HHS Public Access
Author manuscript
Clin Psychol Sci. Author manuscript; available in PMC 2019 September 01.

Published in final edited form as:
Clin Psychol Sci. 2018 September ; 6(5): 744–754. doi:10.1177/2167702618775405.

Making Decisions with Trees: Examining Marijuana Outcomes among College Students using Recursive Partitioning

Wilson AD[1], **Montes KS**[1], **Bravo AJ**[1], **Conner BT**[2], and **Pearson MR**[1] **Marijuana Outcomes Study Team**

[1]University of New Mexico

[2]Colorado State University

Abstract

Exploratory analyses were used to identify a unique constellations of variables that are associated with marijuana use outcomes among college students. We used recursive partitioning to examine over 100 putative antecedents of lifetime marijuana user status, past-month marijuana user status, and negative marijuana-related consequences. Participants (*N*=8141) completed measures online across 11 sites in the USA. Norms (descriptive, injunctive, and internalized norms) and marijuana identity best distinguished marijuana users from non-users (i.e., lifetime/past month), whereas marijuana use frequency, use of protective behavioral strategies, and positive/negative urgency best distinguished the degree to which users reported negative consequences. Our results demonstrate that tree-based modeling is a useful methodological tool in the selection of targets for future clinical research. Additional research is needed to determine if these factors are causal antecedents, rather than consequences or epiphenomena. We hope this large sample study provides the impetus to develop intervention strategies targeting these factors.

Keywords

Recursive Partitioning; Norms; Protective Behavioral Strategies; Marijuana Consequences

Introduction

Machine learning (ML; Michalski et al., 2013), or statistical learning (James et al., 2013), is a branch of quantitative methods borne from the fields of computer science and artificial intelligence. The past decade has witnessed increased applications of ML approaches in the behavioral sciences (Conner et al., 2010; Hellemann et al., 2009; King & Resick, 2014; Pearson et al., 2012). One sub-type of ML, decision tree learning, is particularly well suited to developing parsimonious predictive models from behavioral data (Hellemann et al.,

*Corresponding author: Matthew Pearson, Ph.D., Center on Alcoholism, Substance Abuse, and Addictions, University of New Mexico, 2650 Yale SE MSC 11-6280, Albuquerque, NM, 87106; Pn:(505)925-2322; mateo.pearson@gmail.com.
Author Contributions
A.D.W. performed the data analysis and drafted portions of the introduction/results/discussion. K.S.M. drafted the abstract/table/sections of method. A.J.B./B.T.C. drafted sections of the discussion. M.R.P. developed the research question, created the Figures, and drafted sections of the introduction/method/results/discussion. MOST collected the data. All authors provided critical revisions/approved of final manuscript.

2009). Though not without drawbacks (King & Resick, 2014; Marshall, 2001), tree-based modeling has advantages over traditional correlational and regression-based analyses; namely, tree-based modeling does not rely on assumptive statistical tests and effects determined by *p*-values less than .05. Additionally, tree-based modeling can parse the relative predictive utility of large numbers of variables (i.e., handle high dimensional data), including the >100 variables collected in the present study. Moreover, tree-based modeling is particularly suited for the discovery of hierarchical structure that defines more homogeneous groups within larger heterogeneous samples. This hierarchical structure is rarely hypothesized but likely has more external validity than looking at bivariate relations between independent and dependent variables or examining a single mediator or moderator relation. In other words, using tree-based modeling enables researchers to explore their data to discover hierarchical structures that more accurately predict relative risk on an outcome of interest.

In over 40 years of the Monitoring the Future study (Johnston et al., 2016), marijuana use rates have ebbed and flowed, but two trends have been consistent: 1) after caffeine and alcohol, marijuana is the most commonly used psychoactive substance in the United States, and 2) marijuana use rates peak during the traditional college years. Further, college students are at increased risk of cannabis use disorder (Caldeira et al., 2008) relative to the general population (Hasin et al., 2015). Therefore, it is imperative to determine the factors that confer risk of problematic marijuana use in this population.

Traditional correlation and regression-based approaches have identified many distinct risk/ protective factors for marijuana-related outcomes among college students. These factors range from relatively distal antecedents like personality traits (e.g., impulsivity-like traits, sensation seeking, anxiety sensitivity, hopelessness; Cyders et al., 2007; Galbraith & Conner, 2015; Whiteside & Lynam, 2001; Woicik et al., 2009) and more proximal antecedents like normative perceptions (e.g., descriptive/injunctive norms, internalization of college marijuana use culture; Napper et al., 2015; Pearson et al., 2017c), reasons for using marijuana (e.g., coping/enhancement/conformity/social/expansion motives; Simons et al., 1998), and use of protective behavioral strategies to offset the negative consequences associated with marijuana use (Pedersen et al., 2016).

