A preliminary risk prediction model for cannabis use disorder

Rajapaksha Mudalige Dhanushka S. Rajapaksha [a], Ryan Hammonds [b], Francesca Filbey [b], Pankaj K. Choudhary [a,*], Swati Biswas [a,*]

[a] *Department of Mathematical Sciences, University of Texas at Dallas, Richardson, TX, USA*
[b] *School of Behavioral and Brain Sciences, University of Texas at Dallas, Richardson, TX, USA*

A B S T R A C T

The ongoing trend toward legalization of cannabis for medicinal/recreational purposes is expected to increase the prevalence of cannabis use disorder (CUD). Thus, it is imperative to be able to predict the quantitative risk of developing CUD for a cannabis user based on their personal risk factors. Yet no such model currently exists. In this study, we perform preliminary analysis toward building such a model. The data come from n = 94 regular cannabis users recruited from Albuquerque, New Mexico during 2007–2010. As the data are cross-sectional, we only consider risk factors that remain relatively stable over time. We apply statistical and machine learning classification techniques that allow n to be small relative to the number of predictors. We use predictive accuracy estimated using leave-one-out-cross-validation to evaluate model performance. The final model is a LASSO logistic regression model consisting of the following seven risk factors: age; level of enjoyment from initial cigarette smoking; total score on Impulsive Sensation-Seeking Scale questionnaire; score on cognitive instability factor of Barratt Impulsivity Scale questionnaire; and scores on neuroticism, openness, and conscientiousness personality traits of Neuroticism, Extraversion, and Openness inventory. This model has an overall accuracy of 0.66 and the area under its receiver operating characteristic curve is 0.65. In summary, a preliminary relative risk model for predicting the quantitative risk of CUD is developed. It can be employed to identify users at high risk of CUD who may be provided with early intervention.

1. Introduction

Substance use disorders are currently a major public health crisis in the US (SAMHSA, 2016). Cannabis is the most commonly used illicit substance in the world (NIDA, 2019). With more than 200 million users of cannabis worldwide, its harmful health effects have become a serious global problem (Colizzi et al., 2020; Gunn et al., 2016; Meier et al., 2012). During the past two decades, the laws and policies related to cannabis use have also changed drastically throughout the world. For example, countries such as Canada, Spain, and Germany have legalized cannabis for medical use while some have even legalized its non-medical use, e.g., Uruguay in 2015 and Canada in 2018 (Cox, 2018). Not surprisingly, the legalization trend continues in the US, with 33 states and the District of Columbia legalizing medical marijuana use, and 11 states and the District of Columbia legalizing adult non-medical marijuana use (ProCon.org, 2020).

Regardless of the developing accord about the usefulness of medical marijuana for several serious illnesses, there is a widespread concern that this may cause adverse effects (Brown and Hasin, 2019; Hammond et al., 2020; Wall et al., 2019). According to a study on the effects of medical marijuana laws, the likelihood of current as well as regular use of cannabis among people aged 21 or older has increased after the laws came into effect (Wen et al., 2015). This also appears to have contributed to an increased prevalence of illicit cannabis use and cannabis use disorder (Hasin et al., 2017). In particular, among adult males, arrests due to illegal marijuana possession in major cities have increased by 15–20% and the treatment provided in rehabilitation facilities for such arrests have increased by 10–20% (Chu, 2014).

This article focuses on cannabis use disorder (CUD). Earlier, there was a consensus that CUD is rare, which is no longer true. It is estimated that about 34% of cannabis users develop CUD during their lifetime based on the 4th edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) (Marel et al., 2019). Furthermore, a recent study based on DSM-V criteria found that about 27% of cannabis users develop CUD during their lifetime (Feingold et al., 2020). Another research shows that after legalizing marijuana for recreational use, the

prevalence of CUD among past year cannabis users between the ages of 12 and 17 rose from 22.8% to 27.2% (Cerdá et al., 2019). Thus, given that the prevalence of CUD is expected to increase further, it is imperative to predict the risk of developing CUD for cannabis users, especially for adolescents and emerging adults, based on their personal risk factors. Identifying individuals at high risk of CUD will allow the possibility of applying early intervention, which may potentially help stem the increasing prevalence of the disorder.

Several risk factors have been reported for substance use disorders in general and specifically for CUD. These include male sex, early exposure to traumatic events, early use initiation, family history of substance use, childhood depression, and conduct disorder symptoms (Gray and Squeglia, 2018; Meier et al., 2016; Tomko et al., 2019). High impulsivity and certain personality traits are also associated with the disorders (Beaton et al., 2014; Ketcherside et al., 2016). In particular, work by co-author Filbey's lab showed that openness distinguishes cannabis-only users from nicotine-only users, co-morbid marijuana and nicotine users, and non-users (Ketcherside et al., 2016). The results from this study also indicate that conscientiousness is lower among cannabis users.

