*NOTE THIS PAPER IS UNDER REVIEW. PLEASE CONTACT FRANK SCHWEBEL (FSCHWEBEL@UNM.EDU) FOR MORE INFORMATION

Using Decision Trees to Identify Salient Predictors of Cannabis-Related Outcomes

Frank J. Schwebel, Ph.D.
Center on Alcohol, Substance use, And Addictions
University of New Mexico
Dylan K. Richards, Ph.D.
Center on Alcohol, Substance use, And Addictions
University of New Mexico
Rory A. Pfund, Ph.D.
Center on Alcohol, Substance use, And Addictions
University of New Mexico
Verlin W. Joseph, MPH, Ph.D.
Center on Alcohol, Substance use, And Addictions
University of New Mexico
Matthew R. Pearson, Ph.D.
Center on Alcohol, Substance use, And Addictions
University of New Mexico
Marijuana Outcomes Study Team*

Corresponding author: Matthew R. Pearson (mateo.pearson@gmail.com)

This work was supported by the National Institute of Alcoholism and Alcohol Abuse and the National Institute on Drug Abuse of the National Institutes of Health, award numbers T32 AA018108, F32 AA028712, and K01 AA023233. The content is the sole responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The authors have no conflicts of interest.

Author Note

*This project was completed by the Marijuana Outcomes Study Team (MOST), which includes the following investigators (in alphabetical order): Amber M. Anthenien, University of Houston; Adrian J. Bravo, University of New Mexico; Bradley T. Conner, Colorado State University; Robert D. Dvorak, University of Central Florida; Gregory A. Egerton, University at Buffalo; James M. Henson, Old Dominion University; John T. P. Hustad, Pennsylvania State University College of Medicine; Kevin M. King, University of Washington; Bruce S. Liese, University of Kansas; James G. Murphy, The University of Memphis; Clayton Neighbors, University of Houston; Xuan-Thanh Nguyen, University of California, Los Angeles; Jamie E. Parnes, Colorado State University; Matthew R. Pearson, University of New Mexico; Eric R. Pedersen, RAND; Mark A. Prince, Colorado State University; Sharon A. Radomski, University at Buffalo; Lara A. Ray, University of California, Los Angeles; Jennifer P. Read, University at Buffalo.

*NOTE THIS PAPER IS UNDER REVIEW. PLEASE CONTACT FRANK SCHWEBEL (FSCHWEBEL@UNM.EDU) FOR MORE INFORMATION

Abstract

Cannabis use continues to escalate among young adults and college attendance may be a risk factor for use. Severe cases of cannabis use can escalate to a cannabis use disorder which is associated with worse psychosocial functioning. Predictors of cannabis use consequences and cannabis use disorder symptom severity have been identified; however, they typically employ a narrow set of predictors and rely on linear models. Machine learning is well-suited for exploratory data analyses of high-dimensional data. This study applied decision tree learning to identify predictors of cannabis user status, negative cannabis-related consequences, and cannabis use disorder symptoms. Undergraduate college students (N=7000) were recruited from nine universities in nine states across the US. Among the 11 trees, 36 splits created by 17 distinct predictors were identified. Consistent with prior research, one's beliefs about cannabis were strong predictors of user status. Negative reinforcement cannabis use motives were the most consistent predictors of cannabis use disorder symptoms and past month cannabis use was the most consistent predictor of probable cannabis use disorder. Typical frequency of cannabis use was the only predictor of negative cannabis-related consequences. Our results demonstrate that decision trees are a useful methodological tool for identifying targets for future clinical research.

Keywords: machine learning, decision tree, recursive partitioning, cannabis, college students

*NOTE THIS PAPER IS UNDER REVIEW. PLEASE CONTACT FRANK SCHWEBEL (FSCHWEBEL@UNM.EDU) FOR MORE INFORMATION

1.0 Introduction

A recent national survey demonstrates continued escalation of the prevalence of cannabis use among young adults (Schulenberg et al., 2020). Specifically, 40% and 27% of young adults reported using cannabis in the past year and past month, respectively (Schulenberg et al., 2020). Although rates of cannabis use are similar for college students and their same-age noncollege peers, college attendance may be a risk factor for cannabis use. College attendance has been associated with a 51% increased probability of using cannabis in the past year for young adults who had not used cannabis by the 12th grade (Miech et al., 2017). Further, a large, multi-site (11 U.S. universities) survey of cannabis use among college students found that the vast majority (90.8%) of past month users experience at least one negative consequence related to their use (Pearson et al., 2017b).

