Published in final edited form as:

Psychol Addict Behav. 2014 December ; 28(4): 1127–1143. doi:10.1037/a0037235.

Applying an Ensemble Classification Tree Approach to the Prediction of Completion of a 12-Step Facilitation Intervention with Stimulant Abusers

Suzanne R. Doyle, Ph.D.^a and Dennis M. Donovan, Ph.D.^{a,b}

^aAlcohol and Drug Abuse Institute, University of Washington, Seattle, WA, USA

^bDepartment of Psychiatry and Behavioral Sciences, University of Washington School of Medicine, Seattle, WA, USA

Abstract

Aims—The purpose of this study was to explore the selection of predictor variables in the evaluation of drug treatment completion using an ensemble approach with classification trees. The basic methodology is reviewed and the subagging procedure of random subsampling is applied.

Methods—Among 234 individuals with stimulant use disorders randomized to a 12-Step facilitative intervention shown to increase stimulant use abstinence, 67.52% were classified as treatment completers. A total of 122 baseline variables were used to identify factors associated with completion.

Findings—The number of types of self-help activity involvement prior to treatment was the predominant predictor. Other effective predictors included better coping self-efficacy for substance use in high-risk situations, more days of prior meeting attendance, greater acceptance of the Disease model, higher confidence for not resuming use following discharge, lower ASI Drug and Alcohol composite scores, negative urine screens for cocaine or marijuana, and fewer employment problems.

Conclusions—The application of an ensemble subsampling regression tree method utilizes the fact that classification trees are unstable but, on average, produce an improved prediction of the completion of drug abuse treatment. The results support the notion there are early indicators of treatment completion that may allow for modification of approaches more tailored to fitting the needs of individuals and potentially provide more successful treatment engagement and improved outcomes.

Keywords

Classification trees; subagging; drug treatment completion

1. Introduction

Retention has long been identified as one of the major obstacles in the treatment of cocaine abuse and dependence, especially over the past two decades (Fernandez-Montalvo & López-Goñi, 2010; Lorenzoni et al., 2013; Polling, Kosten, & Sofuoglu, 2007; Santonja-Gomez et al., 2010; Stotts et al., 2007). “Ever since cocaine abusers began seeking treatment, researchers and clinicians have noted how difficult it is to retain them in treatment; so much so that retention has become more or less a proxy for treatment ‘success’ (Carrol, 1999, p. 505).” Attrition rates in clinical trials of treatment for cocaine dependence are high, often in the 75% range (Sayre et al., 2002). In a meta-analytic study of 34 controlled psychosocial treatment conditions and 2,340 adult participants across all studies for illicit substance use disorders, Dutra et al. (2008) found approximately 1/3 dropped out before treatment completion compared to 44.6% for control conditions, and that across all substance use groups (cannabis, cocaine, opiates, and polysubstances), cocaine patients tended to have higher mean dropout rates (42.0%). Similarly, in a report detailing trends in treatment admissions for over 378,500 methamphetamine or amphetamine abusers from 1992 to 2002 in California, Brecht, Greenwell, and Anglin (2005) found low treatment completion rates ranging from 27% to 33%.

Given that treatment dropout among stimulant abusers has been found in prior research to be associated with relapse and continued substance use (Aharonovich et al., 2006), identifying variables that best predict treatment completion for particular subgroups among stimulant abusers may aid clinicians in targeting dropout prevention strategies. To fully understand the basis of the complexities of treatment retention, it is important to recognize the critical factors involved, and to understand the complex relations between them.

Stotts et al. (2007) conducted a review of 23 published studies to identify the frequency of predictors studied and their relation to treatment retention among cocaine-dependent samples. They found inconsistently selected variables of convenience, widely varying statistical procedures and conflicting results of significance. A total of 201 predictor variables were evaluated, and it was reported that the majority of variables were arbitrarily selected post-hoc and devoid of theory, with notable variability found in the number of predictors utilized in the analyses (median = 7, range = 1 - 45). The most frequent categories of predictors across studies were in the following order: substance abuse, demographic, psychological, other, legal, family, and psychotherapy. Furthermore, diverse statistical techniques were utilized with most analyses being regression based (34.8% logistic, 21.7% survival analysis, and 8.7% linear), followed by ANOVA/ANCOVA (26.1%) and bivariate tests (8.7%), such as correlations or t-tests. Their review found a lack of unanimity across studies on the statistical significance of various predictor variables.

In addition to the qualitative literature review, Stotts et al. (2007) demonstrated an empirical quantitative analysis by testing a common set of 24 variables across three similar 12-week cocaine clinical trials conducted by the same investigators in the same research clinic within a five-year period. Predictors included demographics (6), drug use (11), psychosocial (4) and legal (3) variables. These analyses also resulted in discrepancies in variables identified as significant predictors of retention, providing further evidence of the instability of post-hoc

analyses conducted to identify retention predictors. Their overall conclusion was that the cumulative results of the qualitative and quantitative studies emphasizes that the inconsistencies call into question the meaningfulness or generalizability of the literature on cocaine treatment retention. "Thus, in the cocaine treatment field, failure to reliably replicate findings regarding predictors of retention indicates that little progress has been made in identifying key characteristics associated with treatment drop out, precluding the development of more efficient and effective strategies to enhance retention (Stotts et al., 2007, p. 2830)."

Their literature review revealed that researchers tend to analyze a small set of predictor variables because of their availability or because the general use of bivariate or standard regression techniques (e.g., 74% reported by Stotts et al., 2007) were not suitable for the exploration of complex relations among a large number of variables. It is frequently assumed that to analyze such high-dimensional data requires very large sample sizes and/or data reduction techniques, such as factor analysis to derive factor scores that can be utilized in subsequent statistical procedures. However, many researchers are unaware of the fact that binary recursive partitioning procedures, which are becoming increasingly popular in many scientific fields, can deal with the situation of small samples with large numbers of variables and their complex interactions. Unfortunately, these procedures have rarely been used in the general field of psychology and more specifically in substance abuse treatment research.

In providing an overview of binary recursive partitioning, also referred to as classification and regression trees (CART), Merkle and Shaffer (2011) point out there have been few applications of CART within the discipline of psychology and that the procedure is relatively unknown in the field. This is unfortunate since prediction, classification and the assessment of the relative importance of variables are fundamental to psychological research, and CART is a gateway method in the advancement of models and the discovery of results that may not be readily interpretable under existing theory and detectable by alternative statistical methods (Breiman, Friedman, Olshen, & Stone, 1984). It can be utilized to efficiently segment populations into meaningful groups and subgroups whose members share attributes that are important facilitators or barriers to the behavior of interest. In turn, this allows researchers to identify segments of a population and to predict the likelihood of engaging or not engaging in a particular behavior and to modify treatment approaches accordingly.

Strobl, Tutz, and Malley (2009) recognized that CART methods have gained popularity as an approach to multivariate data exploration in various scientific fields, including the analysis of microarray data, DNA sequencing, and many other applications in genetics, epidemiology, and medicine. Hellemann, Conner, Anglin, and Longshore (2009) also rightly called attention to the fact that the methodology is popular in genetics research and the examination of treatment outcomes in other medical fields, and that there is no reason why it could not be used to identify potential risk factors and suggest treatment guidelines in the area of drug use disorders research. They demonstrated the utility of using this procedure in the prediction of drug treatment retention with the inclusion of 195 potential predictors on a sample of 315 heroin-dependent individuals randomly assigned to two treatment conditions of narcotic replacement therapy. Their battery of intake assessment variables covered an

expansive range of psychosocial, demographic, and histographical measures. Severity of drug use, criminal behavior, level of insight, social network, and age were found to be predictive of treatment retention.

Although the results of their study adequately demonstrated the utility of the procedure for the detection of complex relations among diverse and interrelated variables, Helleman et al. (2009) constructed only one tree and the stability of the single derived tree may be in question. Instability occurs when even small changes in the data analyzed can result in the entire tree structure being altered, with concomitant changes in the type and number of predictors selected and in the resultant predicted values. The employment of cross-validation methods with the same data, using a resampling approach, would increase the generalizability and allow for investigation of the potential instability of tree structure. If the classification trees are unstable, the average over an ensemble of trees, rather than a single tree, will provide more reliable prediction and interpretability of predictor variable importance.

CART is a nonparametric and nonlinear computationally intensive statistical methodology formalized by Breiman et al. (1984). It is a relatively newer computer-assisted data analytic method based on machine learning and can be used to identify mutually exclusive and exhaustive subgroups of a population with members sharing characteristics associated with the outcome variable of interest. It has increased in popularity over the past several decades, principally due to the number of desirable properties and advantages it has over other classification and modeling methods such as discriminant analysis, logistic regression or ordinary least squares. In particular, CART (1) can be used to classify data involving either categorical, ordinal or continuous variables, both outcome and predictors; (2) can deal efficiently with large data sets, either in terms of the number of cases or number of variables, even including the analysis of data with more variables than observations; (3) requires no distributional assumptions for predictor variables, being effectively resistant to outliers, multicollinearity, and heteroskedasticity; (4) is invariant to monotonic transformations of variables; (5) missing data can be handled without complex, statistically-based multiple imputation techniques; (6) is a procedure free of probabilistic significance tests with the resultant tree-structure of predictors being easy to interpret; and, (7) it allows for numerous and/or complex interactions.

The term interaction in CART indicates the combination or cluster of characteristics or attributes that leads to the formulation of decision rules for classification, but does not reveal whether a variable moderates the effect of another. In previous studies of substance abuse treatment retention, little effort has been made to evaluate such predictor interactions, so there is little information about whether selected predictors interact with one another to produce multiplicative patterns of risk or identify subgroups of individuals who might be at higher risk for treatment dropout. Through identification of the interplay of factors or situations that precipitate non-completion, treatments might be designed or applied to mitigate dropout, and the ability to distinguish those at risk could additionally provide insight into mechanisms of continued substance use frequently associated with treatment dropout (e.g., see Brewer, Catalano, Haggerty, Gainey, & Bleming, 1998).