Some brief screening tools such as BSTAD (Brief Screener for Tobacco, Alcohol, and Other Drugs) (Levy et al., 2014) and S2BI (Screening to Brief Intervention) have been developed for adolescents (Levy et al., 2014). For example, the cutoff for CUD based on BSTAD is at least two days of marijuana use in the past one year. A relatively lengthy tool, Transmissible Liability Index, assesses the inherited risk for disorders based on a 45-item questionnaire (Tarter et al., 2015; Vanyukov et al., 2009). Also, a recent study has developed a simple cumulative risk index for substance dependence in adulthood using risk factors in childhood and adolescence (Meier et al., 2016). It can be used to screen adolescents who are likely to develop persistent disorder in adulthood. A similar study has developed a risk score by counting the number of early life risk factors present in an individual and associating it with cannabis use and CUD in early adulthood (Hayatbakhsh et al., 2009). However, a key limitation of the existing tools is that none of them provides a quantitative risk of developing the disorder based on personal risk factors, which restricts their practical utility. Models for predicting such risks have been developed for several diseases, including breast cancer (Gail et al., 1989; NCI, 2020), contralateral breast cancer (Chowdhury et al., 2018, 2017), heart disease (D'Agostino et al., 2008), depression (Cattelani et al., 2019; King et al., 2008), and psychiatric disorders (Bernardini et al., 2017), and they are in wide clinical use. However, currently there is no such quantitative risk prediction tool for CUD.

In this study, we build upon the findings of Ketcherside et al. (2016) in a cannabis-using adult population and perform a secondary analysis of the data. More specifically, we build a preliminary quantitative risk prediction model to estimate the chance that a cannabis user will develop CUD based on various demographic, behavioral, psychiatric, and cognitive risk factors.

2. Methods

2.1. Participants

The study participants are cannabis users who were recruited from the general population in the Albuquerque metro area during 2007–2010 (Filbey and Yezhuvath, 2013). The study was approved by the University of New Mexico and University of Texas at Dallas institutional review boards. All participants signed an informed consent form. The intent of the original study was to determine the neurobiological antecedents of substance use disorders (Filbey and Yezhuvath, 2013; Ketcherside and Filbey, 2015). For the secondary analysis reported in this article, the inclusion criterion was regular use of cannabis, i.e., at least 4 times a week for at least 6 months.

2.2. Data preparation

The initial data set obtained after applying the inclusion criterion consisted of 118 cannabis users. We used CUD as the outcome (response) variable, which was derived based on the DSM-IV criteria for dependence. DSM-IV is a multi-dimensional measure for diagnosing CUD and is well established in the literature (Hayatbakhsh et al., 2009; Marel et al., 2019; Meier et al., 2016). The variable selection process to identify potential risk factors was the following. First, the variables with more than 50% missing values were discarded. Then, among the remaining variables, only those that remain relatively stable over time were chosen. Given the cross-sectional nature of the data, focusing attention on such type of variables protects against using risk factors that may actually be an effect of CUD. This resulted in 30 variables. These included measures of impulsivity and personality traits. The former were obtained using two questionnaires, namely, Impulsive Sensation-Seeking Scale (ImpSS), a 19-item self-reported questionnaire from the Zuckerman-Kuhlman Personality Questionnaire (Zuckerman et al., 1993) and Barratt Impulsivity Scale (BIS), a 30-item self-reported questionnaire where the items can be grouped into six first-order factors that measure different aspects of impulsivity (Beaton et al., 2014; Stanford et al., 2009). Both ImpSS and BIS were considered because there are some characteristics of impulsivity that are captured by ImpSS but not by BIS and vice versa, and the two have been used together in several studies (Beaton et al., 2014; Martínez-Loredo et al., 2015, 2018). The personality traits were obtained using Neuroticism, Extraversion, and Openness inventory (NEO), a five-factor inventory for measuring five different dimensions of personality (Ketcherside et al., 2016). The actual measures derived from these questionnaires were total score on the ImpSS questionnaire, scores on the six factors from the BIS questionnaire, and scores on the five factors from the NEO questionnaire.

Only 46 of the 118 subjects had complete data on all 30 variables. To guard against loss of subjects due to missing data on potentially unimportant variables, univariate logistic regression models were fitted with each of these variables as a predictor. Thereafter, the predictors with univariate model p-value less than or equal to 0.3 were selected into the final set of potential predictors for a multivariate model (Hosmer et al., 2013). The resulting data set had 12 potential risk factors and 94 subjects with complete observations on them. This final data set was used for the rest of the model building exercise.

2.3. Risk factors

Table 1 presents the 12 risk factors. They fall in three groups. Group 1 consists of three variables, namely, age at the time of data collection, age at first use of cannabis, and the level of enjoyment from initial cigarette smoking (measured on an ordinal scale from 0 to 10 through the question "How much did you enjoy smoking at first?," with higher value indicating more enjoyment). Group 2 consists of six measures of impulsivity, namely, total score on the ImpSS questionnaire (ImpSS-T) (Beaton et al., 2014; Zuckerman et al., 1993) and scores on five of the six first-order factors from the BIS questionnaire (Beaton et al., 2014; Stanford et al., 2009), namely, attention (BIS-A), cognitive instability (BIS-I), motor impulsiveness (BIS-M), perseverance (BIS-P), and cognitive complexity (BIS-C). For all variables in this group, higher values imply greater impulsivity (Patton et al., 1995). Group 3 consists of scores on three of the five personality dimensions measured using the NEO inventory (Stanford et al., 2009), namely, neuroticism (general tendency to experience negative feelings; NEO-N), openness (open to new experiences and imaginative; NEO-O), and conscientiousness (forward planning, organization and ability to carry out tasks; NEO-C). Higher values for these variables imply greater neuroticism, openness, and conscientiousness, respectively.

Table 1

Characteristics of subjects in terms of mean (SD) on various risk factors whose p-values from univariate logistic regression models are less than 0.3.