In the present study, several constructs were selected based on previous marijuana research or the extant substance use literature. For example, the following variables (along with additional variables examined in the current study) were selected based on the following research findings: (a) frequency of marijuana use has been found to be positively associated with marijuana-related consequences (Simons et al., 2012); (b) positive urgency has been found to be positively associated with substance use (Zapolski et al., 2009); (c) the use of marijuana-related protective behavioral strategies has been found to be negatively associated with frequency of marijuana use and marijuana-related consequences (Pedersen et al., 2016); (d) marijuana-related injunctive (i.e., typical college student) and descriptive norms (i.e., heaviest descriptive norms) have been found to be positively associated with frequency of marijuana use (Buckner, 2013); (e) identification with being a drinker has been found to positively associated with frequency of alcohol use and alcohol-related consequences (Lindgren et al., 2013); thus, identification with being a marijuana user should be positively associated with frequency of marijuana use and marijuana-related consequences; and (f) internalization of college alcohol use culture has been found to be positively associated with frequency of marijuana use (Osberg et al., 2010). What is not known is the extent to which there are hierarchical constellations of the variables associated with specific marijuana-related outcomes (e.g., ever using marijuana, current marijuana use, and marijuana-related consequences). The current study attempts to discover hierarchical structures of correlates of marijuana use and outcomes using an exploratory data analysis technique to increase our understanding of what accounts for marijuana user status and marijuana-related negative consequences.

Clinically, research on risk factors for marijuana-related outcomes may be limited by methods of investigation and analysis. For example, most studies focus on a narrow set of risk factors (or even a single risk factor) that are selected *a priori* by a given research team. Further, researchers explicitly choose the functional relation between independent variables and a given outcome, with the majority of the field electing to use additive, linear models that generally have limited classification ability (see Strobl et al., 2009; Yarkoni & Westfall, 2017). We posit that recursive partitioning (RP; Zhang & Singer, 2013), a decision tree-based analytic technique that facilitates the modeling of various functional forms between multiple independent variables with a given outcome of interest, can overcome some of the limitations of current approaches for identifying risk and protective factors. RP provides simple and easy-to-visualize decision rules for predicting categorical (i.e., classification trees) or continuous outcomes (i.e., regression trees). RP finds the split on a predictor variable that best distinguishes between low vs. high risk on an outcome. This partitioning approach is 'recursive' in that following each split, the same algorithm using all possible predictor variables, including variables that determined the previous split, determines the next split. This process continues until each 'terminal node' contains a relatively homogenous subsample. One advantage of RP over regression-based techniques is that it does not assume any particular functional form (e.g., linear, log-linear) for the associations between statistical predictors and outcome variables (Strobl et al., 2009). Another advantage is the determination of optimal cut-points on variables that predict the outcome of interest. Lastly, RP is ideal for high dimensional data, whether a product of a large number of independent variables or a large number of higher-order interactions among those variables.

Given that marijuana use rates peak during the traditional college years and college student marijuana users report experiencing negative marijuana-related consequences (Pearson et al., 2017b), we wanted to better classify risk/protective factors for marijuana use and marijuana-related problems into hierarchical structures that better differentiate those at risk for experiencing these problems from the broader population of users. Specifically, we developed intuitive decision trees that indicate unique contributions of variables that best account for lifetime marijuana use, past month marijuana use, and the experience of negative marijuana-related consequences among college students.

Method

Participants, Procedure, and Measures

Participants were 8,141 college students recruited from psychology department participant pools at 11 participating universities to complete an online survey. Our sample included 4339 lifetime users and 2129 past month users. Demographic information is reported elsewhere (Bravo et al., 2017b; Pearson et al., 2017b). The study was IRB-approved and conformed to World Medical Association Declaration of Helsinki provisions. Measures are summarized in Table 1.

Statistical Analyses

We tested 3 models using RP. In the full sample, Model 1 considered 76 variables as correlates of ever having used marijuana (lifetime marijuana status). Among lifetime marijuana users, Model 2 considered these variables as correlates of having used marijuana in the past month (past-month marijuana status). For these models, we excluded marijuana use indicators and other variables (e.g., marijuana use motives) that can only be assessed among users because they could result in artificial, perfect classification (e.g., individuals with scores on motives were all past-month marijuana users). For these binary outcomes, we calculated sensitivity/specificity and positive/negative predictive value. Among past-month marijuana users, Model 3 considered 119 variables as potential correlates of a total score on the Marijuana Consequences Questionnaire, including marijuana use indicators. Full RP models were trained on the data for each outcome, then to guard against overfitting, 10-fold cross-validation testing using the 1-minus standard error rule advised by Breiman et al. (1984) was implemented to create parsimonious pruned models. That is, we pruned each tree to the smallest number of splits and the smallest cross-validation error given that the cross-validation error plus its standard error is less than 1 (cf. Conner et al., 2010). Please note that 1 is the relative error of a model with no splits.