In particular, with a binary outcome CART is advantageous in the presence of many variable interactions that the researcher may not have predicted, with interactions too complex to be captured by parametric regression models or that demonstrate nonlinear effects.

Furthermore, it is an effective method when numerous predictor variables are also binary and the combination of many main and interaction effects may result in cell counts too sparse for a reliable solution with alternative statistical methods, such as logistic regression. In this article, we place an emphasis on classification trees with a binary outcome. The specific solution to deriving appropriate predictors depends on whether the outcome of interest is categorical (including binary) or quantitative; that is, whether a classification tree or a regression tree, respectively, is appropriate. More specific information regarding regression tree analysis with a continuous outcome can be found in related literature (cf., e.g., Breiman et al., 1984; Berk, 2006; Strobl et al., 2009; Zhang & Singer, 1998).

The purpose of the study was (1) to demonstrate a single application of the classification tree methodology in terms of the interactive nature of various predictor variables in profiling treatment completion, and (2) present an exploration of the selection of predictor variables in the evaluation of drug treatment completion using the ensemble subagging approach.

2. Methods

2.1. Procedure

Data was collected from a multisite randomized controlled trial evaluating the effectiveness of an 8-week combined group plus individual 12-Step facilitative intervention in leading to a significant greater reduction in stimulant drug use among individuals with stimulant use disorders. Stimulant use included cocaine, amphetamines, methamphetamines and other stimulants. Participant recruitment, assessment, and treatment took place at 10 community-based treatment programs (CTPs) affiliated with the NIDA Clinical Trials Network (CTN). Individuals with stimulant use disorders ($n = 471$) were randomly assigned to TAU (treatment-as-usual) or TAU into which the Stimulant Abuser Groups to Engage in 12-Step (STAGE-12) intervention was integrated. TAU consisted of the CTPs' standard outpatient psychosocial treatment (minimum of 5–15 hours weekly). The STAGE-12 sessions replaced 8 TAU sessions, and therefore were integrated into rather than added onto the standard TAU, to allow for the active treatment period of the study to be comparable across conditions within a CTP.

The STAGE-12 intervention of this study was designed to facilitate involvement and participation in 12-Step activities as well as attendance at 12-Step self-help meetings. It consisted of 5 group sessions, each lasting approximately 90 minutes, plus 3 individual sessions lasting approximately 45-75 minutes each. Group sessions focused on increasing acceptance of 12-Step principles; whereas, individual sessions incorporated an intensive referral procedure connecting participants to 12-Step volunteers. The group sessions included 5 topics: (1) acceptance; (2) people, places, and things; (3) the process of surrender; (4) getting active in 12-step programs; and, (5) managing emotions. Individual sessions augmented the group sessions, with the first emphasizing active participation in 12-Step activities. The second individual session varied in content, depending on whether the participant attended the first session, and if so, was concerned with reactions to the meeting,

and if not, focused on perceived and actual barriers to attendance. The third individual session involved reviewing recovery assignments and experiences, finding a meeting sponsor or group volunteer, and discussions on future goals and plans. More detailed descriptions of the STAGE-12 intervention can be found elsewhere in published articles (Daley, Baker, Donovan, Hodgkins, & Perl, 2011; Donovan et al., 2013) and the intervention manual (Baker, Daley, Donovan & Floyd, 2009).

An a priori defined criterion of “completion” was having received a therapeutic dose of the 12-Step intervention (e.g., for completers versus non-completers) for participants in the STAGE-12 intervention only. The delineation of “completion” as active involvement or participation in the 12-Step therapy included attendance in 2 or more individual sessions plus 3 or more group sessions. Selection of this criterion was based on the experience of the investigators and treatment program representatives with similar interventions of this type and length. Prior studies of group or combined group plus individual 12-Step facilitative interventions had found participants to have completed between only 35-40% of scheduled treatment sessions (e.g., Maude-Griffin et al., 1998; Wells, Peterson, Gainey, Hawkins, & Catalano, 1994). The threshold used in the a priori definition employed in the parent study from which the current data are derived was substantially higher than these rates (e.g., 60.00% of group sessions, 66.67% of individual sessions, and 62.50% of total sessions).

Completion status was available only on participants of the STAGE-12 intervention and not obtainable for subjects in the TAU arm. Although some CTPs may have reflected a 12-Step treatment philosophy and supported involvement in 12-step activities, CTP counselors in the TAU arm did not provide participants the specific components of the STAGE-12 facilitative intervention, including project-prepared handouts, reading material, or journal writing encouragement concerning plans and reactions regarding 12-Step meeting attendance. Additionally, participants in TAU were not connected to a community-based 12-Step volunteer nor did they receive a list of potential sponsors. Therefore, these preclusions for the TAU group did not allow completion status in terms of exposure to the specific content of the 12-Step intervention to be assessed.

The study by Wells et al. (2014), with the same data used in this current study, revealed those with high exposure (identically defined as completers here) showed evidence of more self-reported abstinence from and lower rates of stimulant drug use when compared to those with low exposure (noncompleters). Comparable to the parent study (Donovan et al., 2013), the primary outcome in the Wells et al. (2014) study was the number of days of stimulant use within six, 30-day windows of assessment. These included the 30 days between randomization and mid-treatment, between mid-treatment and end of treatment, and 4 additional 30-day windows over the subsequent four month follow-up period. Based on a zero-inflated negative binomial random-effects regression model, their results indicated that for the 30-day window from randomization to mid-treatment, the odds of abstinence of stimulant use were 41.31 times greater for those with high STAGE-12 treatment exposure (completers) versus those with low exposure (non-completers) to the treatment. The odds ratios (OR) declined over time, but remained statistically significant ($p < .05$) at the 30-day windows between mid-treatment and end-of-treatment ($OR = 20.38$), from end-of-treatment to the first follow-up ($OR = 10.05$) and at the second follow-up ($OR = 4.96$). The rate of

stimulant use assessed with rate ratios (RR) for those with high exposure to STAGE-12 was less than half that of those with low exposure (RR = 0.42) during the 30-day window assessed between randomization and mid-treatment and about half (RR = .51) by end-of-treatment.

Regarding secondary outcomes in the Wells et al. (2014) study, those achieving high exposure (completers) to STAGE-12 compared with those with less exposure (non-completers), demonstrated lower probabilities of stimulant positive urines; higher odds of self-reported abstinence from and lower rates of non-stimulant drug use; lower odds of not attending and higher rates (days) of attending 12-Step self-help groups; greater likelihood of reporting no drug problems; a greater maximum number of days of self-reported duties at meetings; and more types of other 12-Step activities engaged in during 30-day assessment windows. More detailed information concerning the evaluation of additional outcomes between individuals with high (completers) versus low (non-completers) STAGE-12 treatment exposure can be found in Wells et al. (2014).

As a direct follow-up to the study by Wells et al. (2014), where treatment completion was shown to be related to better treatment outcomes, our current study focuses on the identification of the factors related to the prediction of completion and non-completion of the 12-Step facilitative intervention. Delineating these factors may provide a basis for targeting specialized services to vulnerable subgroups susceptible to non-completion and shorter treatment exposure. Additionally, restricting our attention to recursive partitions for a binary outcome simplifies matters for demonstrating the selection of predictor variables in the evaluation of drug treatment completion using an ensemble approach with classification trees.

2.2. Participants

Of the 234 subjects randomized to the STAGE-12 intervention arm, 91.45% attended 2 or more individual sessions, with 46.58% having attended all three; 67.53% attended 3 or more group sessions, with 40.17% having attended all five sessions. The number meeting the previously defined criterion for intervention completion, attending at least 2 individual plus 3 group sessions was 158 (67.52%). Among all STAGE-12 participants, 88 individuals (37.61%) attended all 8 intervention sessions.

Individuals randomized to STAGE-12 averaged 38.24 years of age (SD = 10.04) and 12.23 years of education (SD = 1.65). Sixty-two percent were female, 46.15% reported Caucasian race and 37.61% reported Black/African American race. Thirty-four percent were unemployed, and only 15.45% were currently married. A majority (72.65%) met DSM-IV criteria for current cocaine dependence, 33.76% for methamphetamine dependence, and 6.84% for amphetamine dependence.

2.3. Measures

A total of 122 baseline variables were used to identify factors associated with treatment completion (see Appendix A). Predictors were selected from a review of the literature, and, therefore, past research played a key role in discerning the important and relevant variables for this population. These included 9 demographic, 5 family/social, 8 medical, 8 psychiatric,

16 legal, 28 substance use, 29 alcohol or drug treatment-related variables, and 19 psychological measures. Of the 122 variables, 86 are binary, 28 are quantitative, 5 are count data, and 3 are ordinal, based on a Likert scale. Variables were obtained from the following 15 instruments.

2.3.1. Demographic Form (DEM)—Information from this form included gender, ethnicity, and age.

2.3.2. The Addiction Severity Index (ASI)—Thirty-eight variables were included from the ASI which is a standardized, multidimensional, semi-structured interview developed by McLellan, Kushner, et al. (1992). It provides clinical information on the severity of potential treatment problems in 7 areas commonly affected by alcohol and drug dependence. These areas are reflected in 7 ASI composite scores of Medical, Employment, Alcohol, Drug, Legal, Family/Social, and Psychiatric problems. Thirty-one additional variables that did not comprise any composite score were also included.

2.3.3. Alcohol Use Disorder Identification Test, 3-Item Version (AUDIT-C)—The AUDIT-C is made up of three questions about alcohol consumption (Bradley et al., 2007; Bush, Kivlahan, McDonell, Fihn, & Bradley, 1998) derived from the longer 10-item AUDIT (Saunders, Aasland, Babor, de la Fuente, & Grant, 1993), which assesses three main areas: alcohol consumption/hazardous drinking, alcohol-related problems/harmful drinking, and alcohol dependence symptoms. The AUDIT-C, with potential total scores ranging from 0 to 12, is a practical, valid screening test for heavy drinking and/or active alcohol abuse or dependence.