Variable	Controls (n = 36)	Cases (n = 58)	Total (n = 94)	Range	P-Value
Age	25.57 (6.63)	23.16 (6.22)	24.07 (6.45)	18–50	0.250
Age at first cannabis use	14.50 (2.62)	14.72 (1.95)	14.64 (2.22)	7–22	0.221
Level of enjoyment from Initial smoking	4.11 (2.94)	2.62 (3.01)	3.19 (3.05)	0–10	0.038
ImpSS-T	8.47 (3.79)	10.69 (3.72)	9.84 (3.88)	1–17	0.079
BIS-A	8.56 (2.88)	10.09 (3.22)	9.50 (3.17)	5–18	0.032
BIS-M	14.50 (3.07)	15.31 (3.35)	15.00 (3.25)	7–24	0.239
BIS-C	10.47 (2.55)	11.45 (2.58)	11.07 (2.60)	5–19	0.181
BIS-P	7.03 (1.95)	7.48 (1.87)	7.31 (1.90)	4–14	0.189
BIS-I	5.94 (1.71)	7.21 (1.99)	6.72 (1.98)	3–12	0.002
NEO-N	16.36 (7.93)	21.78 (8.53)	19.70 (8.67)	0–38	0.004
NEO-O	32.03 (6.86)	35.52 (6.76)	32.95 (6.80)	17–44	0.173
NEO-C	33.56 (6.85)	29.93 (7.10)	31.32 (7.19)	14–47	0.065

2.4. Data analysis

The data analysis was performed using five common statistical and machine learning models for classification (James et al., 2013), namely, logistic regression with LASSO penalty, K-Nearest Neighbor (KNN), support vector machine (SVM) with radial kernel, random forest, and gradient boosting. We chose these techniques because they work even when the sample size is small relative to the number of predictors, as is the case here (James et al., 2013). The tuning parameters involved in these models were selected using leave-one-out-cross-validation (LOOCV) (James et al., 2013). Moreover, in keeping with the common practice, the performance of these models was evaluated by examining their prediction accuracy as measured using overall accuracy (i.e., the proportion of correct classifications), sensitivity (i.e., the proportion of correct classifications among the CUD subjects), and specificity (i.e., the proportion of correct classifications among the non-CUD subjects). Further, due to the lack of independent test data, the performance measures were computed using LOOCV. By protecting against over-fitting, the LOOCV-based measures provide a more accurate assessment of model performance on future unseen data than those computed directly from the training data. By default, the models use 0.5 as the cutoff for probability, that is, a study subject is classified as having CUD if their probability of CUD exceeds 0.5. If the cutoff is increased, the sensitivity will decrease and specificity will increase.

To evaluate the overall model performance, we used the receiver operating characteristic (ROC) curve, a plot of sensitivity against 1-specificity (both computed using LOOCV) obtained by varying the cutoffs, and computed the corresponding area under the curve (AUC) (James et al., 2013). The models were fit using the statistical software system R (RCoreTeam, 2019) with the following specific packages: glmnet (Friedman et al., 2010) for LASSO logistic regression, knn (Venables and Ripley, 2002), e1071 (Dimitriadou et al., 2019) for SVM, randomForest (Liaw and Wiener, 2002), gbm (Greenwell et al., 2019) for gradient boosting, caret (Kuhn, 2008) for LOOCV, and pROC (Robin et al., 2011) for AUC.

3. Results

3.1. Sample characteristics

The characteristics of cases of CUD and controls (i.e., the non-CUD subjects) with respect to the 12 risk factors under consideration are shown in Table 1. Out of 94 subjects, 67 (71%) were males and 58 (61.7%) were cases. Compared to controls, the cases on average were younger (23.16 vs 25.56 years), reported lower level of enjoyment from initial smoking (2.62 vs 4.11), were more impulsive sensation seeking (10.69 vs 8.47 on ImpSS-T), and had higher scores on BIS impulsivity traits (e.g., 7.21 vs 5.94 on cognitive impulsivity, BIS-I). The cases were also more likely to experience negative feelings and were more open to new experiences than controls as the cases reported higher averages for neuroticism (21.78 vs 16.36 on NEO-N) and openness (33.52 vs 32.03 on NEO-O), respectively. In addition, the cases were less conscientious than controls as reflected by lower average values for NEO-C for cases (29.93 vs 33.56). Among the twelve risk factors, only the following six exhibit statistically significant association with CUD in a univariate logistic regression model at 10% level of significance: level of enjoyment from initial smoking, ImpSS-T, BIS-A, BIS-I, NEO-N, and NEO-C.

3.2. Results from multivariate models

Fig. 1 presents the variables with non-zero regression coefficients from LASSO logistic regression model and the top seven variables based on variable importance measures for the other models. The seven variables selected by LASSO, namely, age, level of enjoyment from initial smoking, ImpSS-T, BIS-I, NEO-N, NEO-O, and NEO-C, were also found to be important by the other models. In particular, except ImpSS-T and NEO-O, the remaining five were selected as important by all other models. Moreover, ImpSS-T was chosen as an important predictor by KNN, random forest, and SVM while NEO-O was indicated to be important by random forest and gradient boosting.

