E-02
	Current age	2.44	1.41E-12
	Fulltime employment	1.84	1.82E-02

## B) European Americans

	Variable	OR	p-value
Drug related	Less recent since last cocaine use*	1.91	3.30E-15
	Older age at first heavy opioid use <sup>§</sup>	0.56	2.67E-12
	Number of years using heroin <sup>§</sup>	0.69	3.01E-07
	Less recent since last cocaine injection*	1.85	2.38E-06
	Less recent since last had alcohol symptoms that last >1 month*	1.34	4.13E-05
	>20 outpatient visits for drug/psychiatric problems in the most recent year	1.76	9.56E-06
	Less recent since opioid treatment initiation*	1.53	2.24E-06
	Used cocaine >11 times last year	0.47	1.09E-05
	Less recent since first used opioid 1/week for >1 month*	1.41	1.10E-04
	Have injected cocaine	2.01	1.66E-04
	Older age when first heavy use of cocaine <sup>§</sup>	1.27	7.40E-04
	Marijuana interfered with work/home	1.67	1.61E-03
	Less recent since one started opioid self help group*	1.28	2.40E-03
	Less recent since last feel high with cocaine for >1 day*	1.33	1.15E-03
	Less recent since last attended cocaine self help group*	0.76	3.27E-04
	Continued using stimulant for its psychological problems	0.54	2.27E-02
	Heart slow down when cutting down tobacco use	0.56	1.52E-02
	Had tobacco but not addicted	0.60	3.72E-03
	Disclosed problems with cocaine usage to professional	1.65	2.45E-03
	Always able to cut down smoking	1.28	3.69E-02
Behavioral	Treated at outpatient drug program for cocaine	1.41	2.81E-02
	Used opioid more than intended to	0.63	1.15E-02
	Craving for cocaine when cut down	1.42	1.05E-02
	N drinks one ever had in 24h <sup>§</sup>	1.15	1.61E-02
	Ever smoked >1 pack of cigarette daily for >1 month	1.16	2.16E-02
	Had stopped using stimulant for >3 month	1.99	2.38E-03
	Drink resulted in objection or problems with family and work	1.48	4.39E-03
	Started answering before one completed question that caused trouble	0.73	2.24E-02
	<3 ASP criteria in 12 month period	0.64	2.69E-03
	Often failed to pay debt since 15th birthday	0.68	2.17E-03
	Suspended or expelled from school	0.67	2.16E-03
	N times in jail <sup>§</sup>	0.85	1.05E-02
	Less recent since last had suicide idea*	1.20	8.08E-03
	Older age last had antisocial behaviors <sup>§</sup>	1.35	1.03E-04
	No fear in most disturbing event	1.93	1.66E-06
	Avoided scenes that could be reminded of PTSD	1.88	7.88E-05
	Had OCD behaviors when depressed	0.49	3.23E-04
	Feeling distracted	1.56	8.15E-04
	Racing car both on drug or clean	0.56	3.79E-03
	Found customers for prostitutes since 15 years old	0.68	4.04E-02
Other Health	Depression always started with drug problems	1.64	2.55E-03
	Have 8 depression symptoms	1.46	8.99E-03
	Have asthma	0.68	1.22E-02
	Less recent since last had depression episode*	1.20	1.65E-02

	Have outstanding emotional problem	1.63	5.55E-03
	Older age of the heaviest weight <sup>\$</sup>	1.21	2.68E-02
Demographics	Body mass index	1.32	3.59E-06
	Being an atheist	1.45	1.34E-02
	Household income	0.87	1.30E-03
	Current age	2.00	5.74E-09

OR=odds ratio

\* Categorical variables. 1- “within the last two weeks”, 2- “two weeks to just under one month ago”, 3-“one month to just under six months ago”, 4-“six months to a year ago”, 5- “more than a year ago”. OR value indicates the impact of per level change.

\$ Continuous variables. OR value indicates the impact of per standard deviation change.