In the most severe cases, cannabis use escalates into a clinically significant concern, cannabis use disorder (CUD). The prevalence of past year CUD is estimated to be 8.6% among U.S. college students (Arterberry et al., 2020) and CUD prevalence rates increase with time since the initiation of cannabis use (Han et al., 2019). A CUD diagnosis is related to significantly worse psychosocial functioning among regular cannabis users (Foster et al., 2018). Given the public health burden of cannabis use and CUD among college students, research is needed to identify salient correlates to target for intervention.

A growing literature seeks to identify the predictors of cannabis use, negative cannabis-related consequences, and CUD symptom severity. Multiple predictors have been identified across studies, including personality traits (e.g., sensation seeking; Galbraith & Connor, 2015), normative perceptions (Pearson et al., 2017a), cannabis use motives (e.g., coping, Moitra et al., 2015), and use of protective behavioral strategies (PBS; Pedersen et al., 2016). However, these

prior studies are limited due to their focus on a narrow set of predictors and reliance on linear models. Focusing on a narrow set of predictors prohibits the functional relation between independent variables, and linear models are limited in their classification ability (Strobl et al., 2009). Instead, this literature would benefit from analyses considering the combination of multiple independent variables to understand their predictive value in explaining cannabis user status, negative cannabis-related consequences, and CUD (symptoms).

Machine learning is a branch of quantitative methodology arising from computer science and artificial intelligence (Michalski et al., 2013). A subtype of machine learning, decision tree learning, is a promising tool for explaining the predictive value of multiple independent variables related to cannabis-related outcomes. Decision tree learning involves developing parsimonious predictive models and provides decision rules for predicting both categorical and continuous outcomes. The algorithm for decision tree learning finds the split on a predictor variable that best distinguishes between two distinct groups on an outcome. Following each split, the same algorithm using all possible predictor variables (including variables from the previous split), determines the next split. The algorithm is repeated until each terminal node contains a relatively homogeneous subsample. Decision tree models are particularly well-suited to handle high-dimensional data and can process large number of predictor variables simultaneously (Strobl et al., 2009). Recursive partitioning is one type of decision tree learning model and is useful for exploratory data analyses.

Decision tree learning has been used previously to identify a set of predictors of lifetime and past month cannabis use status and negative cannabis-related consequences among a large sample (*N*=8,141) of U.S. college students across 11 universities (Wilson et al., 2018). The authors entered more than 100 predictors (e.g., emotion regulation, impulsivity, cannabis use

motives) into the decision tree models. For both lifetime and past month cannabis use status, the best set of predictors were cannabis norms variables and identification as a cannabis user. For negative cannabis-related consequences, the best set of predictors were frequency of cannabis use, use of cannabis protective behavioral strategies, and positive/negative urgency (i.e., impulsivity-like traits when experiencing positive/negative emotions, respectively). These findings suggest that decision trees may provide insight into targets for cannabis intervention and that an examination of predictors of CUD symptoms and of probable CUD is warranted.

1.1 Present Study

The present study applies decision tree learning, an exploratory, easily interpretable data analytic approach, to identify salient predictors of a wide range of cannabis-related outcomes. Wilson et al. (2018) found that predictors of user status were quite distinct from predictors of negative cannabis-related consequences. To extend this work, we also examined predictors of CUD symptoms, measured both continuously and as a binarized outcome that separate individuals with and without a probable CUD.

2.0 Method

2.1 Participants and Procedure

Participants were 7000 U.S. undergraduate college students recruited between Fall 2016 and Spring 2017 from universities in two states where recreational cannabis use was legal (CO, WA), three states where medical cannabis was legal (CA, NM, NY), and four states where cannabis use was illegal (FL, TN, TX, VA). Convenience sampling was used to recruit participants enrolled in psychology courses to complete an online survey for partial course credit. Further characteristics of this sample have been reported elsewhere (Pearson et al., 2020). The IRB at each participating university approved the study procedures.

2.2 Measures

Supplemental Table 1 provides a summary of how each predictor was measured. A planned missingness design (Graham et al., 2006) was used to minimize participant burden so sample sizes varied across measures.

2.2.1 Cannabis User Status

A single item asked whether participants ever used cannabis in their lifetime (*lifetime use*), and a single item asked if participants used cannabis in the past month (*past month use*).