Table 2 presents the accuracy, sensitivity, and specificity of the models based on 0.5 cutoff as computed using LOOCV as well as the AUC of the models. The associated ROC curves and the plots of accuracy versus cutoff are provided in Supplementary Materials. Although the various models performed similarly, which is reassuring, overall we may conclude that LASSO and gradient boosting outperformed the others. For example, the two are tied for the highest AUC. Nevertheless, an advantage of LASSO is that it provides estimates of regression coefficients and hence odds ratios. This allows easy interpretation of the effects of the risk factors. This important and desirable feature is not available in other models. Therefore, we choose the LASSO logistic regression model as our final model.

It may be of interest to quantify the advantage of this model over a random guess classifier that predicts CUD with probability 0.617, the proportion of CUD cases in the data. The accuracy, sensitivity, and specificity of this classifier can be calculated to be 0.527, 0.617, and 0.383, respectively. These are much lower than the corresponding values reported in Table 2 for the LASSO model.

3.3. The final LASSO model

The final LASSO model predicted the CUD status with 66% accuracy. Its sensitivity and specificity were 0.81 and 0.42, respectively. Thus, it does a much better job of correctly identifying the CUD cases than the non-CUD controls at the probability cutoff of 0.5. This cutoff may not be appropriate in all clinical settings. The appropriate cutoff can be chosen by examining its ROC curve, presented in Supplementary Materials, for the tradeoff between sensitivity and specificity. Its AUC is 0.65. The seven variables selected by this model together with their estimated coefficients and the associated odds ratios (OR) are shown in Table 3. The higher probability of CUD was associated with younger age (OR = 0.97), lower level of enjoyment from initial smoking (OR = 0.89), higher
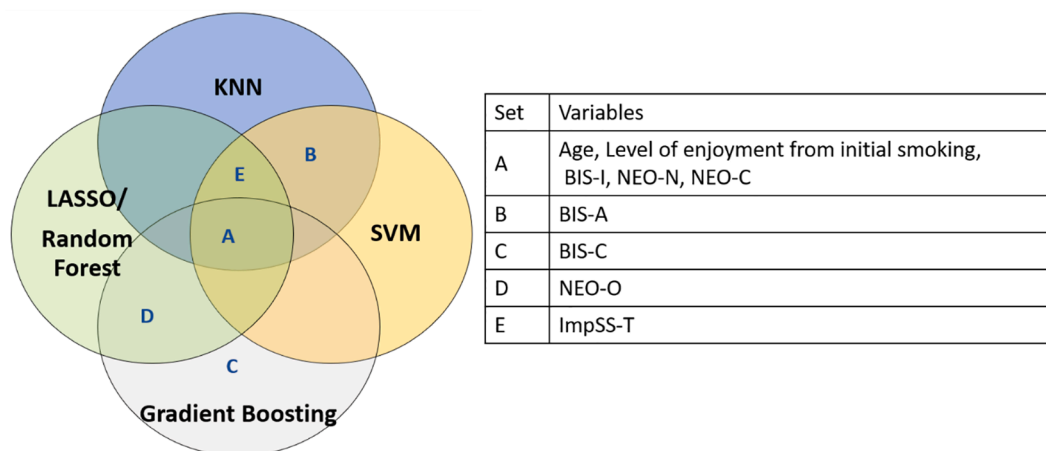
Fig. 1. A Venn diagram showing overlap of important predictors according to various models.

Table 2
Performance measures for various models computed using LOOCV. All measures except AUC are based on 0.5 probability cutoff for classification.

Measure	LASSO	KNN	Random Forest	SVM	Gradient Boosting
Accuracy	0.66	0.66	0.65	0.64	0.68
Sensitivity	0.81	0.90	0.78	0.83	0.79
Specificity	0.42	0.28	0.44	0.33	0.50
AUC	0.65	0.64	0.63	0.63	0.65

score on impulsivity (OR = 1.09), greater cognitive instability (OR = 1.17), higher neuroticism, i.e., more prone to experience negative feelings (OR = 1.03), greater openness to new experiences (OR = 1.01), and lower conscientiousness (OR = 0.99).

To illustrate the model, we considered two subjects from the data who had the largest and the smallest predicted probability of CUD. Their true status is CUD and non-CUD, respectively. The first subject was young (age = 18); received little enjoyment from initial smoking (score = 1); had high scores on impulsivity (ImpSS-T = 12), cognitive instability (BIS-I = 12), and neuroticism (NEO-N = 32); was quite open to new experiences (NEO-O = 38); and had low conscientiousness (NEO-C = 19). The predicted probability of CUD for this subject was 0.93. The second subject was 49 years old; received much enjoyment from initial smoking (score = 10); had low scores on impulsivity (ImpSS-T = 9), cognitive instability (BIS-I = 4), and neuroticism (NEO-N = 11); was also quite open to new experiences (NEO-O = 36); and had high conscientiousness (NEO-C = 39). The predicted probability of CUD for this subject was 0.15.

4. Discussion

Substance use disorders are a growing public health problem and cannabis is the most commonly used illicit substance in the world (NIDA, 2019; WHO, 2020). The legalization of cannabis for medical and recreational purposes worldwide has increased cannabis use and CUD. Therefore, there is a growing need for a CUD risk prediction tool. In this study, we built a preliminary model by identifying risk factors with the help of several statistical and machine learning algorithms.

We eventually chose the LASSO logistic regression model as the final

model for two reasons. First, there was no major difference among the top performing models. Second, LASSO allows the ability to interpret the effects of risk factors quantitatively, a feature unavailable in the other methods. The LASSO model gave seven risk factors with non-zero (important) coefficients. We had also explored the possibility of adding interaction terms to this model but did not eventually add any because the model with interactions had lower predictive accuracy than this model.