Results

Lifetime Marijuana Status

The cross-validated RP tree for Model 1 is shown in panel A of Figure 1. The pruned model explained 25% of the variability in this sample and, on average, 24% in the bootstrap cross-validation procedure (relative error=0.75, cross-validation error=0.76, SE=0.009, Sensitivity=.78, Specificity=.68, PPV=.74, NPV=.73). The strongest correlate was best friend injunctive norms. The next strongest correlate for both the high and low risk groups was mean score on the Perceived Importance of Marijuana to the College Experience Scale (PIMCES). Both high and low risk groups split once more on best friend injunctive norms, such that the chance of ever having used marijuana in the low risk group was 28% among those with mean scores between 0 and 2.17 and 57% among those with scores between 2.17 and 3.42. In the high-risk group, it was 61% among those with scores between 3.42 and 4.50 and 83% with scores above 4.50.

Past-Month Marijuana Status

The cross-validated RP tree for Model 2 is shown in panel B of Figure 1. The pruned model explained 25% of the variability in this sample and, on average, 23% in the bootstrap cross-validation procedure (relative error=0.75, cross-validation error=0.77, SE=0.01, Sensitivity=.71, Specificity=.76, PPV=.74, NPV=.73). The strongest correlate was identification with being a marijuana user, or marijuana identity. This split occurred at 1.1 on a 1–7 response scale with 1=strongly disagree, indicating past-month non-users tended to strongly disagree with all marijuana identity items whereas past-month users did not. The next strongest correlate of past-month use was best friend injunctive norms. Subsequent splits in the model were on descriptive norms (typical use, typical late-night use, heavy use).

Negative Consequences

The cross-validated RP tree for Model 3 is shown in panel C of Figure 1. The pruned model explained 20% of the variability in this sample and, on average, 15% in the bootstrap cross-validation procedure (relative error=0.79, cross-validation error=0.85, SE=0.04). The strongest correlate was total number of periods of marijuana use during a typical week, whereby participants who used in 5 or fewer time periods/week (n=1422) experienced, on average, 6.06 consequences, and participants who used in 6 or more time periods/week (n=706) experienced, on average, 12.2 consequences. Amongst the low risk group, the next most meaningful split was on protective behavioral strategies for marijuana use (PBSM) with those scoring greater than 4.24 (n=888) falling into the lowest of all risk categories (average=4.75 consequences), whereas those scoring less than 4.24 averaged 8.24 consequences. Among the high-risk group, the next most meaningful split was related to the positive urgency facet of the UPPS-P. Participants with positive urgency scores below 2.56 (n=592) averaged 11.1 consequences, and those above 2.56 (n=114) averaged 17.87 consequences. This highest risk group split one more time on the negative urgency facet of the UPPS-P, with those who reported negative urgency scores below 3.09 (n=80) averaging 14.9 consequences and those above 3.09 (n=34) averaging 24.85 consequences.

Supplemental Analyses

To test the consistency in findings, we conducted all analyses in three, roughly equal-sized subsamples. The full results for these analyses are available as supplemental materials (see Supplemental Figures A-I). Overall, our unpruned models demonstrated a remarkable consistency in the identification of important correlates even though the exact cutpoints varied somewhat. For example, marijuana identify was the first (see Figures D and F) or second split (see Figure E) in accounting for past-month marijuana user status across all models, and nearly all the remaining variables were injunctive or descriptive norms. However, we found considerable variability in the number of splits in a particular subsample. For example, in one of the subsamples examining consequences, the best-fitting model included only one split on typical frequency of marijuana use (see Figure G). In another one of the subsamples examining consequences, there were 10 splits (see Figure H).

Discussion

Using RP, we examined a wide range of potential correlates of marijuana-related outcomes and developed intuitive decision trees to uniquely identify salient correlates of lifetime marijuana user status, current (i.e., past month) marijuana user status, and the experience of negative marijuana-related consequences among a large sample of college students. Overall, we found a distinct set of correlates for each of these outcomes. Injunctive norms (risk) followed by internalization of college marijuana use culture, or internalized norms, (risk) were salient correlates of ever having used marijuana. In order, identification with being a marijuana user (risk), injunctive norms (risk), and descriptive norms (risk and protective) were salient correlates of using marijuana in the past month. Finally, in order, marijuana use frequency (risk), use of protective behavioral strategies (protective), positive urgency (risk), and negative urgency (risk) were uniquely associated with marijuana-related negative consequences.