2.2.2 Cannabis Consequences

The 21-item Marijuana Consequences Questionnaire Short Form (MACQ; Simons, Dvorak, Merrill, & Read, 2010) was used to capture domains of negative cannabis-related consequences during the past 30 days. The 21-item scale captures eight domains of cannabis consequences: social-interpersonal consequences, impaired control, negative self-perception, self-care, risk behaviors, academic/occupational consequences, physical dependence, and blackout use. Participants were queried on whether they experienced each of these consequences due to their marijuana use in the past 30 days (0=no, 1=yes). A single total score was calculated to represent the overall number of unique negative cannabis-related consequences experienced in the past 30 days.

2.2.3 DSM-5 Cannabis Use Disorder Symptoms

Past month cannabis users completed distinct measures of CUD symptoms. All completed the Self-Reported Symptoms of Cannabis Use Disorder (SRSCUD, Richards et al., 2020), a 13-item measure developed and validated by the research team to capture symptoms of CUD on a continuum of severity as defined by the DSM-5. Each item was assessed on a 4-point scale (0=*not at all*, 1=*very little*, 2=*somewhat*, 3=*to a great extent*). Sample items for the measure

included: "In the past month, have you used marijuana in larger amounts or over a longer period than was intended?" and "In the past month, have you had a persistent desire or tried unsuccessfully to cut down or control my marijuana use?" We examined the average of all SRSCUD items as a continuous outcome (Model 4) as well as a binarized score reflective of probable CUD (≥6.5 summed score, Model 5).

To minimize response burden/measure redundancy, participants were randomized to complete one of the other three measures of CUD symptoms: the 8-item Cannabis Use Disorder Identification Test-Revised (CUDIT-R; Adamson et al., 2010, *n*=632), the 6-item Cannabis Abuse Screening Test (CAST; Legleye, Karila, Beck, & Reynaud, 2007, *n*=679), and the 5-item Severity of Dependence Scale (SDS; Gossop et al., 1995, *n*=712). For each of these measures, we examined them both as continuous scores (sum for CUDIT-R, Model 6; average for CAST, Model 8; average for SDS, Model 10) as well as binarized scores reflective of those with and without a probable CUD (≥13 for summed CUDIT-R, Model 7; ≥3 for summed CAST, Model 9; ≥3 for summed SDS, Model 11).

2.3 Analysis Plan

We tested 11 models with different baseline predictor variables using decision tree learning using the 'rpart' package (Therneau & Atkinson, 2019) in R (R Core Team, 2019). Decision tree learning was selected as the analytic approach due to the relative ease and intuitiveness of model interpretation. Predictors were classified as being either cannabis use indicators (n=40) (e.g., age of first use, typical frequency, typical quantity) or other psychosocial/cannabis use-related constructs (n=115) (e.g., Protective Behavioral Strategies-Marijuana, UPPS-P) (for a complete list of predictors see Supplemental Table 1). Demographic variables were not included in the models because they are (generally) fixed factors that are not

treatment targets. All predictors were included in each model unless otherwise noted. Measures of CUD were modeled using both continuous scoring and dichotomous coding as being above (1) or below (0) the reported cutoff for probable CUD for each individual measure. Models 1 and 2 were dichotomous measures (0=no, 1=yes) of lifetime cannabis use and past month cannabis use, respectively. For Models 1 and 2, measures that were only assessed among past month cannabis users were excluded as predictors given that they were missing for all non-users. Model 3 was a continuous measure of negative cannabis-related consequences measured using the MACQ. Each measure of CUD symptoms was examined as a continuous outcome (SRSCUD Model 4, CUDIT-R Model 6, CAST Model 8, and SDS Model 10) as well as a dichotomous outcome based on established cutoffs for probable CUD (SRSCUD Model 5, CUDIT-R Model 7, CAST Model 9, and SDS Model 11) (see Supplementary Materials for binary results). To create parsimonious models and to decrease the risk of overfitting the models, branches that did not improve prediction accuracy in cross-validation (k=10) were removed or "pruned". A complexity parameter of 0.03 was selected based on recommendations by Ture and Omurlu (2018).

3.0 Results

3.1 Preliminary Analyses

In the full sample of 7000 participants, 57.3% (*n*=3979) reported using cannabis in their lifetime, and 52.3% (*n*=2077) reported using cannabis in the past month. Among past month users, our sample used cannabis on average 5.86 time periods (i.e., 4-hour blocks of time, *SD*=7.91) during a typical week, averaging around 6.03 grams per week (*SD*=12.04, Winsorized). Our sample reported experiencing an average of 3.64 negative cannabis-related consequences (*SD*=3.90) in the past month. Based on cutoff scores established for each of our CUD symptom measures, 7.7% (CAST) to 34.4% (SRSCUD) of our sample were determined to

have a probable CUD, reflecting the different sensitivity and specificity for each of these measures.