The risk factors identified by our model are consistent with the literature (Dougherty et al., 2013; Kong et al., 2013; Lee-Winn et al., 2018; Pampati et al., 2018; Winters and Lee, 2008). In particular, previous findings indicate that younger people are more likely to develop CUD (Winters and Lee, 2008). Using ImpSS and BIS scales, numerous studies have shown that high impulsivity is prevalent among users of nicotine (Chase and Hogarth, 2011), cocaine (Ball, 1995), and alcohol (Curran et al., 2010). We also found that higher ImpSS-T increases the likelihood of dependence on cannabis. The positive association between cognitive instability and CUD status that we found is also known (Mitchell and Potenza, 2014).

Similarly, the relationship of CUD with personality trait risk factors based on NEO is consistent with the previous findings (Ball, 2005; Fridberg et al., 2011; Kotov et al., 2010). For example, cannabis users have higher openness and lower conscientiousness compared to non-users (Fridberg et al., 2011; Ketcherside et al., 2016). Generally, high neuroticism is reported in nicotine-only users (Tate et al., 1994; Terracciano et al., 2008) and average neuroticism is reported in cannabis-only users (Terracciano et al., 2008). We found that higher neuroticism is associated with higher likelihood of CUD, which is not surprising because our sample consists of co-morbid marijuana and nicotine users.

We also found that less enjoyment from initial smoking is associated with increased likelihood of becoming cannabis dependent. This is in line with the findings from a nationally representative longitudinal study, which was conducted to identify the risk factors associated with different stages of cannabis use (Pampati et al., 2018). This study found that greater quantity of cigarette use decreased the likelihood of re-initiation of cannabis use among participants who were cannabis users prior to reaching adolescence (Pampati et al., 2018). Even though our overall findings are consistent with the literature, we did not find several risk factors for CUD that have been previously reported in the literature.

Table 3
Results from the final LASSO model.

Predictor	Intercept	Age	Level of Enjoyment from Initial Smoking	ImpSS-T	BIS-I	NEO-N	NEO-O	NEO-C
Coefficient	−0.8	−0.03	−0.12	0.08	0.16	0.03	0.01	−0.01
Odd Ratio		0.97	0.89	1.09	1.17	1.03	1.01	0.99

Some of the risk factors such as childhood depression and conduct disorder symptoms were not available in these data. While some other factors such as early exposure to traumatic events had substantial missing data because of which they were excluded. Yet others may not have been identified due to limitations of the study as described in the following.

Our study's first limitation is the cross-sectional and observational nature of the study because of which it is difficult to establish a causal relationship between a risk factor and CUD, especially for the factors that can vary over time. To mitigate the latter issue, we only used risk factors that remain relatively stable over time. However, even then we need to be cautious about drawing any conclusion about causation as this is an observational study. The second is that there are not a large number of subjects and the participating subjects came from a specific metro area in the US, which may not be representative of the entire population of all cannabis users. The third is due to missing values on the variables. When the risk factors are jointly analyzed in a multivariate model, this leads to a loss of some subjects as those with missing values in any of the multiple variables are discarded. We tried to balance the loss of sample size with the inclusion of risk factors. Moreover, to mitigate the issue of small sample size, we chose the statistical and machine learning methods that work even when the sample size is small relative to the number of predictors. Nonetheless, availability of complete data on more subjects would have provided higher power for identifying association.

We also acknowledge that the data used for this study were acquired in 2007–2010 and may be limited in its generalizability to current cannabis use impacts. Nonetheless, New Mexico's cannabis policies may be more historically representative of current national policies (compared to other states) given that medically-indicated cannabis was legalized in New Mexico in 2007 coinciding with the study's data collection. Thus, our findings may provide insights into future trends related to continued changes in cannabis legislation in the US. Also importantly, there has been no change in rate of current marijuana use in New Mexico in recent years, although the rate has remained significantly higher than the US rate (YRBS, 2017; YRRS, 2017). Thus, cannabis use in New Mexico has been stable and should not limit the impact of the current findings. Lastly, the mechanisms that underlie the risk for CUD likely remained relatively unchanged in the last 10 years.

Despite its limitations, this study represents a novel attempt to build a CUD risk prediction tool. To address the limitations, we are working towards building a risk prediction model using longitudinal data from a large number of subjects spread throughout the US. In addition, some people may be dependent on more than one substance (Moss et al., 2014; Richmond-Rakerd et al., 2016, 2017) and in fact, there may be common risk factors for several substance disorders (Oshri et al., 2018; Richmond-Rakerd et al., 2016). Therefore, it would be of interest to model jointly the relationship between multiple substance disorders and potential risk factors. Finally, inclusion of genetic and/or imaging factors can also provide a more personalized model.

5. Conclusion

This study developed a preliminary relative risk model for predicting the risk of CUD based on several risk factors. Higher risk of CUD was associated with younger age, lower level of enjoyment from initial smoking, higher score on impulsivity, greater cognitive instability, higher neuroticism, i.e., more prone to experience negative feelings, greater openness to new experiences, and lower conscientiousness.