2.3.4. DSM-IV Checklist—Substance use disorder diagnoses were derived from the psychoactive substance abuse disorder sections of the DSM-IV Criteria Checklist (Hudziak et al., 1993). This included 16 binary variables on whether or not the subject met DSM-IV diagnostic criteria for current (within the last 6 months) abuse or dependence on 8 substances: cocaine, methamphetamine, amphetamine, other stimulant drugs, alcohol, marijuana, opiates, and benzodiazepines.

2.3.5. Drug Taking Confidence Questionnaire, Short Version (DTCQ-8)—This 8-item measure (Sklar & Turner, 1999) provides a global index of coping self-efficacy across 8 situational areas associated with relapse risk. Individuals rate their degree of confidence that they could resist the urge to use their primary or secondary problem drug across these 8 high-risk relapse situations. One DTCQ-8 global score is obtained, potentially ranging from 0 to 100, with high levels of efficacy being associated with greater confidence that one is able to refrain from using, less difficulty in quitting use, greater motivation for and commitment to quit using.

2.3.6. Substance Use Calendar (SUC)—This measure provides self-report of substance use and is similar to the Timeline Follow-back procedure (Fals-Stewart, O'Farrell, Freitas, McFarlin, & Rutigliano, 2000; Sobell & Sobell, 1992). For the current assessment, one binary variable was included, indicating whether or not the subject reported any

stimulant use (cocaine, methamphetamine, amphetamine, or other stimulant drugs) within the previous 30 days prior to randomization of treatment.

2.3.7. Self-Efficacy Questionnaire – Patient Version (SEQ-P)—This instrument developed by Goldbeck, Myatt, and Aitchison (1997) consists of 5 items in which the individual rates his or her degree of confidence in remaining abstinent over the next 3 months, the expected difficulty in being able to remain abstinent, the amount of help that might be needed to do so, the strength of one's desire to do so, and how confident they are that following discharge they could use drugs once or twice without starting to use heavily again (relapse risk). Each item is assessed on a 7-point Likert scale. Two measures were obtained from this instrument: a composite efficacy score of the first 4 items (possible range of 4 to 28) and a single drug-taking confidence concerning relapse risk item. Higher scores indicate lower confidence in staying clean and sober.

2.3.8. Self-Help Activities Questionnaire (SHAQ)—This self-report instrument assesses the frequency of attendance at a number of self-help groups and the degree of participation in self-help activities in each of these groups (Weiss et al., 1996). Four variables were obtained from this measure: how many days in the previous 30 did they attend Alcoholic Anonymous (AA), Narcotics Anonymous (NA), Cocaine Anonymous (CA), or Crystal Meth Anonymous (CMA) meetings, did they speak at any of these meetings, did they have any duties at a meeting, and other self-help activities they were involved in. This last variable on activities has a potential range of 0 to 6 and includes the *number of different types of activities* they engaged in, such as meeting with members outside of a meeting, meeting with a sponsor, either phoning or receiving a call from a sponsor, receiving at least one phone call from other members, or reading AA/NA/CA/CMA literature for at least 5 minutes.

2.3.9. Spiritual Involvement and Beliefs Scale - Revised (SIBS-R)—A shortened version of the instrument developed by Hatch, Burg, Naberhaus, and Hellmich (1998) was used to evaluate spiritual actions as well as beliefs. Within the context of this measure, spirituality is defined in a broader context than religiosity, and would apply well to individuals who may actively utilize 12-Step principles without participation in any organized religion. The instrument provides 4 scores of Core Spirituality (connection, meaning, faith, involvement, and experience), a Spiritual/Existential perspective, Personal Application/Humility (or spiritual principles in daily activities), and Acceptance/Insight (of things as they are). In addition to the four subscales, a total SIBS-R score was derived.

2.3.10. Stages of Change Readiness and Treatment Eagerness Scale (SOCRATES)—The SOCRATES (Miller & Tonigan, 1996) is a 20-item self-report measure of clients' readiness to change their substance use behaviors. The current version has three subscales: Taking steps, Recognition of a problem, and Ambivalence. In addition to the three subscales a total score was used in the analysis.

2.3.11. Short Understanding of Substance Abuse Scale (SUSS)—The SUSS was developed by Humphreys, Greenbaum, Noke, and Finney (1996) to assess beliefs about the etiology and treatment of both alcohol and drug use. Three subscales are derived (Disease

model, Psychosocial model, and Eclectic orientation), with the Disease subscale having a potential range from 0 to 28, whereas the other two subscales can range from 0 to 20.

2.3.12. Survey of Readiness for Alcoholics Anonymous Participation

(SYRAAP)—This instrument developed by Kingree et al. (2006) assesses issues related to ambivalence and readiness to engage in 12-Step activities. It consists of three, 5-item subscales, each item rated on a 5-point Likert scale, that measure an individuals' perceptions about the Severity of their substance use problems, the Perceived Benefits of their involvement in 12-Step groups, and the Perceived Barriers to participating in 12-Step activities. A total SYRAAP score is also obtained.

2.3.13. Twelve-Step Experiences and Expectations (TSEE)—The TSEE, which was developed for the project, is a brief instrument assessing a participant's prior experiences with 12-Step groups. Individuals are asked to indicate whether they had been involved in any kind of self-help groups for alcohol or drug problems in the past and specifically if they had attended any AA, CA or NA 12-Step groups or other substance-related self-help groups. They are also asked how likely they were to get involved in a 12-Step group during this treatment episode and how helpful they anticipate such involvement would be.

2.3.14. Treatment Services Review (TSR)—This instrument, developed by McLellan, Alterman, Cacciola, Metzger, and O'Brien (1992), evaluates the type and extent of treatment services that a client reports receiving over a given time period (the previous 30 days) during alcohol and drug abuse rehabilitation. The TSR focuses on potential problems and services for seven areas: Medical, Employment and support, Drug use, Alcohol use, Legal, Family/Social, and Psychiatric.

2.3.15. Urine Drug Screen (UDS)—Urine was tested for the presence of six substances: cocaine, opiates, cannabis, benzodiazepines, methamphetamines, and amphetamines at baseline.

2.4. Analysis

The classification tree analysis started with the original data set ($n = 234$), generally referred to as the root or parent node, exploring all possible 122 predictor variables separately and selecting a hierarchy of variables that were, in succession, most related to the treatment completion outcome and maximized predictive accuracy. Beginning with the root node, CART first selected the best possible variable (primary splitter) to partition this parent node into two child nodes. All splits were binary; that is, with ordinal or continuous predictors, CART searched through the full range of values over all binary splits, finding the best cut-point according to the predetermined splitting criterion. In this study, we used the popular Gini Impurity Index splitting method (see Breiman et al., 1984; Hastie, Tibshirani, & Friedman, 2009). The Gini Index ranges from 0 (perfect discrimination with only one class in a node) to .5 (equal number of observations in two classes within a node), and is used throughout the recursive partitioning process to select the best split among the values of a predictor that results in the lowest impurity criterion. The splitting process was continued on

each of the new child nodes until our prespecified stopping rule was reached (i.e., $n = 10$ as the smallest number of observations allowable for a terminal node).

The influence of each predictor was assessed by the total variable importance index. With binary outcomes, the total variable importance index is the reduction in impurity multiplied by a measure of predictive agreement across nodes. With a single tree, the values of all importance measures are summed over each node, totaled then scaled relative to the best performing predictor, which is scored 1, with all other predictors having lower importance index scores descending toward zero. Larger importance values are associated with predictors positioned at or near the tree root or being involved in several different splits. Once the decision tree was constructed, the assignment of model-based predicted values was based on the simple plurality rule: assign the group having the largest number of observations within each terminal node. With predicted and observed group assignment available, the overall strength of a tree was indicated with additional indices of model fit, including misclassification rates, sensitivity, specificity, positive and negative rates, positive and negative predictive values, odds ratios and likelihood ratios. All of the above measures of classification accuracy and their 95% confidence intervals were calculated using the formulae given by Baldessarini, Finklestein, and Arana (1983), Fleiss, Levin, and Paik (2003), and Xhou, Obuchowski, and McClish (2002).

To enhance the predictive ability of the decision trees, the subbagging ensemble approach was utilized. With subbagging, a sobriquet for subsample aggregating (Bühlmann & Yu, 2002; Buja & Stuetzle, 2000), a subsample of smaller size than the original sample was randomly drawn *without* replacement and individual trees were constructed on each subsample. Given the relatively small size of the original sample ($n = 234$), with only 76 cases of non-completers (32.48%), it was decided to obtain 100 subsamples of 80% each ($n = 188$) in order to provide data sets large enough for the estimates of prediction to be sufficiently reliable.

Each random sample differed slightly from the original sample due to random variation, and the resulting trees differed in terms of the number and types of predictor variables selected. With ensembles, in addition to the importance index, a simple measure of variable importance was simply to count the number of times each variable was selected by all individual trees (Strobl et al., 2009), although more elaborate variable importance measures are available (Friedman, 2001; Strobl, Boulesteix, Zeileis, & Hothorn, 2007). Ranking variables by importance and discarding those of low significance provided a useful exploratory analysis technique in the selection of potentially valuable predictor variables.

Establishing predictions from the ensemble of 100 trees was achieved by taking each case and actually *dropping* it through every tree model in the ensemble to obtain a predicted class for that case. The class most trees voted for (i.e., the majority vote) was returned for that case as the prediction across the ensemble of decision trees, with the overall predictive strength indicated by the misclassification rate, sensitivity and specificity, and other related measures previously mentioned. Classification accuracy confidence intervals for the subbagging ensemble method utilized the standard deviation across the 100 trees as an estimate of the standard error.