3.2 Decision Tree Learning

Decision trees are presented as trees in Figures 1-5 (and Supplemental Figures 1-4). These trees are visually intuitive, and the authors suggest viewing the trees while reading through the written results (which are more challenging to interpret without the use of trees as a visual aid).

3.2.1 Lifetime Cannabis Use

Two variables contributed to the final tree predicting lifetime cannabis user status (see Figure 1-Model 1): ease of obtaining cannabis and pros of cannabis. In the overall sample (n=6906), 57.3% of individuals reported lifetime cannabis use. Among individuals who reported it being "probably impossible," "very difficult," or "fairly difficult" to obtain cannabis (n=2142), 27.9% reported lifetime cannabis use (classification: non-cannabis user). Among individuals who reported it being "fairly easy" or "very easy" to obtain cannabis (n=4764), 70.6% reported lifetime cannabis use. This group was split based on perceived pros of cannabis based on the Marijuana Decision Balance scale. The response scale for the pros of cannabis subscale range from 1=*Not Important* to 5=*Extremely Important*. Among individuals (n=1151) who reported potential pros of cannabis as not being very important to their decision to use/not use cannabis (<2.16), 42.6% reported lifetime cannabis use (classification: non-cannabis user). Among individuals (n=3613) with rated pros of cannabis as more important to their decision to use/not use cannabis (≥2.16), 79.5% reported lifetime cannabis use (classification: cannabis user).

3.2.2 Past Month Cannabis Use

Three variables contributed to the final tree predicting past month cannabis user status (see Figure 1-Model 2): time spent engaging in substance use activities, close friends' descriptive norms frequency, and cons of cannabis. Among the overall sample (*n*=3960), 52.5% reported past month cannabis use. Among individuals who reported spending 0 hours per week engaging in substance use activities (*n*=712), 19% reported past month cannabis use (classification: non-past month cannabis user). Among individuals who spent more than 0 hours per week engaging in substance use activities (*n*=3248), 59.8% reported past month cannabis use. This group was split based on their perceptions of how frequently their close friends use cannabis during a typical week. Among individuals (*n*=351) who perceive their close friends to abstain from cannabis during a typical week (<.5), 21.9% reported past month cannabis use (classification: non-past month cannabis user). Among individuals (*n*=2897) who perceive their close friends to use at least once during a typical week (≥0.5), 64.4% reported past month cannabis use. This group was split based on their cons of cannabis use based on the Marijuana Decisional Balance scale. The response scale for the cons of cannabis subscale range from 1=*Not Important* to 5=*Extremely Important*. Among individuals (*n*=1151) who rated cons of cannabis as important to their decision to use/not use cannabis (≥3.13), 47.5% reported past month cannabis use (classification: non-past month cannabis user). Among individuals (*n*=1746) who rated cons of cannabis as less important to their decision to use/not use cannabis (<3.13), 75.5% reported past month cannabis use (classification: past month cannabis user).

3.2.3 Negative Cannabis-Related Consequences (MACQ)

One variable contributed to the final tree predictive negative cannabis-related consequences (see Figure 1-Model 3): typical frequency of cannabis use based on the Marijuana Use Grid. The mean number of consequences among the overall sample (*n*=2036) was 3.64.

Individuals who reported using cannabis on 4 or fewer time periods during a typical week (*n*=1320) reported an average of 2.52 negative cannabis-related consequences (classification: low consequences). Individuals who reported using cannabis on 5 or more time periods during a typical week (*n*=716) reported an average of 5.72 negative consequences (classification: high consequences).