Role of funding source

CRediT authorship contribution statement

Rajapaksha Mudalige Dhanushka S. Rajapaksha: Formal analysis, Investigation, Methodology, Writing - original draft. **Ryan Hammonds:** Data curation, Writing - review & editing. **Francesca Filbey:** Conceptualization, Funding acquisition, Investigation, Writing - review & editing. **Pankaj K. Choudhary:** Conceptualization, Funding acquisition, Investigation, Methodology, Supervision, Writing - original draft. **Swati Biswas:** Conceptualization, Funding acquisition, Investigation, Methodology, Supervision, Writing - original draft.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.pmedr.2020.101228.

References

Ball, S.A., 1995. The validity of an alternative five-factor measure of personality in cocaine abusers. Psychol. Assess 7 (2), 148–154.

Ball, S.A., 2005. Personality traits, problems, and disorders: Clinical applications to substance use disorders. 39(1), 84-102.

Beaton, D., Abdi, H., Filbey, F.M., 2014. Unique aspects of impulsive traits in substance use and overeating: specific contributions of common assessments of impulsivity. Am. J. Drug Alcohol Abuse 40 (6), 463–475.

Bernardini, F., Attademo, L., Cleary, S.D., Luther, C., Shim, R.S., Quartesan, R., Compton, M.T., 2017. Risk prediction models in psychiatry: toward a new frontier for the prevention of mental illnesses. J. Clin. Psychiatry 78 (5), 572–583.

Brown, Q.L., Hasin, D.S., 2019. Possible unintended consequences for pregnant women of legalizing cannabis use. Addiction.

Cattelani, L., Murri, M.B., Chesani, F., Chiari, L., Bandinelli, S., Palumbo, P., 2019. Risk prediction model for late life depression: development and validation on three large european Datasets. IEEE J. Biomed. Health Inform. 23 (5), 2196–2204.

Cerdá, M., Mauro, C., Hamilton, A., Levy, N.S., Santaella-Tenorio, J., Hasin, D., Wall, M.M., Keyes, K.M., Martins, S.S., 2019. Association Between Recreational Marijuana Legalization in the United States and Changes in Marijuana Use and Cannabis Use Disorder From 2008 to 2016. JAMA Psychiatry.

Chase, H.W., Hogarth, L., 2011. Impulsivity and symptoms of nicotine dependence in a young adult population. Nicotine Tob. Res. 13 (12), 1321–1325.

Chowdhury, M., Euhus, D., Onega, T., Biswas, S., Choudhary, P.K., 2017. A model for individualized risk prediction of contralateral breast cancer. Breast Cancer Res. Treat. 161 (1), 153–160.

Chowdhury, M., Euhus, D., Arun, B., Umbricht, C., Biswas, S., Choudhary, P., 2018. Validation of a personalized risk prediction model for contralateral breast cancer. Breast Cancer Res. Treat. 170 (2), 415–423.

Chu, Y.W., 2014. The effects of medical marijuana laws on illegal marijuana use. J. Health Econ. 38, 43–61.

Colizzi, M., Tosato, S., Ruggeri, M., 2020. Cannabis and cognition: connecting the dots towards the understanding of the relationship. Brain Sci. 10 (3).

Cox, C., 2018. The canadian cannabis act legalizes and regulates recreational cannabis use in 2018. Health Policy 122 (3), 205–209.

Curran, M., Fuertes, J., Alfonso, V., Hennessy, J., 2010. The association of sensation seeking and impulsivity to driving while under the influence of alcohol. J. Addict. Offender Counsel.

D'Agostino, R.B., Vasan, R.S., Pencina, M.J., Wolf, P.A., Cobain, M., Massaro, J.M., Kannel, W.B., 2008. General cardiovascular risk profile for use in primary care: the Framingham Heart Study. Circulation 117 (6), 743–753.

Dimitriadou, E., Hornik, K., Leisch, F., Meyer, D., Weingessel, A., 2019. Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071). TU Wien.

Dougherty, D.M., Mathias, C.W., Dawes, M.A., Furr, R.M., Charles, N.E., Liguori, A., Shannon, E.E., Acheson, A., 2013. Impulsivity, attention, memory, and decision-making among adolescent marijuana users. Psychopharmacology 226 (2), 307–319.

Feingold, D., Livne, O., Rehm, J., Lev-Ran, S., 2020. Probability and correlates of transition from cannabis use to DSM-5 cannabis use disorder: Results from a large-scale nationally representative study. Drug Alcohol. Rev. 39 (2), 142–151.

Filbey, F., Yezhuvath, U., 2013. Functional connectivity in inhibitory control networks and severity of cannabis use disorder. Am. J. Drug Alcohol. Abuse 39 (6), 382–391.

Fridberg, D.J., Vollmer, J.M., O'Donnell, B.F., Skosnik, P.D., 2011. Cannabis users differ from non-users on measures of personality and schizotypy. Psychiatry Res. 186 (1), 46–52.

Friedman, J., Hastie, T., Tibshirani, R., 2010. Regularization paths for generalized linear models via coordinate descent. J. Stat. Softw. 33 (1), 1–22.

Gail, M.H., Brinton, L.A., Byar, D.P., Corle, D.K., Green, S.B., Schairer, C., Mulvihill, J.J., 1989. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. J. Natl Cancer Inst. 81 (24), 1879–1886.

Gray, K.M., Squeglia, L.M., 2018. Research review: what have we learned about adolescent substance use? J. Child Psychol. Psychiatry 59 (6), 618–627.