We found that lifetime user status was best accounted for by two distinct types of norms: injunctive norms (i.e., perceived approval of marijuana use by other college students) and internalized norms (i.e., internalizations of the college marijuana use culture; Osberg et al., 2010; Pearson et al., 2017c). Not only do these findings support social norms theory (Perkins & Berkowitz, 1986) in that normative perceptions are important factors among many that contribute to the decision to use marijuana, but they suggest that norms may be the most prominent among such factors. Consistent with previous research on normative perceptions and marijuana use (Napper et al., 2015), both injunctive norms (i.e., best friend marijuana use approval) and descriptive norms (i.e., perception of others' marijuana use behaviors) were critical correlates of past-month marijuana user status. One important consideration is that the strength of this association may be overestimated if marijuana norms and use have a reciprocal relationship. Our cross-sectional design cannot determine the relative extent to which normative perceptions drive the decision to use marijuana (consistent with social norms models), or the degree to which use of marijuana changes one's normative perceptions (consistent with a cognitive dissonance perspective; Festinger, 1967).

With regards to marijuana identity, our findings corroborate a recent study using latent profile analysis among current marijuana users in this sample that found an increase in frequency of marijuana use and marijuana negative consequences across marijuana use classes to be monotonically associated with increased identification with being a marijuana user (Pearson et al., 2017a).

From a harm reduction perspective, the most promising targets for interventions are those factors that are most directly related to experiencing negative consequences from marijuana use. In examining correlates of marijuana-related negative consequences, our findings garner support for marijuana use frequency (Pearson et al., 2017b) and positive urgency as a risk factors for increased marijuana-related negative consequences (Zapolski et al., 2009); whereas protective behavioral strategies (PBS) were shown to be a protective factor associated with decreased negative consequences (Pedersen et al., 2016). In this same sample of current marijuana users, PBS use (i.e., strategies used before, during, or after marijuana consumption that reduce use, intoxication, and/or harm) has been shown to distinguish between marijuana user classes such that the more problematic user classes reported lower use of these strategies (Pearson et al., 2017a); PBS use has been shown to mediate a variety of risk/protective factors (including impulsivity-like traits) on marijuana outcomes (Bravo et al., 2017b); and PBS use has been shown to moderate the relationship between marijuana use frequency and marijuana-related consequences (Bravo et al., 2017a). In the face of such findings, it is imperative that these associations be examined experimentally and longitudinally to garner additional evidence that these are actually causal factors that predict marijuana-related consequences as opposed to epiphenomena that are simply associated with consequences. For example, replicating such findings longitudinally would suggest that marijuana PBS use is one of the most promising intervention targets for college student marijuana users.

Research has consistently shown that RP is a useful analytic technique in determining predictors of outcomes in both theory (e.g., Hellemann et al., 2009) and non-theory driven research (Blumenstein, 2005). One benefit of RP comes from identifying previously unconsidered variables and in identifying nonlinear relations among multiple variables that are associated with the outcomes of interest. For example, in the model examining lifetime marijuana use, the score on the new measure of Perceived Importance of Marijuana to the College Experience (Pearson et al., 2017c) was identified as the second-best correlate of lifetime marijuana user status in both the lower and higher risk groups that were first split on best friends injunctive norms. Thus, in situations in which there is a large number of putative correlates, RP is useful in determining which are most important, and in determining which can most efficiently account for the outcome of interest. Not only does RP allow researchers to identify correlates they may not have hypothesized as most useful in accounting for the outcome of interest, it also allows the examination of population heterogeneity by allowing the same variable to appear in a model multiple times. For example, in the model examining lifetime marijuana user status, we found best friend injunctive norms to be a correlate at two different nodes of the tree indicating four ranges of scores on best friend injunctive norms that resulted in low (28%), medium (58% and 61%), and high (83%) probability of ever having used marijuana. These complex relations in certain subgroups are sometimes more difficult to identify in traditional analyses based in the general linear model, and may highlight surprising patterns in the data worthy of future investigation.

Importantly, the decision trees created by RP analytic techniques have widespread potential utility in clinical practice, including creating risk profiles and identifying treatment effect modifiers that impact clinical decision making. Just as risk profiles are used in the medical field to identify those most at-risk for a disease and as screening tools (e.g., cardiovascular disease; Chiu et al., 2010), colleges interested in preventing those who may have tried marijuana from becoming those who use marijuana problematically might easily identify the subgroup at highest risk by means of a brief assessment of students marijuana injunctive and descriptive norms. Further, machine learning techniques (e.g., RP) allow clinicians to determine psychosocial factors that may impact treatment efficacy. As utilized in the physical therapy field, identifying specific treatment effect modifiers allows physical therapists to “derive treatment decision aids or prediction rules to help match a patient’s characteristics to the interventions available” (Hill & Fritz, 2011; p. 712). The present study serves as a basis for future research to determine a parsimonious set of factors that predict the experience of marijuana-related consequences. If additional research replicates the present study's primary findings, clinical determinations about the psychosocial factors to target during treatment could be made based on simple-to-interpret cut-off scores on a client's typical frequency of using marijuana per week, a simple PBS measure, and the UPPS-P. Although we provide this example utilizing results from the present study, longitudinal studies identifying specific psychosocial factors that impact the efficacy of treatment is needed to provide an accurate portrayal of treatment effect modifiers.