3.2.4 SRSCUD Continuous Symptom Score

Three variables contributed to the final tree predicting CUD symptoms based on the SRSCUD (see Figure 2-Model 4): coping motives, past month cannabis frequency, and conformity motives. The average CUD symptom severity was low (*M*=1.48; 1=*Not at all*, 2=*Very little*, 3=*Somewhat*, 4=*To a great extent*) in the sample of past month cannabis users (*n*=2023). Individuals (*n*=1052) with low average coping motive scores (<1.9, less than "2=*Some of the time*") reported below average CUD symptom severity (1.24). This group was further split between an even lower risk group (1.15) who reported using cannabis on less than 7 or fewer days during the past month (*n*=745; classification: low CUD symptoms), and a slightly higher risk group (1.46) who used cannabis on 8 or more days during the past month (*n*=307; classification: low CUD symptoms). Individuals (*n*=971) with moderate to high average coping motive scores (≥1.9) reported an average of 1.75 symptoms. This group was split based on average conformity motive scores. Individuals (*n*=773) with low to moderate average conformity motive scores (<2.55) reported an average of 1.61 symptoms. This group was split based on past month cannabis frequency. Individuals who used cannabis on 10 or fewer days during the past month (*n*=425), reported an average of 1.42 symptoms (classification: low CUD symptoms), whereas those who used cannabis on 11 or more days during the past month (*n*=348) reported an average of 1.85 symptoms (classification: low CUD symptoms). Individuals (*n*=198) with

moderate to high average conformity motive scores (≥2.55) reported an average of 2.28 symptoms. This group was split based on average coping motive scores. Individuals (*n*=188) with non-maximum average coping motive scores (<4.9) reported an average of 2.20 symptoms (classification: moderate CUD symptoms). Individuals (*n*=10) with maximum average coping motive scores (≥4.9) reported an average of 3.66 symptoms (classification: high CUD symptoms).

3.2.5 CUDIT-R Continuous Symptom Score

Two variables contributed to the final tree predicting CUD symptoms based on the CUDIT-R (see Figure 2-Model 6): past month cannabis frequency and coping motives. The average CUDIT-R score reported among the overall sample (*n*=632) were 8.15. Individuals who used cannabis on 16 or fewer days during the past month (*n*=467) reported an average of 6.15 symptoms. This group was split based on average coping motive score. Individuals (*n*=214) with a low average coping motive score (<1.5) reported an average CUDIT-R score of 4.47 (classification: low symptoms) whereas those (*n*=253) with a moderate to high average coping motive score (≥1.5) reported an average CUDIT-R score of 7.56 (classification: low symptoms). Individuals who used cannabis 17 or more days during the past month (*n*=165) reported an average of 13.81 symptoms. This group was split based on average coping motive score. Individuals (*n*=87) with a low to moderate average coping motive score (<2.7) reported an average CUDIT-R score of 11.66 (classification: moderate symptoms), and Individuals (*n*=78) with a moderate to high average coping motive score (≥2.7) reported an average CUDIT-R score of 16.22 (classification: high symptoms).

3.2.6 CAST Continuous Symptom Score

Four variables contributed to the final tree predicting CUD symptoms based on the CAST (see Figure 2-Model 8): money spent on cannabis, conformity motives, percent time using cannabis alone, and use of cannabis protective behavioral strategies. The average CAST score reported among the overall sample (*n*=679) was 1.85. Individuals who spent less than $32.50 on cannabis during the past month (*n*=468) reported an average of 1.56 symptoms. This group was split based on average conformity motive score. Individuals (*n*=42) with a moderate to high average conformity motive score (≥2.9) reported an average CAST score of 2.35 (classification: moderate symptoms). Individuals (*n*=426) with a low to moderate average conformity motive score (<2.9) reported an average CAST score of 1.48. This group was split based on percent time using cannabis alone. Individuals (*n*=335) who did not report using cannabis alone (<.5%) during the past month reported an average CAST score of 1.37 (classification: low symptoms). Individuals (*n*=91) who did reporting using cannabis alone (>.5%) during the past month reported an average CAST score of 1.89 (classification: low symptoms). Individuals who spent $32.50 or more on cannabis during the past month (*n*=211) reported an average of 2.48 symptoms. This group was split based on average conformity motive score such that individuals (*n*=27) with a moderate to high average conformity motive score (≥2.5) reported an average of 3.33 symptoms (classification: high symptoms) whereas individuals (*n*=184) with a low to moderate average conformity motive score (<2.5) reported an average of 2.36 symptoms. This group was split based on use of protective behavioral strategies. Individuals (*n*=51) who reported higher use of protective behavioral strategies (≥4.32) reported an average CAST score of 1.83 (classification: low symptoms) whereas individuals (*n*=133) who reported lower use of protective behavioral strategies (<4.32) reported an average CAST score of 2.56 (classification: moderate symptoms).