## References

1. Volkow ND, Frieden TR, Hyde PS, Cha SS. Medication-assisted therapies--tackling the opioid-overdose epidemic. *N Engl J Med*. 2014;370(22):2063-6.
2. CDC. Overview of the Drug Overdose Epidemic: Behind the Numbers. 2019.
3. Acion L, Kelmansky D, van der Laan M, Sahker E, Jones D, Arndt S. Use of a machine learning framework to predict substance use disorder treatment success. *PLoS One*. 2017;12(4):e0175383.
4. Che Z, St Sauver J, Liu H, Liu Y. Deep Learning Solutions for Classifying Patients on Opioid Use. *AMIA Annu Symp Proc*. 2017;2017:525-34.
5. Badiani A, Spagnolo PA. Role of environmental factors in cocaine addiction. *Curr Pharm Des*. 2013;19(40):6996-7008.
6. Randall J, Ellis ZW, Nicholas Genes and Avi Ma'ayan. Predicting opioid dependence from electronic health records with machine learning. *BioData Mining*. 2019;12(3).
7. LiX C, CurranG, TilfordJM, FelixH1MartinBC. PSY5 - Using Machine Learning to Predict Opioid Overdoses Among Prescription Opioid Users. *Value in Health*. 2018;21(Supplement 1):S245.
8. Dowell D, Haegerich TM, Chou R. CDC Guideline for Prescribing Opioids for Chronic Pain - United States, 2016. *MMWR Recomm Rep*. 2016;65(1):1-49.
9. Ainscough TS, Brose LS, Strang J, McNeill A. Contingency management for tobacco smoking during opioid addiction treatment: a randomised pilot study. *BMJ Open*. 2017;7(9):e017467.
10. Samet JH, Kertesz SG. Suggested Paths to Fixing the Opioid Crisis: Directions and Misdirections. *JAMA Netw Open*. 2018;1(2):e180218.
11. Pierucci-Lagha A, Gelernter J, Chan G, Arias A, Cubells JF, Farrer L, et al. Reliability of DSM-IV diagnostic criteria using the semi-structured assessment for drug dependence and alcoholism (SSADDA). *Drug Alcohol Depend*. 2007;91(1):85-90.
12. Montalvo-Ortiz JL, Cheng Z, Kranzler HR, Zhang H, Gelernter J. Genomewide Study of Epigenetic Biomarkers of Opioid Dependence in European- American Women. *Sci Rep*. 2019;9(1):4660.
13. Wetherill L, Agrawal A, Kapoor M, Bertelsen S, Bierut LJ, Brooks A, et al. Association of substance dependence phenotypes in the COGA sample. *Addict Biol*. 2015;20(3):617-27.
14. Bzdok D, Altman N, Krzywinski M. Statistics versus machine learning. *Nat Methods*. 2018;15(4):233-4.
15. Beam AL, Kohane IS. Big Data and Machine Learning in Health Care. *JAMA*. 2018;319(13):1317-8.
16. Wong A, Young AT, Liang AS, Gonzales R, Douglas VC, Hadley D. Development and Validation of an Electronic Health Record-Based Machine Learning Model to Estimate Delirium Risk in Newly Hospitalized Patients Without Known Cognitive Impairment. *JAMA Netw Open*. 2018;1(4):e181018.