In summary, the following algorithm was used in the subbagging ensemble method: (1) a random subsample ($n = 188$) without replacement was drawn from the original data ($n = 234$); (2) a classification tree was constructed considering all 122 predictors until terminal nodes had a minimum of 10 cases; (3) these first two steps were repeated 100 times; (4) for each randomly sampled data set, the importance index of each variable included in the tree model and indices of predictive accuracy were calculated; (5) the original data set ($n = 234$) was applied to, or dropped down, each of the 100 ensemble models and the number of times over trees that each case was classified as a completer or a non-completer was counted; and (6) each case was assigned to either a predicted treatment completion or non-completion category by a majority vote over the set of ensemble trees. Comparison of observed and the majority vote predicted completion/non-completion of treatment allowed for predictive accuracy over the ensemble.

Missing data on any of the 122 pre-randomization variables was replaced by the tree imputation method. This allowed for the replacement of missing values through estimation and the analysis of each potential predictor as an outcome with the remaining variables as its predictor in a similar tree-building regression or classification process. The SAS Enterprise Miner software (SAS, 2011) was employed to build all classification trees.

3. Results

3.1. Demonstration of CART with Total Sample of Completers

Figure 1 presents the single classification tree for the total sample. Each node contains the number of cases and the associate group (C = completion, N = non-completion) percentage. It consists of one primary splitter, 9 internal secondary splitters (circles) and 11 terminal nodes (squares). The primary splitter was found on the number of types of self-help activities prior to treatment, divided into two internal or child nodes of either < 2 or ≥ 2 activities. As previously indicated, these activities include meeting with members outside of a meeting, meeting with a sponsor, either phoning or receiving a call from a sponsor, receiving at least one phone call from other members, or reading AA/NA/CA/CMA literature for at least 5 minutes. Inspection of the distribution of this primary splitter revealed $n = 66$ reported no prior activity and of those who responded only one activity, $36/42 = 85.1\%$ indicated only baseline reading of AA/NA/CA/CMA literature at least for 5 minutes.

Of the participants indicating 2 or more types of self-help activities ($n = 126$), 77.8% are completers of the STAGE-12 intervention. Of these, those scoring in the upper 44% of the distribution of the SUSS Disease model ($n = 84$), a large percentage (85.7%) are completers. In other words, a two-way interaction indicates that among individuals involved in two or more types of self-help activities and who also score higher on acceptance of the Disease model at baseline, the predicted probability of treatment completion is 85.7%.

In comparison, a three-way interaction shows that participants indicating only 1 or no type of self-help activities ($n = 108$), who scored in the upper 55% of the distribution of the DTCQ-8 Global Confidence score concerning coping self-efficacy in refraining from use in high-risk situations ($n = 46$), and who also scored in the upper 80% of the distribution on acceptance of the Disease model, the predicted probability of completion is 82.9%. Those

indicating 1 or no activities, who scored in the lower 45% of the distribution of the DTCQ-8 ($n = 62$), and with an ASI Alcohol composite score in the upper 71% ($n = 18$) had only a 38.9% completion rate. More complex interactions can be interpreted similarly by considering the links of internal nodes until a terminal node is reached.

This single model suggests that completion of STAGE-12 treatment is associated with prior experience of involvement in a number of different types of self-help activities, greater acceptance of the Disease model, lower values on baseline ASI Alcohol and Drug composite scores, and greater confidence (DTCQ-8) they could resist the urge to use their problem drug in high-risk relapse situations. The reverse would be true for non-completers who in addition may have a greater confidence that following discharge they could use drugs once or twice without starting to use heavily again (SEQ-P drug taking confidence).

Table 1 presents descriptive indices and classification accuracy with this classification tree utilizing the total sample. The highest importance indices were associated with acceptance of the Disease model, the DTCQ-8 coping self-efficacy in refraining from use in high-risk situations, and involvement in a number of different types of self-help activities prior to treatment. Note, that although the number of different types of self-help activities was the primary splitter, acceptance of the Disease model had the highest importance index due to its interaction with the primary splitter and two different measures of drug-taking confidence. This is because interactions between predictor variables serve to increase the importance of those interacting variables, making them more likely to be given high importance relative to other variables. The misclassification rate was low, at 20.51% (95% CI = 15.34%, 25.69%). Additional indices included sensitivity = 0.866 (95% CI = 0.827, 0.927), specificity = 0.605 (95% CI = 0.493, 0.708), likelihood ratio positive = 2.245 (95% CI = 1.690, 2.982), likelihood ratio negative = 0.188 (95% CI = 0.117, 0.302), and odds ratio = 11.926 (95% CI = 6.087, 23.367). With an accuracy of approximately 80%, the model was slightly better at predicting completion than non-completion.

3.1. CART and Subagging: 100 Randomly Selected Subsamples ($n = 188$ each)

Of the 122 variables, 52 (46.62%) were never selected as a predictor of STAGE-12 treatment completion with the CART ensemble procedure, and 34 (27.87%) variables were minor splitters selected by the model with less than 5% of the 100 random subsamples (see Appendix A, the column designated as K = the number of times the variable was selected for a model).

The results of the subagging are presented in Table 2 for 36 predictor variables present in models in at least 5% of the random subsamples. It includes the number of times the variable was selected for a model (K), the number of times the variable was the primary splitter (P), and the direction of effect (D); where > and < indicate higher scores and lower scores, respectively, are associated with completion, and “n” indicates no and “y” indicates yes to binary items in relation to completion. The direction of effect was determined by reviewing the individual variables across each of the 100 trees. Table 2 also includes descriptive indices of the importance index and classification accuracy. Descriptive indices calculated are based on the number of times (K) the variable was selected, and dichotomous items are *italicized*.

Across the 100 subsamples, the number of variables in the decision tree models ranged from 5 to 11, with the following percentages: 5 (9%), 6 (25%), 7 (23%), 8 (22%), 9 (14%), 10 (6%), and 11 (1%). The average misclassification rate over all 100 trees was 0.209, with little variability across the ensemble (minimum = 0.165, maximum = 0.255, SD = 0.018). This reveals that selecting alternative predictors that produce slightly different splits can result in divergent trees that have similar predictive accuracy. The three most important predictors, occurring in over 50% of the random subsamples, are the number of types of self-help activities in the previous 30 days prior to randomization as measured by the SHAQ, the DTCQ-8 Global Confidence and the SUSS Disease model scores.

The number of types of self-help activities was the predominant predictor, present in 90% of the subsamples, and of these being the primary splitter in 39. As the initial splitter, 46% of the trees separated at 2 and 46% at 3 types of activities, with an increase of the number of types of activities linked to other variables that were associated with completion of treatment. The DTCQ-8 global score was the next most utilized predictor in the decision trees, and in the majority (76%) of the trees where it was the primary splitter, the division was at the median value of the score. That is, greater confidence that one could refrain from using and greater motivation for and commitment to quit using was related to treatment completion. Other effective predictors included better coping self-efficacy for substance use in high-risk situations, lower ASI Drug and Alcohol composite scores, concern about addiction and hurting others by their drug use (SOCRATES Ambivalence subscale score), negative pre-treatment urine screens for cocaine or marijuana, and fewer employment problems.

Applying the majority vote rule, all of the 234 cases were dropped through every tree model in the ensemble to obtain a single predicted class for each case for each tree, and the class most trees voted for was returned for that case as the prediction across the ensemble of 100 decision trees. The misclassification rate was low (0.111), with an extremely low false negative rate (0.013), and resulting high sensitivity (0.987) in the overall prediction of the completion of the STAGE-12 treatment (see Table 3). That is, the proportion of actual completers correctly identified as such with the majority vote was extremely high (98.73%) with only 2 false negative cases predicted.

4. Discussion

In this study, a single classification tree was used to demonstrate the interactive nature of various predictor variables to profile treatment completion among stimulant users in a 12-Step facilitation intervention. The subbagging approach was applied to enhance the predictive ability of utilizing decision trees and the relative influence of each predictor variable given apparent instability of tree structures across the ensemble, but unfortunately at the sacrifice of the more precise interpretability that a single tree allows.

As an exploratory procedure, the strength of CART is that it can effectively deal with a large number of variables, selecting those whose interactions are most important in predicting an outcome of interest. With the ensemble method, 70% of 122 potential predictor variables were either not included as predictors or were included as minor predictors in models in less

than 5% of the 100 random subsamples. Their absence or negligible influence in a model indicates they may not be strong discriminating characteristics for this sample, and their exclusion provides a method of data reduction in profiling treatment completion. Additionally, the range of the number of variables selected for a model across the ensemble of trees (5 to 11) and the variety of variables and primary splitters selected indicate the instability of tree structures with changes in the subsamples of data analyzed. Therefore, the average results of the ensemble approach would be more valid than the interpretation of a single classification tree.

In this exploratory study with a binary outcome we would be unable to use a more traditional p-value-based stepwise logistic regression to derive pertinent predictors because the number of model parameters, including both main effects and all possible two-way interactions, exceeds the sample size; that is, with 122 predictors used in a single model, there are $(122 \times 121)/2 = 7,381$ two-factor interaction terms. Furthermore, selection of a p-value as a stepwise stopping rule is problematic and the frequent selection of the default $p = 0.05$ has been criticized by many authors as inadequate for both prediction and interpretation (see Shtatland, Cain, & Barton (2001) and references therein). Unless there are a priori model specifications on specific interactions that may be theoretically relevant, assessing only a small number of predicted interactions limits the ability to explore the data completely, and the CART methodology, relative to logistic regression, would be better suited to an exploratory analysis. Without prior theoretical justification, a conventional logistic regression would have missed interactions that are in nature triple, quadruple and so on, as noted in our example of a single classification tree.