3.2.7 SDS Continuous Symptom Score

Three variables contributed to the final tree predicting CUD symptoms based on the SDS (see Figure 2-Model 10): conformity motives, past month cannabis frequency, and percent of black-market purchases of cannabis (i.e., bought not from a dispensary). The average number SDS score reported among the overall sample (*n*=712) was 1.55. Individuals (*n*=602) with a low average conformity motive score (<2.10) reported an average SDS score of 1.06. This group was split based on days of past month cannabis use. Individuals who used cannabis 14 or fewer days during the past month (*n*=429) reported an average of 0.64 symptoms (classification: low symptoms). Individuals who used cannabis on 15 or more days during the past month (*n*=173) reported an average of 2.09 symptoms (classification: moderate symptoms). Individuals (*n*=110) with a moderate to high average conformity motive score (≥2.10) reported an average SDS score of 4.25. This group was split based on percentage of cannabis purchased through black market sources during the past month. Individuals (*n*=60) who purchased less than 15% of cannabis from the black market during the past month reported an average SDS score of 2.73 (classification: moderate symptoms). Individuals (*n*=50) who purchased at least 15% of cannabis from the black market during the past month reported an average SDS score of 6.06 (classification: high symptoms).

4.0 Discussion

This study applied decision tree learning to identify clinically relevant predictors of cannabis-related outcomes. Across all models, 17 distinct predictors were identified: ease of obtaining cannabis, pros of cannabis use, cons of cannabis use, time spent engaging in substance use activities, close friends descriptive norms frequency, typical frequency of cannabis use, conformity motives, coping motives, past month cannabis frequency, money spent on cannabis,

percent time using cannabis alone, use of cannabis protective behavioral strategies, percent of black-market purchases of cannabis, global negative effect expectancies, typical college student injunctive norms quantity, typical quantity of cannabis used (grams), and perceived helpfulness of protective behavioral strategies (11 trees, 36 splits).

Consistent with Wilson et al. (2018), predictors of user status were largely related to one's beliefs about cannabis. Although Wilson et al. (2018) identified largely normative beliefs as predictors of user status, we also found higher pros about the effects of cannabis (for lifetime user status) and lower cons about the effects of cannabis (for past month user status) from a marijuana decisional balance measure (Elliott et al., 2011) as predictors of user status. In terms of normative variables, we found the perceived frequency of cannabis use by close friends (i.e., descriptive norms) measured by the Marijuana Norms Grid (Montes et al., 2020) predicted past month user status.

From a behavioral economic measure of time allocation, we found time spent engaging in substance use activities was predictive of past month user status. Not surprisingly, this predictor was a strong indicator of user status despite the measure broadly assessing engagement in substance use activities (as opposed to substance-specific engagement). Similarly, perceived ease of obtaining cannabis was a strong predictor of user status. Individuals who easily obtained cannabis were more likely to be lifetime users as opposed to those who reported difficulty obtaining cannabis.

To our surprise, typical frequency of cannabis use as assessed by the Marijuana Use Grid (Pearson et al., 2020) was the only salient predictor of negative cannabis-related consequences. This result may provide further support for the consistent relationship found between frequency of cannabis use and consequences (Looby & Earleywine, 2007; Pearson, 2019).

Across the models predicting CUD symptoms (models 4, 6, 8, 10), negative reinforcement cannabis use motives (i.e., coping and conformity motives) were the most consistent predictors across models (8 splits). Individuals with higher coping and conformity motive scores typically experienced more CUD symptoms. Past month cannabis use was the next most consistent predictor across models, with individuals who reported more frequent use experiencing more symptoms (4 splits).

Across the models predicting probable CUD (models 5, 7, 9, 11), we found significantly more heterogeneity in predictors. The most consistent predictor was past month cannabis use (3 splits), followed by negative reinforcement cannabis motives (2 splits), and measures of time spent engaging in substance use activities (2 splits). Each predictor split was in the expected direction, with higher risk factor scores (e.g., higher conformity motives) being associated with an increased likelihood of probable CUD.

The predictors contributing to the binary and continuous models of the SRSCUD were the same and a similar consistent pattern of predictors (with different predictors) was found with the CUDIT-R. Although the model splits and shapes were not identical across models, they were quite similar. The predictors contributing to the binary and continuous models of the CAST and SDS were not the same, although there were some shared predictors and similar tree shapes. This may be related to the individual measure content. The SRSCUD was developed as a direct measure of CUD symptoms and although the CUDIT-R measures consumption, it is the most well-validated self-report measure of CUD symptoms. The CAST includes 2 (of a total of 6) items that measure consumption content ("Have you ever smoked cannabis… before midday? when you were alone?"). Given the stem of the CAST trees assessed money spent on cannabis, it is possible the item assessing using alone is driving the results (i.e., likely need to spend money

on cannabis to use alone). The SDS was validated with a sample of individuals who use heroin, cocaine, and amphetamine and includes an item that clearly assesses withdrawal ("Did the prospect of missing a fix (or dose) or not chasing make you anxious or worried?"). Although there is evidence of cannabis withdrawal (Budney et al., 2004), it appears to be less severe and more focused on emotional and behavioral symptoms than other substances (Budney & Hughes, 2006).