17. Zhang H, Zheng Y, Yoon G, Zhang Z, Gao T, Joyce B, et al. Regularized estimation in sparse high-dimensional multivariate regression, with application to a DNA methylation study. *Stat Appl Genet Mol Biol*. 2017;16(3):159-71.
18. Rondina JM, Hahn T, de Oliveira L, Marquand AF, Dresler T, Leitner T, et al. SCoRS--A Method Based on Stability for Feature Selection and Mapping in Neuroimaging [corrected]. *IEEE Trans Med Imaging*. 2014;33(1):85-98.
19. Zhang X, Lu X, Shi Q, Xu XQ, Leung HC, Harris LN, et al. Recursive SVM feature selection and sample classification for mass-spectrometry and microarray data. *BMC Bioinformatics*. 2006;7:197.
20. Darst BF, Malecki KC, Engelman CD. Using recursive feature elimination in random forest to account for correlated variables in high dimensional data. *BMC Genet*. 2018;19(Suppl 1):65.
21. Pinaya WH, Gadelha A, Doyle OM, Noto C, Zugman A, Cordeiro Q, et al. Using deep belief network modelling to characterize differences in brain morphometry in schizophrenia. *Sci Rep*. 2016;6:38897.
22. Gelernter J, Panhuysen C, Weiss R, Brady K, Hesselbrock V, Rounsaville B, et al. Genomewide linkage scan for cocaine dependence and related traits: significant linkages for a cocaine-related trait and cocaine-induced paranoia. *Am J Med Genet B Neuropsychiatr Genet*. 2005;136B(1):45-52.
23. Sherva R, Wang Q, Kranzler H, Zhao H, Koesterer R, Herman A, et al. Genome-wide Association Study of Cannabis Dependence Severity, Novel Risk Variants, and Shared Genetic Risks. *JAMA Psychiatry*. 2016;73(5):472-80.
24. Malison RT, Kalayasiri R, Sanichwankul K, Sughondhabirrom A, Mutirangura A, Pittman B, et al. Inter-rater reliability and concurrent validity of DSM-IV opioid dependence in a Hmong isolate using the Thai version of the Semi-Structured Assessment for Drug Dependence and Alcoholism (SSADDA). *Addict Behav*. 2011;36(1-2):156-60.
25. Mohamed Hebiri JCL. How Correlations Influence Lasso Prediction. *arXiv*. 2012.
26. Chih-wei Hsu C-cC, Chih-jen Lin. A practical guide to support vector classification. 2010.
27. Thais Mayumi Oshiro PSP, José Augusto Baranauskas. How Many Trees in a Random Forest? Machine learning and data mining in pattern recognition 8th international conference,. 2012;July.
28. Debaditya Roy SRMK, Krishna Mohan Chalavadi. Feature selection using Deep Neural Networks. 2015 International Joint Conference on Neural Networks (IJCNN). 2015.
29. Halligan S, Altman DG, Mallett S. Disadvantages of using the area under the receiver operating characteristic curve to assess imaging tests: a discussion and proposal for an alternative approach. *Eur Radiol*. 2015;25(4):932-9.
30. Mohamed Bekkar HKD, Taklit Akrouf Alitouche, . Evaluation Measures for Models Assessment over Imbalanced Data Sets. *Journal of Information Engineering and Applications* 2013;3(10).
31. Weng WH, Waghlikar KB, McCray AT, Szolovits P, Chueh HC. Medical subdomain classification of clinical notes using a machine learning-based natural language processing approach. *BMC Med Inform Decis Mak*. 2017;17(1):155.

32. Saito T, Rehmsmeier M. Precrec: fast and accurate precision-recall and ROC curve calculations in R. *Bioinformatics*. 2017;33(1):145-7.
33. Fabian Pedregosa GV, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, Édouard Duchesnay. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*. 2011;12:2825-30.
34. Gowin JL, Ball TM, Wittmann M, Tapert SF, Paulus MP. Individualized relapse prediction: Personality measures and striatal and insular activity during reward-processing robustly predict relapse. *Drug Alcohol Depend*. 2015;152:93-101.
35. Erguzel TT, Noyan CO, Eryilmaz G, Unsalver BO, Cebi M, Tas C, et al. Binomial Logistic Regression and Artificial Neural Network Methods to Classify Opioid-Dependent Subjects and Control Group Using Quantitative EEG Power Measures. *Clin EEG Neurosci*. 2019;1550059418824450.
36. Ahn WY, Vassileva J. Machine-learning identifies substance-specific behavioral markers for opiate and stimulant dependence. *Drug Alcohol Depend*. 2016;161:247-57.
37. Ahn WY, Ramesh D, Moeller FG, Vassileva J. Utility of Machine-Learning Approaches to Identify Behavioral Markers for Substance Use Disorders: Impulsivity Dimensions as Predictors of Current Cocaine Dependence. *Front Psychiatry*. 2016;7:34.
38. Buu A, Johnson NJ, Li R, Tan X. New variable selection methods for zero-inflated count data with applications to the substance abuse field. *Stat Med*. 2011;30(18):2326-40.
39. Oliveto AH, Feingold A, Schottenfeld R, Jatlow P, Kosten TR. Desipramine in opioid-dependent cocaine abusers maintained on buprenorphine vs methadone. *Arch Gen Psychiatry*. 1999;56(9):812-20.
40. Krantz MJ, Mehler PS. Treating opioid dependence. Growing implications for primary care. *Arch Intern Med*. 2004;164(3):277-88.
41. Schottenfeld RS, Pakes J, Ziedonis D, Kosten TR. Buprenorphine: dose-related effects on cocaine and opioid use in cocaine-abusing opioid-dependent humans. *Biol Psychiatry*. 1993;34(1-2):66-74.
42. Mannelli P, Wu LT, Peindl KS, Gorelick DA. Smoking and opioid detoxification: behavioral changes and response to treatment. *Nicotine Tob Res*. 2013;15(10):1705-13.
43. Thylstrup B, Schroder S, Hesse M. Psycho-education for substance use and antisocial personality disorder: a randomized trial. *BMC Psychiatry*. 2015;15:283.
44. Robins LN. The intimate connection between antisocial personality and substance abuse. *Soc Psychiatry Psychiatr Epidemiol*. 1998;33(8):393-9.
45. Brook JS, Whiteman M, Finch SJ, Cohen P. Young adult drug use and delinquency: childhood antecedents and adolescent mediators. *J Am Acad Child Adolesc Psychiatry*. 1996;35(12):1584-92.
46. Dragisic T, Dickov A, Dickov V, Mijatovic V. Drug Addiction as Risk for Suicide Attempts. *Mater Sociomed*. 2015;27(3):188-91.