These results do not exist in a vacuum, and the application of these findings, and indeed the findings of any machine learning approach toward prediction of behavior, depend on replication in other samples. Cross-validation by means of splitting data into training and testing samples is a good first step when large datasets are available, but ideally multiple research groups will employ machine learning to build consensus towards clinically useful prediction tools, informing diagnosis, intervention and prevention as was envisioned 60 years ago by Meehl (1957). Given our non-experimental, cross-sectional design, we are unable to make causal inferences. Given that we did not assess cannabis use disorder in our sample, the clinical implications of our findings may be limited to harm reduction approaches targeting more typical marijuana users rather than those that have developed a cannabis use disorder. Future studies should examine whether classification tools like recursive partitioning can be used to significantly improve the selection of individuals needing a marijuana intervention and/or tailoring specific intervention components to specific individual characteristics. One limitation of RP is that there is the tendency for classification/regression trees to overfit the data, and that exact cut-points for specific splits may be influenced by small changes in the data (Strobl et al., 2009), which can be seen in the present sample by examining the subsample analyses (see Supplemental Figures). Although other machine learning approaches, like random forests, may be well-suited to address this limitation (Strobl et al., 2009), they lose the straightforward, decision rules that make RP so easily translated to being used in applied contexts.

Conclusion

The present study describes the development of intuitive models with factors that may help account for marijuana use and consequences using recursive partitioning. We found norms, use of protective behavioral strategies, identification with being a marijuana user, and impulsivity-like traits to be factors associated with marijuana user status and/or marijuana-related consequences. Findings support the use of exploratory data analytic techniques to better understand variables that may predict marijuana-related outcomes. Additional research is needed to support these factors as causal antecedents, rather than consequences or epiphenomena. We expect this large sample study will serve as an impetus to develop intervention strategies targeting these factors.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

Blumenstein BA (2005). A comment on the utility of recursive partitioning. Journal of Clinical Oncology, 23, 4254–4255. [PubMed: 15781879]

Bravo AJ, Anthenien AM, Prince MR, Pearson MR, & Marijuana Outcomes Study Team (2017a). Marijuana protective behavioral strategies as a moderator of the effects of risk/protective factors on marijuana-related outcomes. Addictive Behaviors, 69, 14–21. [PubMed: 28110137]

Bravo AJ, Prince MA, Pearson MR, & Marijuana Outcomes Study Team. (2017b). Can I use marijuana safely? An examination of distal antecedents, marijuana protective behavioral strategies, and marijuana outcomes. Journal of Studies on Alcohol and Drugs, 78, 203–212. [PubMed: 28317500]

Breiman L, Friedman J, Stone CJ, & Olshen RA (1984). Classification and regression trees. CRC press.

Buckner JD (2013). College cannabis use: The unique roles of social norms, motives, and expectancies. Journal of Studies on Alcohol and Drugs, 74(5), 720–726. [PubMed: 23948531]

Caldeira KM, Arria AM, O'Grady KE, Vincent KB, & Wish ED (2008). The occurrence of cannabis use disorders and other cannabis-related problems among first-year college students. Addictive Behaviors, 33, 397–411. [PubMed: 18031940]

Chiu M, Austin PC, Manuel DG, & Tu JV (2010). Comparison of cardiovascular risk profiles among ethnic groups using population health surveys between 1996 and 2007. Canadian Medical Association Journal, 182(8), E301–E310. [PubMed: 20403888]

Conner BT (2015). Sensation Seeking Personality Trait Questionnaire. Unpublished scale.

Conner BT, Hellemann GS, Ritchie TL, & Noble EP (2010). Genetic, personality, and environmental predictors of drug use in adolescents. Journal of Substance Abuse Treatment, 38(2), 178–190. [PubMed: 19717274]

Cyders MA, Smith GT, Spillane NS, Fischer S, Annus AM, & Peterson C (2007). Integration of impulsivity and positive mood to predict risky behavior: development and validation of a measure of positive urgency. Psychological Assessment, 19, 107–118. [PubMed: 17371126]

Dvorak RD (2016). Beliefs Regarding Marijuana Scale (BRMS). Unpublished scale.

Festinger L (1962). *A theory of cognitive dissonance* (Vol. 2). Stanford university press.

Galbraith T, & Conner BT (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. [PubMed: 25347013]

Gratz KL, & Roemer L (2004). Multidimensional assessment of emotion regulation and dysregulation: Development, factor structure, and initial validation of the difficulties in emotion regulation scale. Journal of Psychopathology and Behavioral Assessment, 26, 41–54.