Clinically, these results can be applied by considering salient variables alone or in conjunction with the use of existing measures of CUD. For example, if an individual completes the SDS and has a score near the clinical cutoff for probable CUD, assessing the impact conformity motives has on their cannabis use may help clarify the result/diagnosis. Individuals reporting low average conformity motives are considerably less likely to meet the cutoff for probable CUD relative to those reporting moderate to high average conformity motives.

Predictor variables can also be applied during therapy by assessing salient variables in a conversational manner. In lieu of (or in conjunction with) a measure of motives for use, a clinician might ask, "What are some of the reasons you use cannabis?" and then follow-up with relevant questions/reflections about using to cope ("Using after experiencing stressful situations helps you feel better") or conform ("You mainly use when you're around your friends or people you look up to").

Given the sociodemographic and geographical diversity of our large sample, we expect our results are likely to generalize to the broader U.S. college student population. However, it is unclear whether such findings would be similar among non-college attending adults, or in distinct cultural milieu (i.e., across cultures). Although our analytic methods were very similar to Wilson et al. (2018), we could not easily ascertain the degree of replication given the assessment

batteries across these studies differed substantially. In other words, differences in predictors identified between these two studies could be due to differences in these assessment batteries.

4.1 Conclusion

The present study applied decision tree learning to identify clinically relevant predictors of cannabis user status, negative cannabis-related consequences, and CUD symptoms. This study provides further support for beliefs about cannabis predicting user status and found that negative reinforcement cannabis use motives are associated with increased CUD symptoms. Findings support the continued use of exploratory data analytic techniques to identify variables that might predict CUD symptoms and probable-CUD diagnosis. Findings may be helpful clinically to diagnose CUD and inform case conceptualization and treatment targets.

References

Adamson, S. J., Kay-Lambkin, F. J., Baker, A. L., Lewin, T. J., Thornton, L., Kelly, B. J., & Sellman, J. D. (2010). An improved brief measure of cannabis misuse: the Cannabis Use Disorders Identification Test-Revised (CUDIT-R). *Drug and Alcohol Dependence, 110*(1-2), 137-143.

Arterberry, B. J., Boyd, C. J., West, B. T., Schepis, T. S., & McCabe, S. E. (2020). DSM-5 substance use disorders among college-age young adults in the United States: Prevalence, remission and treatment. *Journal of American College Health*, *68*, 650-657.

Budney, A. J., & Hughes, J. R. (2006). The cannabis withdrawal syndrome. *Current Opinion in Psychiatry*, *19*(3), 233-238.

Budney, A. J., Hughes, J. R., Moore, B. A., & Vandrey, R. (2004). Review of the validity and significance of cannabis withdrawal syndrome. *American Journal of Psychiatry*, *161*(11), 1967-1977.

Elliott, J. C., Carey, K. B., & Scott-Sheldon, L. A. (2011). Development of a decisional balance scale for young adult marijuana use. *Psychology of Addictive Behaviors, 25*(1), 90-100.

Foster, K. T., Arterberry, B. J., Iacono, W. G., McGue, M., & Hicks, B. M. (2018). Psychosocial functioning among regular cannabis users with and without cannabis use disorder. *Psychological Medicine*, *48*, 1853-1861.

Galbraith, T., & Conner, B. T. (2015). Religiosity as a moderator of the relation between sensation seeking and substance use for college-aged individuals. *Psychology of Addictive Behaviors, 29*, 168–175.

Gossop, M., Darke, S., Griffiths, P., Hando, J., Powis, B., Hall, W., & Strang, J. (1995). The Severity of Dependence Scale (SDS): psychometric properties of the SDS in English and Australian samples of heroin, cocaine and amphetamine users. *Addiction, 90*(5), 607-614.

Graham, J. W., Taylor, B. J., Olchowski, A. E., & Cumsille, P. E. (2006). Planned missing data designs in psychological research. *Psychological Methods, 11*(4), 323.