47. Phillips J, Carpenter KM, Nunes EV. Suicide risk in depressed methadone-maintained patients: associations with clinical and demographic characteristics. *Am J Addict.* 2004;13(4):327-32.
48. Harned MS, Najavits LM, Weiss RD. Self-harm and suicidal behavior in women with comorbid PTSD and substance dependence. *Am J Addict.* 2006;15(5):392-5.
49. Tremereau F, Darreys A, Staner L, Correa H, Weibel H, Khidichian F, et al. Suicidality in opioid-dependent subjects. *Am J Addict.* 2008;17(3):187-94.
50. Rounsaville BJ, Kleber HD. Untreated opiate addicts. How do they differ from those seeking treatment? *Arch Gen Psychiatry.* 1985;42(11):1072-7.
51. Johnson RE, Chutuape MA, Strain EC, Walsh SL, Stitzer ML, Bigelow GE. A comparison of levomethadyl acetate, buprenorphine, and methadone for opioid dependence. *N Engl J Med.* 2000;343(18):1290-7.
52. Fiellin DA, Weiss L, Botsko M, Egan JE, Altice FL, Bazerman LB, et al. Drug treatment outcomes among HIV-infected opioid-dependent patients receiving buprenorphine/naloxone. *J Acquir Immune Defic Syndr.* 2011;56 Suppl 1:S33-8.
53. Kazakova OB, Tret'iakova EV, Kukovinets OS, Tolstikov GA, Nazyrov TI, Chudov IV, et al. [Synthesis and pharmacological activity of amides and ozonolysis product of maleopimaric acid]. *Bioorg Khim.* 2010;36(6):832-40.
54. Liang W, Chikritzhs T. Reduction in alcohol consumption and health status. *Addiction.* 2011;106(1):75-81.
55. Carpenter CS. Workplace drug testing and worker drug use. *Health Serv Res.* 2007;42(2):795-810.
56. Kosten TR, George TP. The neurobiology of opioid dependence: implications for treatment. *Sci Pract Perspect.* 2002;1(1):13-20.
57. Moscati A, Mezuk B. Losing faith and finding religion: religiosity over the life course and substance use and abuse. *Drug Alcohol Depend.* 2014;136:127-34.
58. Smyth BP, Barry J, Keenan E, Ducray K. Lapse and relapse following inpatient treatment of opiate dependence. *Ir Med J.* 2010;103(6):176-9.
59. Scherbaum N, Specka M. Factors influencing the course of opiate addiction. *Int J Methods Psychiatr Res.* 2008;17 Suppl 1:S39-44.
60. Junghwan Cho KL, Ellie Shin, Garry Choy, Synho Do. How much data is needed to train a medical image deep learning system to achieve necessary high accuracy? *ICLR2016.* 2016.