Greenwell, B., Boehmke, B., Cunningham, J., GBM Developers, 2019. Generalized Boosted Regression Models.

Gunn, J.K., Rosales, C.B., Center, K.E., Nuñez, A., Gibson, S.J., Christ, C., Ehiri, J.E., 2016. Prenatal exposure to cannabis and maternal and child health outcomes: a systematic review and meta-analysis. BMJ Open 6 (4), e009986.

Hammond, C.J., Chaney, A., Hendrickson, B., Sharma, P., 2020. Cannabis use among U.S. adolescents in the era of marijuana legalization: a review of changing use patterns, comorbidity, and health correlates. Int. Rev. Psychiatry 1–14.

Hasin, D.S., Sarvet, A.L., Cerdá, M., Keyes, K.M., Stohl, M., Galea, S., Wall, M.M., 2017. US adult illicit cannabis use, cannabis use disorder, and medical marijuana laws: 1991–1992 to 2012–2013. JAMA Psychiatry 74 (6), 579–588.

Hayatbakhsh, M.R., Najman, J.M., Bor, W., O'Callaghan, M.J., Williams, G.M., 2009. Multiple risk factor model predicting cannabis use and use disorders: a longitudinal study. Am. J. Drug Alcohol Abuse 35 (6), 399–407.

Hosmer, D.W., Lemeshow, S., Sturdivant, R.X., 2013. Applied Logistic Regression, 3rd ed.

James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. An Introduction to Statistical Learning: with Applications in R.

Ketcherside, A., Filbey, F.M., 2015. Mediating processes between stress and problematic marijuana use. Addict. Behav. 45, 113–118.

Ketcherside, A., Jeon-Slaughter, H., Baine, J.L., Filbey, F.M., 2016. Discriminability of personality profiles in isolated and Co-morbid marijuana and nicotine users. Psychiatry Res. 238, 356–362.

King, M., Walker, C., Levy, G., Bottomley, C., Royston, P., Weich, S., Bellón-Saameño, J. A., Moreno, B., Svab, I., Rotar, D., Rifel, J., Maaroos, H.I., Aluoja, A., Kalda, R., Neeleman, J., Geerlings, M.I., Xavier, M., Carraça, I., Gonçalves-Pereira, M., Vicente, B., Saldivia, S., Melipillan, R., Torres-Gonzalez, F., Nazareth, I., 2008. Development and validation of an international risk prediction algorithm for episodes of major depression in general practice attendees: the PredictD study. Arch. Gen. Psychiatry 65 (12), 1368–1376.

Kong, G., Smith, A.E., McMahon, T.J., Cavallo, D.A., Schepis, T.S., Desai, R.A., Potenza, M.N., Krishnan-Sarin, S., 2013. Pubertal status, sensation-seeking, impulsivity, and substance use in high school-aged boys and girls. J. Addict. Med. 7 (2), 116–121.

Kotov, R., Gamez, W., Schmidt, F., Watson, D., 2010. Linking "big" personality traits to anxiety, depressive, and substance use disorders: a meta-analysis. Psychol. Bull. 136 (5), 768–821.

Kuhn, M., 2008. Building predictive models in R using the caret package. J. Stat. Softw.

Lee-Winn, A.E., Mendelson, T., Johnson, R.M., 2018. Associations of personality traits with marijuana use in a nationally representative sample of adolescents in the United States. Addict. Behav. Rep. 8, 51–55.

Levy, S., Weiss, R., Sherritt, L., Ziemnik, R., Spalding, A., Van Hook, S., Shrier, L.A., 2014. An electronic screen for triaging adolescent substance use by risk levels. JAMA Pediatr 168 (9), 822–828.

Liaw, A., Wiener, M., 2002. classification and Regression by randomForest. R News.

Marel, C., Sunderland, M., Mills, K.L., Slade, T., Teesson, M., Chapman, C., 2019. Conditional probabilities of substance use disorders and associated risk factors: progression from first use to use disorder on alcohol, cannabis, stimulants, sedatives and opioids. Drug Alcohol. Depend. 194, 136–142.

Martínez-Loredo, V., Fernández-Hermida, J.R., Fernández-Artamendi, S., Carballo, J.L., García-Cueto, E., García-Rodríguez, O., 2015. The association of both self-reported and behavioral impulsivity with the annual prevalence of substance use among early adolescents. Subst. Abuse Treat. Prev. Policy 10, 23.

Martínez-Loredo, V., Fernández-Hermida, J.R., La Torre-Luque, A., Fernández-Artamendi, S., 2018. Polydrug use trajectories and differences in impulsivity among adolescents. Int. J. Clin. Health Psychol. 18 (3), 235–244.

Meier, M.H., Caspi, A., Ambler, A., Harrington, H., Houts, R., Keefe, R.S., McDonald, K., Ward, A., Poulton, R., Moffitt, T.E., 2012. Persistent cannabis users show neuropsychological decline from childhood to midlife. Proc. Natl. Acad. Sci. U.S.A. 109 (40), E2657–E2664.

Meier, M.H., Hall, W., Caspi, A., Belsky, D.W., Cerdá, M., Harrington, H.L., Houts, R., Poulton, R., Moffitt, T.E., 2016. Which adolescents develop persistent substance dependence in adulthood? Using population-representative longitudinal data to inform universal risk assessment. Psychol. Med. 46 (4), 877–889.