In the absence of theoretical models, it would be difficult to narrow down the large number of possible interactions with traditional regression approaches. Thus, researchers are often left with rather simplistic models with only main effects which are insufficient in explaining the complexities of behavior associated with treatment retention. For example, in a research study utilizing both classification tree analysis and logistic regression models to identify risk factors for relapse to smoking, it was noted that using a model with only main effects was not sufficient in that the CART models with the interactions included revealed the interaction terms to be the best predictors (Piper, Loh, Smith, Japuntich, & Baker, 2011). In a similar manner, Swan, Javitz, Jack, Curry, and McAfee (2004) used CART to explore gender differences in a 12-month cessation outcome from a bupropion smoking cessation trial. Their results revealed six subgroups for women and six different subgroups for men. They report these findings differed from the logistic regression results previously reported by the same research group using the same sample (Swan et al., 2003), and that this supported the premise there are predictors specifically relevant to subgroups of smokers and regression analysis may not be able to identify them in the absence of a priori predictions or interactions along with model building designed to test them.

Classification trees have several strengths compared to model-based methods like logistic regression. With a significant continuous predictor, logistic regression provides for its influence that is constant across all levels of that variable, but the tree-based method provides for identifying that level at which the predictor best splits the sample in relation to the outcome. CART analysis with a binary outcome can uncover nonlinear relations and

complex interactions, and differs from exploratory logistic regression models in that interactions between variables need not be pre-specified, for unknown interactions, if they exist, will be revealed within the tree structures and the variable importance scores will reflect their usefulness. Using an ensemble of trees increases the probability that some trees will capture the interactions among variables, especially in situations where predictor variables have no main effect on the outcome. Also, with a binary outcome, CART can adequately handle a dozen or more binary predictors, but logistic regression, with the combination of all main and interaction effects of interest, may often lead to sparse cell counts that preclude accurate parameter estimates. Furthermore, where CART works well with a large ratio of variables to observations, classical logistic regression is robust when the number of observations is distinctly larger, but with more variables than observations it breaks down unless shrinkage methods are used like ridge regression (Le Cessie & Van Houwelingen, 1992) or the lasso (Hastie et al., 2009, p. 661; Tibshirani, 1996). Previous research has also demonstrated that logistic regression tends to identify individuals homogeneous in outcome but heterogeneous in predictors, whereas CART identifies individuals homogeneous on both variables and that the latter method may be more useful than logistic regression for designing distinct tailored interventions for subgroups of high-risk individuals (Kiernan, Kraemer, Winkleby, King, & Taylor, 2001). More in-depth comparisons of the statistical properties of CART analysis and logistic regression can be found in published articles by Kiernan et al. (2001) and Lemon, Roy, Clark, Friedmann, and Rakowski, (2003), and the monograph by Michie, Spiegelhalter, and Taylor (1994).

The analyses conducted in this study are exploratory, which is necessary for theory development, and they do not represent any attempt to identify a causal model. In an exploratory stage of research, one of the best benefits of CART is that it can be used to efficiently assess a large number of predictors even in the face of limited sample sizes. Research does not need to be restricted to the examination of a small handful of predictors concerning drug abuse as has typically been conducted when analyses are based on more traditional statistical techniques (see Stotts et al., 2007). Ensemble techniques can be used to decrease the number of potential predictors. Identified important predictors can then be utilized to create a more final decision tree with an exploratory sample and subsequently cross-validated with an independent sample by evaluating the predictive value of the classification tree, and examining indices of classification accuracy. Deriving a final, more stable classification tree will allow the evaluation of the relative importance of potential interrelated risk factors and to predict outcomes for specific subgroups of individuals through a set of simple, easily understood yes or no questions that closely exemplify the decision-making process in real-life clinical settings.

The findings in this study provide a unique contribution to the delineation of factors and their relative importance in the prediction of the completion of a 12-Step facilitation intervention for individuals with stimulant use disorders. The utility of this method is that from 122 potential variables, 36 were selected that would be most useful for predicting treatment completion. Important predictors that occurred in at least 20% of the decision trees include prior experience and attendance in self-help activities and meetings; self-efficacy and confidence in abstinence and against relapse; greater alignment with the disease model philosophy; concern about addiction and hurting others by their drug use; less severity of

issues with alcohol and various other drugs; fewer employment problems; and, pre-treatment negative urine screens for cocaine and/or marijuana. The results of the CART ensemble revealed not only some of the salient predictors identified in previous studies concerning retention in stimulant-use treatment (see review by Stott et al., 2007), but also expanded the list of plausible predictors as preeminent from a set of 122 possible factors and all of their potential interactions, while simultaneously providing a descriptive measure of importance for each variable. This adds confidence in the significance of previously identified predictors while revealing less acknowledged relations that may have been overlooked.

While the focus of the present paper was primarily on the methodological aspects of the tree classification process, the results obtained concerning the predictors of 12-Step treatment completion are consistent with prior literature and are clinically meaningful. As an example, Kaskutas and colleagues (Kaskutas, Subbaraman, Witbrodt, & Zembre, 2009) found that the greatest effect of their 12-Step facilitative intervention was in the subgroup of alcohol and drug dependent individuals who had more prior experience with 12-Step meeting exposure. Kelly and Moos (2003) found individuals were less likely to drop out of 12-Step programs if they believed more strongly in the Disease model of addiction and had more prior involvement with 12-Step self-help groups. In samples of stimulant dependent individuals in treatment, urine screens positive for cocaine or marijuana were predictive of poorer stimulant use outcome, while less severe drug use patterns and current employment were related to treatment completion and positive treatment outcome (Maglione, Chao, & Anglin, 2000; Pierce et al., 2009).

The use of the CART classification tree procedure could be used to inform clinical work by deriving a flow chart for predicting potential drop-out, in this instance, from a 12-Step facilitative intervention or from other types of treatment. If the results presented in this paper are replicated or extended by future studies, this ultimately could lead to the development of an evidence-based attrition prevention intervention. Once a more definitive decision tree has been validated across samples and demonstrated relative stability, the predictors can be examined to identify mutually exclusive subgroups of at-risk individuals in similar 12-Step facilitative interventions on stimulant drug use. The segmentation into subgroups sharing characteristics that are important barriers or facilitators of a treatment outcome allows researchers to distinguish individuals that are most likely to either engage or not engage in particular behaviors and an adjunct treatment can be used to supplement the primary treatment or more intensive treatment can be provided that is tailored to their specific risk profiles in an effort to retain them in treatment rather than a one-size-fits-all approach. By distinguishing specific subgroups for which a specific treatment is essential, the method may be useful for creating detailed treatment algorithms.

Given the results of this study, in addition to considering mostly static variables like race, gender or marital status, researchers should attend to the selection of more dynamic variables, such as confidence and self-efficacy, that may provide better implications for the development of treatment interventions. For example, Moeller et al. (2001) found impulsivity to be a significant predictor of cocaine use and treatment retention and suggested targeting impulsivity in cocaine dependence treatment, such as contingency management therapy with short-term rewards for abstinence aimed at those with impulse control

disorders commonly exhibited by cocaine dependent individuals. In a similar manner, investigating the relation of cognitive functioning to commitment language and its shifts during a cognitive behavioral treatment session to treatment outcomes among cocaine-dependent patients, Aharonovich, Amrhein, Bisage, Nunes, and Hasin (2008) found commitment strength scores predicted treatment retention, suggesting engagement in the behavioral intervention process may depend on cognitive abilities. They recommended future clinical studies should further investigate the relations between cognition and verbal commitment to change and their differential contributions to treatment processes and outcome.

With a large enough sample, an ideal method for assessing the extent to which an ensemble of trees fits the characteristics of the dataset would have been to evaluate the results with a new independent sample or to split the current data into an exploratory and separate validation test sample. To evaluate forecasting adequacy, data from the validation sample would be dropped through the set of 100 trees already constructed by the ensemble method with the exploratory data. The predictive adequacy on the validation sample data across the derived ensemble of trees through a majority vote could then be assessed by considering indices of classification accuracy, as demonstrated in this study, with 95% confidence intervals used to statistically evaluate hypothesized values when applied to the validation sample data.

However, in the current study, predictive adequacy or forecasting was not attempted due the fact a new independent sample was not available. And, because of potential imbalance on the binary outcome variable (i.e., one class is represented by a large number of examples while the other is represented by only a few), a split-sample validation technique was not considered feasible. When the current data ($n = 234$) was randomly divided into equal-sized exploratory and validation samples ($n = 117$ each), with ensemble subsampling at 80%, some randomly drawn subsamples demonstrated non-completion as a rare event of few or almost none in this class resulting in a tree with poor performance for predicting the minority group of non-completers. Unfortunately, CART may not provide stable tree models of a rare outcome and typically attempts to minimize the overall error rate by predicting the non-rare event, especially with small samples, and requires special statistical approaches to tackle the problem of extremely imbalanced data (Chen, Liaw, & Breiman, 2004; Japkowicz & Stephens, 2002; Provost, 2000). It is important to note that this situation is problematic not only for CART procedures, but also for traditional logistic regression approaches when one is interested in correctly classifying rare or unique cases (King & Zeng, 2001).

It should be acknowledged that one potential problem with CART is a bias toward the selection of predictor variables having more values and therefore a greater number of split points (Kim & Loh, 2001; Shih, 2004; Strobl et al., 2007). This problem was recognized by Brieman et al. (1984), although a solution was not provided. However, Strobl et al. (2007) found that the random forest approach (see Brieman, 2001) on bootstrap samples with replacement artificially induced associations among variables and preferred those with more values, while there was no bias when subsampling without replacement was utilized. The difference between random samples with replacement, or bagging (Breiman, 1996), and subbagging (random subsamples without replacement, used in this study) and random forest

is that with the former two, variable selection is the same as that with a single classification tree, whereas with random forests an extra source of diversity is introduced when the variables are selected through random selection from the set of all predictor variables.

Finally, our analyses utilized only the subagging ensemble approach and to simplify matters, we restricted attention to recursive partitions for a binary outcome. While space prohibits a comprehensive overview, there are additional statistical ensemble methods extending the principles of CART, such as random forests, boosting (Hastie et al, 2009), Adaboost (Freund & Schapire, 1997), and Arcing classifiers (Breiman, 1998) that can also produce enhanced predictive accuracy (see also Berk, 2006, for a brief review of ensemble methods). Such ensemble methods often perform extremely well and have been shown to have desirable statistical properties (Bauer & Kohavi, 1999; Breiman, 1996, 2001; Bühlmann & Yu, 2002; Dietterich, 2000; Grandvalet, 2004; Mannor, Meir, & Zhang, 2003).