Hasin DS, Saha TD, Kerridge BT, Goldstein RB, Chou SP, Zhang H, ... & Huang B (2015). Prevalence of marijuana use disorders in the United States between 2001–2002 and 2012–2013. JAMA Psychiatry, 72, 1235–1242. [PubMed: 26502112]

Hellemann G, Conner BT, Anglin MD, & Longshore D (2009). Seeing the trees despite the forest: Applying recursive partitioning to the evaluation of drug treatment retention. Journal of Substance Abuse Treatment, 36, 59–64. [PubMed: 18599252]

Hill JC, & Fritz JM (2011). Psychosocial influences on low back pain, disability, and response to treatment. Physical Therapy, 91(5), 712–721. [PubMed: 21451093]

James G, Witten D, Hastie T, & Tibshirani R (2013). An introduction to statistical learning (Vol. 6). New York: Springer.

Johnston LD, O'Malley PM, Bachman JG, Schulenberg JE & Miech RA (2016). Monitoring the Future national survey results on drug use, 1975–2015: Volume 2, College students and adults ages 19–55. Ann Arbor, MI: Institute for Social Research, The University of Michigan.

King MW, & Resick PA (2014). Data mining in psychological treatment research: A primer on classification and regression trees. Journal of Consulting and Clinical Psychology, 82(5), 895–905. [PubMed: 24588404]

Lindgren KP, Neighbors C, Teachman BA, Wiers RW, Westgate E, & Greenwald AG (2013). I drink therefore I am: validating alcohol-related implicit association tests. Psychology of Addictive Behaviors, 27, 1–13. [PubMed: 22428863]

Marshall RJ (2001). The use of classification and regression trees in clinical epidemiology. Journal of Clinical Epidemiology, 54(6), 603–609. [PubMed: 11377121]

Meehl PE (1957). When shall we use our heads instead of the formula? Journal of Counseling Psychology, 4(4), 268–273.

Michalski RS, Carbonell JG, & Mitchell TM (Eds.). (2013). Machine learning: An artificial intelligence approach. Springer Science & Business Media.

Napper LE, Hummer JF, Chithambo TP, & LaBrie JW (2015). Perceived parent and peer marijuana norms: the moderating effect of parental monitoring during college. Prevention Science, 16, 364–73. [PubMed: 24838776]

Osberg TM, Atkins L, Buchholz L, Shirshova V, Swiantek A, Whitley J, … Oquendo N (2010). Development and validation of the College Life Alcohol Salience Scale: A measure of beliefs about the role of alcohol in college life. Psychology of Addictive Behaviors, 24(1),1–12. [PubMed: 20307107]

Pearson MR, Bravo AJ, Conner BT, & Marijuana Outcomes Study Team. (2017a). Distinguishing subpopulations of marijuana users with latent profile analysis. Drug and Alcohol Dependence, 172, 1–8. [PubMed: 28081515]

Pearson MR, Kholodkov T, Henson JM, & Impett EA (2012). Pathways to early coital debut for adolescent girls: A recursive partitioning analysis. Journal of Sex Research, 49(1), 13–26. [PubMed: 21512947]

Pearson MR, Liese BS, Dvorak RD, & Marijuana Outcomes Study Team (2017b). College student marijuana involvement: Perceptions, use, and consequences across 11 college campuses. Addictive Behaviors, 66, 83–89. [PubMed: 27894029]

Pearson MR, Kholodkov T, Gray MJ, & Marijuana Outcomes Study Team. (2017c). Perceived Importance of Marijuana to the College Experience Scale (PIMCES): initial development and validation. Journal of Studies on Alcohol and Drugs, 78(2), 319–324. [PubMed: 28317514]

Pedersen ER, Hummer JF, Rinker DV, Traylor ZK, & Neighbors C (2016). Measuring protective behavioral strategies for marijuana use among young adults. Journal of Studies on Alcohol and Drugs, 77, 441–450. [PubMed: 27172576]

Perkins HW, & Berkowitz AD (1986). Perceiving the community norms of alcohol use among students: Some research implications for campus alcohol education programming. International Journal of the Addictions, 21, 961–976. [PubMed: 3793315]

Simons J, Correia CJ, Carey KB, & Borsari BE (1998). Validating a five-factor marijuana motives measure: Relations with use, problems, and alcohol motives. Journal of Counseling Psychology, 45, 265–273.

Simons JS, Dvorak RD, Merrill JE, & Read JP (2012). Dimensions and severity of marijuana consequences: development and validation of the Marijuana Consequences Questionnaire (MACQ). Addictive Behaviors, 37, 613–21. [PubMed: 22305645]

Shadel WG, & Mermelstein R (1996). Individual differences in self-concept among smokers attempting to quit: Validation and predictive utility of measures of the smoker self-concept and abstainer self-concept. Annals of Behavioral Medicine, 18, 151–156. [PubMed: 24203766]

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. [PubMed: 19968396]

Whiteside SP, & Lynam DR (2001). The five factor model and impulsivity: Using a structural model of personality to understand impulsivity. Personality and Individual Differences, 30, 669–689.