Han, B., Compton, W. M., Blanco, C., & Jones, C. M. (2019). Time since first cannabis use and 12-month prevalence of cannabis use disorder among youth and emerging adults in the United States. *Addiction*, *114*, 698-707.

Legleye, S., Karila, L., Beck, F., & Reynaud, M. (2007). Validation of the CAST, a general population Cannabis Abuse Screening Test. *Journal of Substance Use, 12*(4), 233-242.

Looby, A., & Earleywine, M. (2007). Negative consequences associated with dependence in daily cannabis users. *Substance Abuse Treatment, Prevention, and Policy*, *2*(1), 3.

Michalski, R. S., Carbonell, J. G., & Mitchell, T. M. (Eds.). (2013). *Machine learning: An artificial intelligence approach*. New York, NY: Springer Science+Business Media.

Miech, R. A., Patrick, M. E., O'Malley, P. M., & Johnston, L. D. (2017). The influence of college attendance on risk for marijuana initiation in the United States: 1977 to 2015. *American Journal of Public Health*, *107*, 996-1002.

Moitra, E., Christopher, P. P., Anderson, B. J., & Stein, M. D. (2015). Coping-motivated marijuana use correlates with DSM-5 cannabis use disorder and psychological distress among emerging adults. *Psychology of Addictive Behaviors, 29*, 627–632.

Montes, K. S., Richard, D. K., Pearson, M. R., & Marijuana Outcomes Study Team (2020). A novel approach to assess descriptive and injunctive norms for college student marijuana use. *Addictive Behaviors*.

Pearson, M. R. (2019). A meta-analytic investigation of the associations between cannabis use and cannabis-related negative consequences. *Psychology of Addictive Behaviors*, *33*(3), 190.

Pearson, M. R., Kholodkov, T., Gray, M. J., & Marijuana Outcomes Study Team. (2017a). Perceived Importance of Marijuana to the College Experience Scale (PIMCES): Initial development and validation. *Journal of Studies on Alcohol and Drugs, 78*, 319–324.

Pearson, M. R., Liese, B. S., Dvorak, R. D., & Marijuana Outcomes Study Team. (2017b). College student marijuana involvement: Perceptions, use, and consequences across 11 college campuses. *Addictive Behaviors, 66*, 83–89.

Pearson M. R., Marijuana Outcomes Study Team, & Protective Strategies Study Team. (2020). Marijuana Use Grid: A brief, comprehensive measure of cannabis use. Unpublished manuscript.

Pedersen, E. R., Hummer, J. F., Rinker, D. V., Traylor, Z. K., & Neighbors, C. (2016). Measuring protective behavioral strategies for marijuana use among young adults. *Journal of Studies on Alcohol and Drugs, 77*, 441–450.

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Richards, D. K., Schwebel, F. J., Sotelo, M., Pearson, M. R., & Marijuana Outcomes Study Team (2020). Self-Reported Symptoms of Cannabis Use Disorder (SRSCUD): Psychometric Testing and Validation. *Experimental and Clinical Psychoharmacology*.

Schulenberg, J. E., Johnston, L. D., O'Malley, P. M., Bachman, J. G., Miech, R. A. & Patrick, M. E. (2020). *Monitoring the Future national survey results on drug use, 1975–2019:*

Volume II, College students and adults ages 19–60. Ann Arbor: Institute for Social Research, The University of Michigan. Available at http://monitoringthefuture.org/pubs.html#monographs

Simons, J. S., Dvorak, R. D., Merrill, J. E., & Read, J. P. (2012). Dimensions and severity of marijuana consequences: Development and validation of the Marijuana Consequences Questionnaire (MACQ). Addictive Behaviors, 37(5), 613-621.

Strobl, C., Malley, J., & Tutz, G. (2009). An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological Methods, 14*, 323–348.

Therneau, T. & Atkinson, B. (2019). rpart: Recursive Partitioning and Regression Trees. R package version 4.1-15. https://CRAN.R-project.org/package=rpart

Ture, M., & Kurt Omurlu, I. (2018). Determining of complexity parameter for recursive partitioning trees by simulation of survival data and an application on breast cancer data. *Journal of Statistics and Management Systems, 21*(1), 125-138.

Wilson, A. D., Montes, K. S., Bravo, A. J., Conner, B. T., Pearson, M. R., & Marijuana Outcomes Study Team. (2018). Making decisions with trees: Examining marijuana outcomes among college students using recursive partitioning. *Clinical Psychological Science*, *6*, 744-754.