In this paper, we have briefly described the subagging ensemble CART method and illustrated its use in psychological research, specifically drug abuse treatment completion, as well as serving as a reminder to researchers to use caution when interpreting a single tree or regression model. This method places more emphasis on the exploration of accurate prediction and less attention on a well-defined model, with interest not in discerning specific main effects and interactions, but in the identification of important variables in the prediction of treatment completion. Since high-dimensional problems are common not only in genetics and medical research, but also in areas of psychology, we encourage researchers to recognize the utility of CART procedures and to continue the consideration of ways in which ensemble methods can improve the analysis of psychological data.

It is also important to point out the examples provided in this article are quite simple, and that, in fact, classification and regression trees can become very complex and difficult to interpret, especially with many terminal nodes with multifold predictor interactions, and if a priori stopping rules that place limits on the size of the tree have not been stipulated. Moreover, due to the ease of some CART methods, one must also avoid the pitfall of “data dredging” and such analyses should be conducted with a priori evaluation of the predictors assessed along with reasonable consideration of the direction of the association with the outcome (Marshall, 2001).

In conclusion, we have tried to highlight features of classification trees and a simple ensemble subagging approach as an introduction to the framework, terminology, and application of the technique. Our intent was not to present a comprehensive overview of the complex mathematical properties of the available recursive partitioning approaches, nor to provide a comparison of these approaches with more traditional regression-based methods and the advantages and limitations associated with each. For more in-depth examinations of these topics, interested readers should review references previously provided within the content of this manuscript.

Acknowledgments

Data collection and analysis were supported by a series of grants from NIDA as part of the Cooperative Agreement on National Drug Abuse Treatment CTN: Appalachian/Tri-States Node (U10DA20036), Florida Node Alliance

(U10DA13720), Ohio Valley Node (U10DA13732), Oregon Node (U10DA13036), Pacific Region Node (U10DA13045), Pacific Northwest Node (U10DA13714), Southern Consortium Node (U10DA13727), and Texas Node (U10DA20024). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of NIDA.

We would like to express our appreciation to the patients and the administrative, clinical, and research staff of the 10 CTPs that participated in the present study: Center for Psychiatric & Chemical Dependency Services, Pittsburgh, PA; ChangePoint, Inc. Portland, OR; Dorchester Alcohol & Drug Commission, Summerville, SC; Evergreen Manor, Everett, WA; Gateway Community Services, Jacksonville, FL; HinaMauka, Kaneohe, HI; Maryhaven, Columbus, OH; Nexus Recovery Center, Dallas, TX; Recovery Centers of King County, Seattle, WA; Willamette Family Treatment Services, Eugene, OR.

Appendix A

Instrument	Variable Description	K
Demographics		
DEM	<i>Gender (female)</i>	10
DEM	<i>Ethnicity: Caucasian vs. non-Caucasian</i>	3
DEM	<i>Age in years</i>	5
ASI	<i>Married vs. not married</i>	0
ASI	<i>In a controlled environment in past 30 days</i>	4
ASI	<i>Education: 12+ years vs. less 12</i>	0
ASI	<i>How long was the longest fulltime job in years</i>	31
ASI	<i>Experienced employment problems in past 30 days</i>	1
ASI	<i>Employment composite score</i>	18
Family/Social Relations		
ASI	<i>Do you live with anyone who has a current alcohol problem</i>	4
ASI	<i>Do you live with anyone who uses non-prescribed drugs</i>	0
ASI	<i>Spend most of free time with friends vs. family or being alone</i>	0
ASI	<i>Spend most of free time alone vs. family or friends</i>	8
ASI	<i>Family composite score</i>	0
Medical		
TSR	<i>Received any in-program/out-program services in past 30 days</i>	2
TSR	<i>Hospitalized for physical medical problem in past 30 days</i>	0
TSR	<i>Received medication for physical medical problem in past 30 days</i>	0
TSR	<i>Seen physician, assistant-physician or nurse for medical care in past 30 days</i>	1
TSR	<i>Had significant discussion about medical problem in an individual or group session in past 30 days</i>	0
ASI	<i>Chronic medical problem that continues to interferes with life</i>	3
ASI	<i>Taking any prescribed medication on a regular basis for a physical problem</i>	3
ASI	<i>Medical composite score</i>	14
Psychiatric		
TSR	<i>Been hospitalized for an emotional or psychological problem in past 30 days</i>	0
TSR	<i>Received medication for emotional or psychological problem in past 30 days</i>	8
TSR	<i>Received testing for an emotional or psychological problem in past 30 days</i>	0
TSR	<i>Had a significant discussion about your psychological or emotional problems with a psychologist or psychiatrist (individual or group) in past 30 days</i>	0
TSR	<i>Had a significant discussion about your psychological or emotional problems with a counselor or social worker (individual or group) in past 30 days</i>	0

Instrument	Variable Description	K
ASI	<i>Ever treated for psychological or emotional problems in an inpatient setting</i>	0
ASI	<i>Ever treated for psychological or emotional problems in an outpatient setting</i>	3
ASI	Psychiatric composite score	1
Legal		
ASI	<i>Ever been arrested and charged for shoplifting or vandalism</i>	0
ASI	<i>Ever been arrested and charged for parole or probation violations</i>	3
ASI	<i>Ever been arrested and charged for drug charges</i>	7
ASI	<i>Ever been arrested and charged for forgery</i>	1
ASI	<i>Ever been arrested and charged for weapons offense</i>	10
ASI	<i>Ever been arrested and charged for burglary, larceny, B&E</i>	2
ASI	<i>Ever been arrested and charged for robbery</i>	0
ASI	<i>Ever been arrested and charged for assault</i>	0
ASI	<i>Ever been arrested and charged for prostitution</i>	4
ASI	<i>Ever been arrested and charged for contempt of court</i>	9
ASI	<i>Ever been arrested and charged for other offenses</i>	6
ASI	<i>Ever charged for disorderly conduct/ vagrancy/public intoxication</i>	0
ASI	<i>Ever been charged for driving while intoxicated (DWI)</i>	2
ASI	<i>Ever been charged for major driving offense</i>	2
ASI	<i>In the past 30 days were you detained or incarcerated</i>	1
ASI	Legal composite score	12
Substance Use		
SUC	Number of days of stimulant use within the 30 days prior to randomization.	3
DSM	<i>Dependent Amphetamines</i>	0
DSM	<i>Dependent Methamphetamines</i>	0
DSM	<i>Dependent Cocaine</i>	2
DSM	<i>Dependent Other Stimulants</i>	0
DSM	<i>Dependent Alcohol</i>	0
DSM	<i>Dependent Marijuana/Hashish</i>	0
DSM	<i>Dependent Opiates</i>	0
DSM	<i>Dependent Benzodiazepines</i>	0
DSM	<i>Abuse Amphetamines</i>	0
DSM	<i>Abuse Methamphetamines</i>	0
DSM	<i>Abuse Cocaine</i>	0
DSM	<i>Abuse Other Stimulants</i>	0
DSM	<i>Abuse Alcohol</i>	5
DSM	<i>Abuse Marijuana/Hashish</i>	1
DSM	<i>Abuse Opiates</i>	7
DSM	<i>Abuse Benzodiazepines</i>	0
UDS	<i>Urine screen positive for cocaine</i>	22
UDS	<i>Urine screen positive for amphetamines</i>	0
UDS	<i>Urine screen positive for methamphetamine</i>	0
UDS	<i>Urine screen positive for opiates</i>	0

Instrument	Variable Description	K
UDS	<i>Urine screen positive for marijuana</i>	22
UDS	<i>Urine screen positive for benzodiazepines</i>	0
ASI	<i>Spent money on drugs in past 30 days</i>	13
ASI	<i>Days experienced drug problems in past 30 days</i>	0
ASI	AUDIT-C	2
ASI	Drug composite score	22
ASI	Alcohol composite score	25
Alcohol or Drug Treatment		
ASI	<i>Treated in outpatient setting for alcohol/drugs in the past 30 days</i>	6
TSR	<i>Inpatient treatment for alcohol problem in past 30 days</i>	0
TSR	<i>Received medication to help detox from alcohol problem in past 30 days</i>	0
TSR	<i>Received medication to prevent drinking in past 30 days</i>	0
TSR	<i>Received a blood alcohol test in past 30 days</i>	0
TSR	<i>Received medication to prevent drinking in past 30 days</i>	0
TSR	<i>Received a blood alcohol test in past 30 days</i>	2
TSR	<i>Attended alcohol education session in past 30 days</i>	0
TSR	<i>Attended alcohol relapse prevention meeting in past 30 days</i>	1
TSR	<i>Had significant discussion about alcohol problem in an individual or group session in past 30 days</i>	0
TSR	<i>Been in inpatient treatment for drug problem in past 30 days</i>	0
TSR	<i>Received medication to detox or come off drugs in past 30 days</i>	0
TSR	<i>Received medication to maintain or stabilize drug use in past 30 days</i>	0
TSR	<i>Received medication to block the effects of drugs in past 30 days</i>	0
TSR	<i>Received urinalysis or other test for drug use in past 30 days</i>	0
TSR	<i>Attended drug education session in past 30 days</i>	0
TSR	<i>Attended drug relapse prevention group or session in past 30 days</i>	3
TSR	<i>Had significant discussion about drug problem in an individual or group session in past 30 days</i>	2
SHAQ	How many days in the past 30 did you attend AA, NA, CA, or CMA meetings	24
SHAQ	<i>In the past 30 days did you ever speak at an AA, NA, CA, or CMA meeting</i>	2
SHAQ	<i>In the past 30 days did you have duties at an AA, NA, CA, or CMA meeting</i>	0
SHAQ	Number of types of self-help activities in the past 30 days (range 0 to 6)	90
TSEE	<i>Ever attended AA meetings</i>	0
TSEE	<i>Ever attended CA meetings</i>	3
TSEE	<i>Ever attended NA meetings</i>	0
TSEE	<i>Ever attended other self-help meetings</i>	0
TSEE	<i>Have you ever been involved in any kind of self-help groups for alcohol/drug problems in the past?</i>	1
TSEE	How likely are you to get involved in self-help group meetings during your current treatment?	3
TSEE	How helpful do you think involvement in a self-help group would be to you as part of your current treatment and recovery process?	3
Psychological Assessments		
DTCQ-8	Drug Taking Confidence Questionnaire: Global score of relapse risk	79