Mitchell, M.R., Potenza, M.N., 2014. Addictions and personality traits: impulsivity and related constructs. Curr. Behav. Neurosci. Rep. 1 (1), 1–12.

Moss, H.B., Chen, C.M., Yi, H.Y., 2014. Early adolescent patterns of alcohol, cigarettes, and marijuana polysubstance use and young adult substance use outcomes in a nationally representative sample. Drug Alcohol Depend. 136, 51–62.

NCI, 2020. The Breast Cancer Risk Assessment Tool. National Cancer Institute. https://bcrisktool.cancer.gov/. (Accessed 23 Sep, 2020).

NIDA, 2019. Media Guide: Most Commonly Used Additive Drugs. National Institute on Drug Abuse. https://www.drugabuse.gov/publications/media-guide/most-commonly-used-addictive-drugs. (Accessed 23 Sep, 2020).

Oshri, A., Kogan, S.M., Kwon, J.A., Wickrama, K.A.S., Vanderbroek, L., Palmer, A.A., MacKillop, J., 2018. Impulsivity as a mechanism linking child abuse and neglect with substance use in adolescence and adulthood. Dev. Psychopathol. 30 (2), 417–435.

Pampati, S., Buu, A., Hu, Y.H., Mendes de Leon, C.F., Lin, H.C., 2018. Effects of alcohol and cigarette use on the initiation, reinitiation, and persistence of cannabis use from adolescence to emerging adulthood. Addict. Behav. 79, 144–150.

Patton, J.H., Stanford, M.S., Barratt, E.S., 1995. Factor structure of the Barratt impulsiveness scale. J. Clin. Psychol. 51 (6), 768–774.

ProCon.org, 2020. Legal Medical Marijuana States and DC. https://medicalmarijuana.procon.org/legal-medical-marijuana-states-and-dc/. (Accessed 23 Sep, 2020).

R Core Team, 2019. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Richmond-Rakerd, L.S., Fleming, K.A., Slutske, W.S., 2016. Investigating progression in substance use initiation using a discrete-time multiple event process survival mixture (MEPSUM) approach. Clin. Psychol. Sci. 4 (2), 167–182.

Richmond-Rakerd, L.S., Slutske, W.S., Wood, P.K., 2017. Age of initiation and substance use progression: a multivariate latent growth analysis. Psychol. Addict. Behav. 31 (6), 664–675.

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.C., Müller, M., 2011. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinf. 12, 77.

SAMHSA, 2016. Facing Addiction in America: The Surgeon General's Report on Alcohol, Drugs. and Health.

Stanford, M.S., Mathias, C.W., Dougherty, D.M., Lake, S.L., Anderson, N.E., Patton, J.H., 2009. Fifty years of the barratt impulsiveness scale: an update and review. Personal. Indiv. Diff. 47.

Tarter, R.E., Kirisci, L., Reynolds, M., Horner, M., Zhai, Z., Gathuru, I., Vanyukov, M., 2015. Longitudinal modeling of the association between transmissible risk, affect during drug use and development of substance use disorder. J. Addict. Med. 9 (6), 464–469.

Tate, J.C., Pomerleau, C.S., Pomerleau, O.F., 1994. Pharmacological and non-pharmacological smoking motives: a replication and extension. Addiction.

Terracciano, A., Löckenhoff, C.E., Crum, R.M., Bienvenu, O.J., Costa, P.T., 2008. Five-factor model personality profiles of drug users. BMC Psychiatry 8, 22.

Tomko, R., Williamson, N.A., McRae-Clark, A., Gray, K.M., 2019. Cannabis Use Disorder as a Developmental Disorder, in: Montoya, I., D., Weiss, S., R. B. (Eds.), Cannabis Use Disorders. Springer, pp. 189-199.

Vanyukov, M.M., Kirisci, L., Moss, L., Tarter, R.E., Reynolds, M.D., Maher, B.S., Kirillova, G.P., Ridenour, T., Clark, D.B., 2009. Measurement of the risk for substance use disorders: phenotypic and genetic analysis of an index of common liability. Behav. Genet. 39 (3), 233–244.

Venables, W.N., Ripley, B.D., 2002. Modern Applied Statistics with S, 4 ed. Springer-Verlag, New York.

Wall, M.M., Liu, J., Hasin, D.S., Blanco, C., Olfson, M., 2019. Use of marijuana exclusively for medical purposes. Drug Alcohol Depend. 195, 13–15.

Wen, H., Hockenberry, J.M., Cummings, J.R., 2015. The effect of medical marijuana laws on adolescent and adult use of marijuana, alcohol, and other substances. J. Health Econ. 42, 64–80.

WHO, 2020. Management of substance abuse: Cannabis. World Health Organization. https://www.who.int/substance_abuse/facts/cannabis/en/. (Accessed 23 Sep, 2020).

Winters, K.C., Lee, C.Y., 2008. Likelihood of developing an alcohol and cannabis use disorder during youth: association with recent use and age. Drug Alcohol. Depend. 92 (1–3), 239–247.

YRBS, 2017. Youth Risk Behavior Survey, Centers for Disease Control and Prevention.

YRRS, 2017. New Mexico Youth Risk and Resiliency Survey, Centers for Disease Control and Prevention.

Zuckerman, M., Kuhlman, D.M., Joireman, J., Teta, P., 1993. A comparison of three structural models for personality: the big three, the big five, and the alternative five. J. Pers. Soc. Psychol. 65.