Woicik PA, Stewart SH, Pihl RO, & Conrod PJ (2009). The Substance Use Risk Profile Scale: A scale measuring traits linked to reinforcement-specific substance use profiles. Addictive Behaviors, 34, 1042–55. [PubMed: 19683400]

Yarkoni T, & Westfall J (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. Perspectives on Psychological Science, 12(6), 1100–1122. [PubMed: 28841086]

Zapolski TC, Cyders MA, & Smith GT (2009). Positive urgency predicts illegal drug use and risky sexual behavior. Psychology of Addictive Behaviors, 23, 348–354. [PubMed: 19586152]

Zhang H, & Singer B (2013). Recursive partitioning in the health sciences. New York: Springer Science & Business Media.

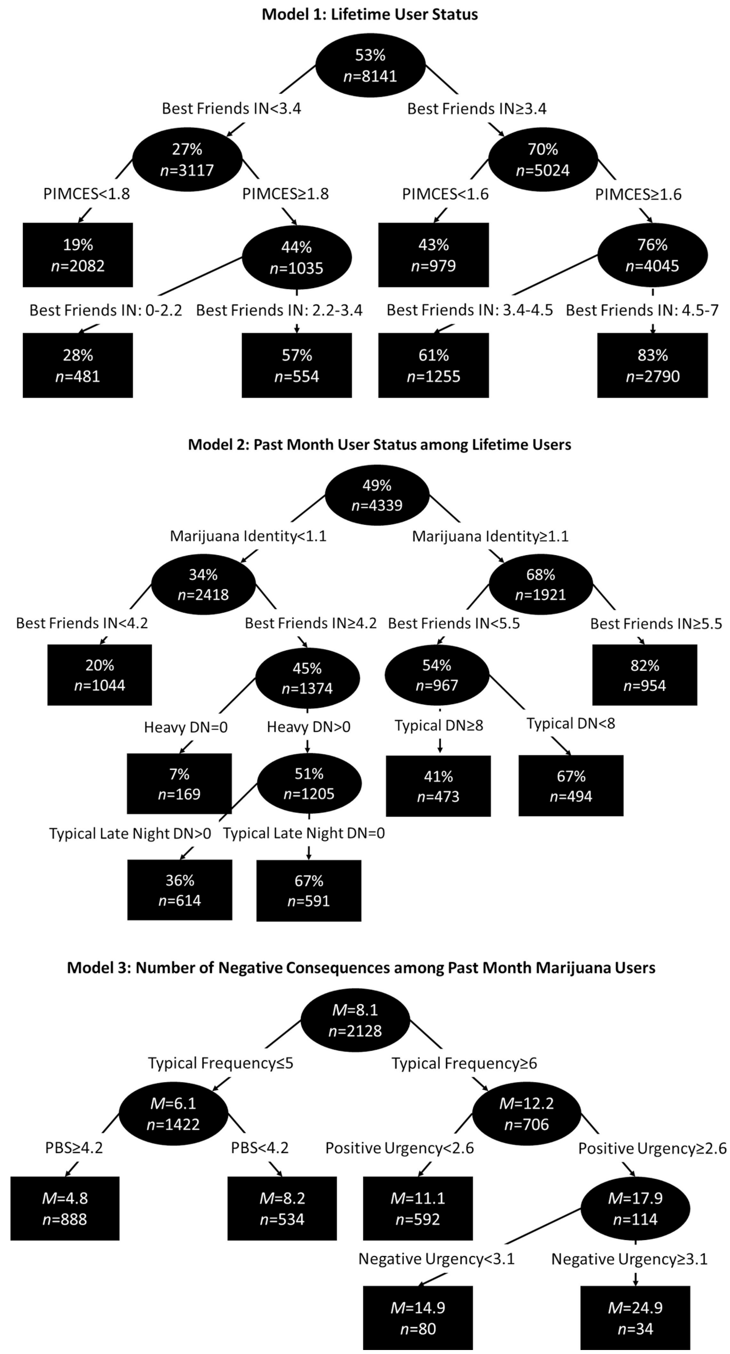

Figure 1.
Recursive partitioning models predicting lifetime marijuana user status (panel a), predicting past-month marijuana user status among lifetime users (panel b), and predicting number of negative consequences experienced among past-month marijuana users (panel c)

Table 1.