Instrument	Variable Description	K
SEQ-P	Composite Self-efficacy score	20
SEQ-P	At present, how confident are you that following discharge you could use drugs once or twice without starting to use heavily again?	21
SIBS-R	Core Spirituality subscale score	13
SIBS-R	Perspective Existential subscale score	6
SIBS-R	Application/Humility subscale score	1
SIBS-R	Acceptance/Insight subscale score	8
SIBS-R	Total Spiritual Involvement and Beliefs score	7
SOCRATES	Recognition subscale score	2
SOCRATES	Ambivalence subscale score	20
SOCRATES	Taking Steps subscale score	18
SOCRATES	Readiness to Change Total score	5
SUSS	Psychosocial model subscale score	0
SUSS	Disease model subscale score	51
SUSS	Eclectic model subscale score	2
SYRAAP	Perceived Benefit subscale score	12
SYRAAP	Perceived Severity subscale score	10
SYRAAP	Perceived Barriers subscale score	8
SYRAAP	Total score	3

K = number of times the variable was selected for a model Dichotomous items are *italicized*.

References

- Aharonovick R, Amrhein PC, Bisaga A, Nunes EV, Hasin DS. Cognition, commitment language, and behavioral change among cocaine-dependent patients. *Psychology of Addictive Behaviors*. 2008; 22:557–562. [PubMed: 19071981]
- Aharonovich E, Hasin DS, Brooks AC, Liu X, Bisaga A, Nunes EV. Cognitive deficits predict low treatment retention in cocaine dependent patients. *Drug and Alcohol Dependence*. 2006; 81:313–322. [PubMed: 16171953]
- Baker, S.; Daley, DC.; Donovan, DM.; Floyd, AS. STAGE-12: Stimulant Abuser Groups to Engage in 12-Step Programs: A Combined Group and Individual Treatment Program, v. 3.0. 2009. Retrieved from the Clinical Trials Network, Dissemination Library, website: <http://ctndisseminationlibrary.org/display/888.htm>
- Baldessarini RJ, Finklestein S, Arana GW. The predictive power of diagnostic tests and the effect of prevalence of illness. *Archives of General Psychiatry*. 1983; 40:569–573. [PubMed: 6838334]
- Bauer E, Kohavi R. An empirical comparison of voting classification algorithms: Bagging, boosting, and variants. *Machine Learning*. 1999; 36:105–139.
- Berk RA. An introduction to ensemble methods for data analysis. *Sociological Methods & Research*. 2006; 34:263–295.
- Bradley KA, DeBenedetti AF, Volk RJ, Williams EC, Frank D, Kivlahan DR. AUDIT-C as a brief screen for alcohol misuse in primary care. *Alcoholism: Clinical and Experimental Research*. 2007; 31:1208–1217.
- Brecht ML, Greenwell L, Anglin MD. Methamphetamine treatment: Trends and predictors of retention and completion in a large state treatment system (1992-2002). *Journal of Substance Abuse Treatment*. 2005; 29:295–306. [PubMed: 16311182]
- Breiman L. Random Forests. *Machine Learning*. 2001; 45:5–32.

- Breiman L. Bagging predictors. *Machine Learning*. 1996; 24:123–140.
- Breiman L. Arcing classifiers. *The Annals of Statistics*. 1998; 26:801–849.
- Breiman, L.; Friedman, JH.; Olshen, RA.; Stone, CJ. *Classification and Regression Trees*. Wadsworth; Monterey, CA: 1984.
- Brewer DD, Catalano RF, Haggerty K, Gainey RR, Bleming CB. A meta-analysis of predictors of continued drug use during and after treatment for opiate addiction. *Addiction*. 1998; 93:73–92. [PubMed: 9624713]
- Bühlmann P, Yu B. Analyzing bagging. *The Annals of Statistics*. 2002; 30:927–961.
- Buja A, Stuetzle W. Observations on bagging. *Statistica Sinica*. 2006; 16:323–352.
- Bush K, Kivlahan DR, McDonell MB, Fihn SD, Bradley KA. The AUDIT alcohol consumption questions (AUDIT-C): an effective brief screening test for problem drinking. Ambulatory Care Quality Improvement Project (ACQUIP). Alcohol Use Disorders Identification Test. *Archives of Internal Medicine*. 1998; 158:1789–1795. [PubMed: 9738608]
- Carroll KM. Old psychotherapies for cocaine dependence revisited. *Archives of General Psychiatry*. 1999; 56:505–506. [PubMed: 10359463]
- Chen, C.; Liaw, A.; Breiman, L. Using random forest to learn imbalanced data. 2004. (Technical Report No. 666). Retrieved from Department of Statistics, University of California, Berkeley, website: <http://www.stat.berkeley.edu/tech-reports/666.pdf>
- Daley DC, Baker S, Donovan DM, Hodgkins CC, Perl H. A combined group and individual 12-step facilitative intervention targeting stimulant abuse in the NIDA Clinical Trials Network: STAGE-12. *Journal of Groups in Addiction and Recovery*. 2011; 6:228–244. [PubMed: 22859917]
- Dietterich TG. An experimental comparison of three methods of constructing ensembles of decision trees: Bagging, boosting, and randomization. *Machine Learning*. 2000; 40:139–157.
- Donovan DM, Daley DC, Brigham GS, Hodgkins CC, Perl HI, Garrett SB, Zammarelli L. Stimulant abuser groups to engage in 12-Step: A multisite trial in the National Institute on Drug Abuse Clinical Trials Network. *Journal of Substance Abuse Treatment*. 2013; 44:103–114. [PubMed: 22657748]
- Dutra L, Stathopoulou G, Basden SL, Leyro TM, Powers MB, Otto MW. A meta-analytic review of psychosocial interventions for substance use disorders. *American Journal of Psychiatry*. 2008; 165:179–187. [PubMed: 18198270]
- Fals-Stewart W, O'Farrell TJ, Freitas TT, McFarlin SK, Rutigliano P. The timeline followback reports of psychoactive substance use by drug-abusing patients: Psychometric properties. *Journal of Consulting and Clinical Psychology*. 2000; 68:134–144. [PubMed: 10710848]
- Fernandez-Montalvo J, López-Goñi JJ. Comparison of completers and dropouts in psychological treatment for cocaine addiction. *Addiction Research & Theory*. 2010; 18:433–441.
- Fleiss, JL.; Levin, B.; Paik, MC. *Statistical Methods for Rates and Proportions*. 3rd ed. John Wiley & Sons, Inc.; Hoboken, NJ: 2003.
- Freund Y, Schapire RE. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*. 1997; 55:119–139.
- Friedman J. Greedy function approximation: A gradient boosting machine. *Annals of Statistics*. 2001; 29:1189–1232.
- Goldbeck R, Myatt P, Aitchison T. End-of-treatment self-efficacy: A predictor of abstinence. *Addiction*. 1997; 92:313–324. [PubMed: 9219393]
- Grandvalet Y. Bagging equalizes influence. *Machine Learning*. 2004; 55:251–70.
- Hastie, T.; Tibshirani, R.; Friedman, JH. *The Elements of Statistical Learning*. 2nd ed. Springer; New York: 2009.
- Hatch RL, Burg MA, Naberhaus DS, Hellmich LK. The Spiritual Involvement and Beliefs Scale. Development and testing of a new instrument. *Journal of Family Practice*. 1998; 46:476–486. [PubMed: 9638112]
- Helleman G, Conner BT, Anglin MD, Longshore D. Seeing the trees despite the forest: Applying recursive partitioning to the evaluation of drug treatment retention. *Journal of Substance Abuse Treatment*. 2009; 36:59–64. [PubMed: 18599252]