Summary of Measures

Measures Subscales	# of Variables	Items	Response Scale	"Example Items" or Subscale Summarization
Lifetime User Status (Model 1 Outcome)	1	1	0=no 1=yes	"In your lifetime, have you ever used marijuana in any form?"
Past Month Use/User Status (Model 2 Outcome)	2	1	0–30 (dichotomized to 0=none, 1=any when modeled as an outcome	"On how many days during the last 30 days did you use marijuana?"
Marijuana Consequences Questionnaire, MACQ; Simons et al., 2012) (Model 3 Outcome)	1	50	0=no 1=yes	"While using marijuana I have said or done embarrassing things." "I have not gone to work, or have missed classes or school because of using marijuana, being high, or after effects (feeling hung-over). "
Marijuana Use Indicators				
Frequency of Use in Past 30 days	1	1	0–30	"On how many days during the last 30 days did you use marijuana?"
Frequency of Getting High in Past 30 days	1	1	0–30	"On how many days during the last 30 days did you use marijuana to the point of being high?"
Time Periods Used in Past 30 Days-Typical	16	7	Checked time periods used	"During a week of typical marijuana use in the past 30 days, please indicate times and days that you used marijuana." Monday – Sunday 12a-4a 4a-8a 8a-12p 12p-4p 4p-8p 8p-12a Scales: Total number of time periods of marijuana use during the week (1 variable), number of time periods of use on each day of the week (7 variables), number of days of use during each time period (6), average number of time periods used per day (1 variable), number of days with any use per week (1 variable)
Time Periods Used in Past 30 Days-Heaviest	16	7	Checked time periods used	"During the week of your heaviest marijuana use in the past 30 days, please indicate times and days that you used marijuana." Same scales as above
UPPS-P (Cyders et al. 2007; Whiteside & Lynam, 2001)	5	(59)	1=disagree strongly 4=agree strongly	
Positive Urgency		14		"When I am very happy, I tend to do things that may cause problems in my life."
Negative Urgency		12		"In the heat of an argument, I will often say things that I later regret."
Premeditation		11		"My thinking is usually careful and purposeful."
Perseverance		10		"I finish what I start."
Sensation Seeking		12		"I generally seek new and exciting experiences and sensations."

Measures Subscales	# of Variables	Items	Response Scale	"Example Items" or Subscale Summarization
Substance Use Risk Profile (Woicik et al., 2009)	4	(23)	1=strongly disagree 4=strongly agree	
Anxiety Sensitivity		5		"It's frightening to feel dizzy or faint"
Hopelessness		7		"I am content"
Sensation Seeking		6		"I would like to skydive"
Impulsivity		5		"I often don't think things through before I speak"
Sensation Seeking Personality Type Quest (Conner, 2015)	2	(14)	1=strongly disagree 5=strongly agree	
Experience seeking		5		"I think it is important to try as many new things as I can."
Risk seeking		9		"I like to do things that other people think are dangerous."
Difficulty With Emotion Regulation Scale (Gratz & Roemer, 2004)	6	(36)	1=almost never (0– 10%) 5=almost always (91– 100%)	
Non-Acceptance of Emotional Responses		6		"When I'm upset, I feel guilty for feeling that way"
Difficulty engaging in goal-direct behavior		5		"When I'm upset, I have difficulty concentrating"
Impulsive control difficulties		6		"When I'm upset, I lose control over my behaviors"
Lack of emotional awareness		6		"I am attentive to my feelings" (revise code)
Limited access to emotion regulation strategies		8		"When I'm upset, I believe that I'll end up feeling very depressed"
Lack of Emotional Clarity		5		"I have no idea how I'm feeling"
Marijuana Protective Behavioral Strategies (Pedersen et al., 2016)	1	(39)	1=never 6=always	"Avoid use while spending time with family"
Marijuana Motives Measure (Simons et al., 1998)	5	(25)	1=almost never 5=almost always/always	
Enhancement		5		"To enjoy a party"
Coping		5		"To forget about my problems"
Social		5		"To be sociable"
Conformity		5		"To be liked"
Expansion		5		"To be more open to experiences"
Marijuana Injunctive Norms	6	9	1=strongly disapproving 7=strongly approving	"To what extent do you feel the following individuals approve of using marijuana? Your best friends Typical college students Your parents"
Marijuana Descriptive Norms	32	14	Six, 4-hour blocks of time (12a-4a, 4a-8a, 8a-12p, etc.),	Reported at which times they believed the typical college student used marijuana during a "typical week" and their "heaviest use week" in the past 30 days

Measures Subscales	# of Variables	Items	Response Scale	"Example Items" or Subscale Summarization
Marijuana Identity (Shadel & Mermelstein, 1996)	1	5	1=strongly disagree 7=strong agree	"Using marijuana is part of my self-image."
Identification with typical student (Shadel & Mermelstein, 1996)	1	5	1=strongly disagree 7=strong agree	"Being a college student is part of "who I am"."
Protective Behavioral Strategies for Marijuana (PBSM; Pedersen et al., 2016)	1	50	1=Never 6=Always	"Avoid using marijuana before work or school"
Perceptions of Marijuana Users/Non-Users (Dvorak, 2016)	1	10	1=completely disagree 5=completely agree	"Marijuana users are more fun to be around than nonusers."
Perceived Importance of Marijuana to the College Experience (Pearson et al., 2016)	1	13	1=strongly disagree 5=strong agree	"To get high on marijuana is a college rite of passage.."
Demographics	15	9		Age, gender, year in school, race, ethnicity, marital status, sexual orientation, political views, fraternity/sorority affiliation
Marijuana use indicators	37			See text