- Hudziak JJ, Helzer JE, Wetzel WW, Kessel KB, McGee B, Janca A, Przybeck T. The use of the DSM-III-R Checklist for initial diagnostic assessment. *Comprehensive Psychiatry*. 1993; 34:375–383. [PubMed: 8131381]
- Humphreys K, Greenbaum MA, Noke JM, Finney JW. Reliability, validity, and normative data for a short version of the Understanding of Alcoholism Scale. *Psychology of Addictive Behaviors*. 1996; 10:38–44.
- Japkowicz N, Stephen S. The class imbalance problem: A systematic study. *Intelligent Data Analysis*. 2002; 6:429–449.
- Kaskutas LA, Subbaraman MS, Witbrodt J, Zemore SE. Effectiveness of making Alcoholics Anonymous easier: A group format 12-step facilitation approach. *Journal of Substance Abuse Treatment*. 2009; 37:228–239. [PubMed: 19339148]
- Kelly JF, Moos R. Dropout from 12-step self-help groups: Prevalence, predictors, and counteracting treatment influences. *Journal of Substance Abuse Treatment*. 2003; 24:241–250. [PubMed: 12810145]
- Kiernan M, Kraemer HC, Winkleby MA, King AC, Taylor CB. Do logistic regression and signal detection identify different subgroups at risk? Implications for the design of tailored interventions. *Psychological Methods*. 2001; 6:35–48. [PubMed: 11285811]
- Kim H, Loh W. Classification trees with unbiased multiway splits. *Journal of the American Statistical Association*. 2001; 96:589–604.
- King, G.; Zeng, L. Logistic regression in rare events data. 2001. Retrieved from: <http://GKing.Harvard.edu>
- Kingree JB, Simpson A, Thompson M, McCrady B, Tonigan JS, Lautenschlager G. The development and initial evaluation of the Survey of Readiness for Alcoholics Anonymous Participation. *Psychology of Addictive Behaviors*. 2006; 20:453–462. [PubMed: 17176180]
- Le Cessie S, Van Houwelingen JC. Ridge estimators in logistic regression. *Journal of the Royal Statistical Society, Series C*. 1992; 41:191–201.
- Lemon SC, Roy J, Clark MA, Friedmann PD, Rakowski W. Classification and regression tree analysis in public health: Methodological review and comparison with logistic regression. *Annals of Behavior Medicine*. 2003; 26:172–181.
- Lorenzoni V, Curzio O, Karakachoff N, Saponaro A, Sanza M, Mariani F, Molinaro S. The effects of the macro-environment on treatment retention for problem cocaine users. *International Journal of Drug Policy*. 2013; 24:52–59. [PubMed: 22884542]
- Maglione M, Chao B, Anglin MD. Correlates of outpatient drug treatment drop-out among methamphetamine users. *Journal of Psychoactive Drugs*. 2000; 32:221–228. [PubMed: 10908011]
- Mannor S, Meir R, Zhang T. Greedy algorithms for classification - consistency, convergence rates, and adaptivity. *Journal of Machine Learning Research*. 2003; 4:713–742.
- Marshall RJ. The use of classification and regression trees in clinical epidemiology. *Journal of Clinical Epidemiology*. 2001; 54:603–609. [PubMed: 11377121]
- Maude-Griffin PM, Hohenstein JM, Humfleet GL, Reilly PM, Tusel DJ, Hall SM. Superior efficacy of cognitive-behavioral therapy for urban crack cocaine abusers: Main and matching effects. *Journal of Consulting & Clinical Psychology*. 1998; 66:832–837. [PubMed: 9803702]
- McLellan AT, Alterman AI, Cacciola J, Metzger D, O'Brien CP. A new measure of substance abuse treatment: Initial studies of the Treatment Services Review. *Journal of Nervous & Mental Disease*. 1992; 180:101–110. [PubMed: 1737971]
- McLellan AT, Kushner H, Metzger D, Peters R, Smith I, Grissom G, Argeriou M. The fifth edition of the Addiction Severity Index. *Journal of Substance Abuse Treatment*. 1992; 9:199–213. [PubMed: 1334156]
- Merkle EC, Shaffer VA. Binary recursive partitioning: Background, methods, and application to psychology. *British Journal of Mathematical & Statistical Psychology*. 2011; 64:161–181. [PubMed: 21506949]
- Michie, D.; Spiegelhalter, DJ.; Taylor, CC. *Machine Learning, Neural and Statistical Classification*. Ellis Horwood; New York: 1994.

- Miller WR, Tonigan JS. Assessing drinkers' motivations for change: The Stages of Change Readiness and Treatment Eagerness Scale (SOCRATES). *Psychology of Addictive Behaviors*. 1996; 10:81–89.
- Moeller FG, Dougherty DM, Barratt ES, Schmitz JM, Swann AC, Grabowski J. The impact of impulsivity on cocaine use and retention in treatment. *Journal of Substance Abuse Treatment*. 2001; 21:193–198. [PubMed: 11777668]
- Peirce JM, Petry NM, Roll JM, Kolodner K, Krasnansky J, Stabile PQ, Stitzer ML. Correlates of stimulant treatment outcome across treatment modalities. *American Journal of Drug & Alcohol Abuse*. 2009; 35:48–53. [PubMed: 19152207]
- Piper ME, Loh W-Y, Smith SS, Japuntich SJ, Baker TB. Using decision tree analysis to identify risk factors for relapse to smoking. *Substance Use & Misuse*. 2011; 46:492–510. [PubMed: 20397871]
- Polling J, Kosten TR, Sofuoglu M. Treatment outcome predictors for cocaine dependence. *The American Journal of Drug & Alcohol Abuse*. 2007; 33:191–206. [PubMed: 17497542]
- Provost, F. Machine learning from imbalanced data sets 101. *Proceedings of the AAAI-2000 Workshop on Imbalanced Data Sets*. 2000. Retrieved from New York University Faculty Digital Archive, website: <https://archive.nyu.edu/handle/2451/27763>
- Santonja-Gomez F-J, Sanchez-Hervas E, Secades-Villa R, Zacaes-Romaguera F, Garcia-Rodriguez O, Garcia-Fernandez G. Pretreatment characteristics as predictors of retention in cocaine-dependent outpatients. *Addictive Disorders & Their Treatment*. 2010; 9:93–98.
- SAS Enterprise Miner Release 7.1. Software system for Windows. SAS Institute, Inc.; Cary, NC: 2011.
- Saunders JB, Aasland OG, Babor TF, de la Fuente JR, Grant M. Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO collaborative project on early detection of persons with harmful alcohol consumption: II. *Addiction*. 1993; 88:791–804. [PubMed: 8329970]
- Sayre SL, Schmitz JM, Stotts AL, Averill PM, Rhoades HM, Grabowski JJ. Determining predictors of attrition in an out-patient substance abuse program. *American Journal of Drug and Alcohol Abuse*. 2002; 28:55–72. [PubMed: 11853135]
- Shih Y-S. A note on split selection bias in classification trees. *Computational Statistics & Data Analysis*. 2004; 45:457–466.
- Shtatland, ES.; Cain, E.; Barton, MB. SUGI'26 Proceedings, Paper 222-26. SAS Institute Inc.; Cary, NC: 2001. The perils of stepwise logistic regression and how to escape them using information criteria and the Output Delivery System.
- Sklar SM, Turner NE. A brief measure for the assessment of coping self-efficacy among alcohol and other drug users. *Addiction*. 1999; 94:723–729. [PubMed: 10563037]
- Sobell, LC.; Sobell, MB. Timeline followback: A technique for assessing self-reported alcohol consumption. In: Litten, RZ.; Allen, JP., editors. *Measuring alcohol consumption: Psychosocial and biological methods*. Humana Press; Totowa, NJ: 1992. p. 41-72.
- Stotts AL, Mooney ME, Sayre SL, Novy M, Schmitz JM, Grabowski J. Illusory predictors: Generalizability of findings in cocaine treatment retention research. *Addictive Behaviors*. 2007; 32:2819–2836. [PubMed: 17543472]
- Strobl C, Tutz G, Malley J. An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological Methods*. 2009; 14:323–348. [PubMed: 19968396]
- Strobl C, Boulesteix A-L, Zeileis A, Hothorn T. Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics*. 2007; 8:25. [PubMed: 17254353]
- Swan GE, Jack LM, Curry S, Chorost M, Javitz H, McAfee T, Dacey S. Bupropion SR and counseling for smoking cessation in actual practice: Predictors of outcome. *Nicotine & Tobacco Research*. 2003; 5:911–921. [PubMed: 14668075]
- Swan GE, Javitz HS, Jack LM, Curry SJ, McAfee T. Heterogeneity in 12-month outcome among female and male smokers. *Addiction*. 2004; 99:237–250. [PubMed: 14756716]
- Tibshirani R. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society*. 1996; 58:267–288. Series B

- Weiss RD, Griffin ML, Najavits LM, Hufford C, Kogan J, Thompson HJ, Siqueland L. Self-help activities in cocaine dependent patients entering treatment: Results from NIDA Collaborative Cocaine Treatment Study. *Drug & Alcohol Dependence*. 1996; 43:79–86. [PubMed: 8957146]
- Wells EA, Donovan DM, Daley DC, Doyle S, Brigham G, Garret SB, Walker R. Is completion of level of exposure to a 12-Step facilitation therapy associated with treatment outcome? *Journal of Substance Abuse Treatment*. 2014 in press.
- Wells EA, Peterson PL, Gainey RR, Hawkins JD, Catalano RF. Outpatient treatment for cocaine abuse: A controlled comparison of relapse prevention and Twelve-Step approaches. *American Journal of Drug and Alcohol Abuse*. 1994; 20:1–17. [PubMed: 8192128]
- Zhang, H.; Singer, B. *Recursive Partitioning in the Health Sciences*. Springer-Verlag; New York, NY: 1998.
- Zhou, X-H.; Obuchowski, NA.; McClish, DK. *Statistical Methods in Diagnostic Medicine*. John Wiley & Sons, Inc.; New York: 2002.

Figure 1.

A single classification tree based on the total sample ($n = 234$), where C = completion and N = non-completion.

Table 1

Descriptive indices of the CART variable importance index and classification accuracy for a single classification tree (see Figure 1) with the total sample (n = 234).

Variable	Number of Splitting Rules	Importance Index
SUSS Disease Model	3	1.000
DTCQ-8 Global Confidence score (relapse risk)	2	0.863
SHAQ Number of types of self-help activities	1	0.839
ASI Drug Composite score	1	0.645
ASI Alcohol Composite score	2	0.639
SEQ-P Drug-taking confidence item	1	0.600

Classification Accuracy with 95% Confidence Intervals	
Misclassification Rate = 0.205 (0.153, 0.257)	Likelihood Ratio Positive = 2.245 (1.690, 2.982)
Sensitivity = 0.886 (0.827, 0.927)	Likelihood Ratio Negative = 0.188 (0.117, 0.302)
Specificity = 0.605 (0.493, 0.708)	Odds Ratio = 11.926 (6.087, 23.367)
False Negative Rate = 0.114 (0.064, 0.163)	False Negatives (n = 18)
False Positive Rate = 0.395 (0.285, 0.505)	True Negatives (n = 46)
Positive Predictive Value = 0.824 (0.766, 0.881)	False Positives (n = 30)
Negative Predictive Value = 0.719 (0.609, 0.829)	True Positives (n = 140)

Table 2

Descriptive indices of the CART variable selection, variable importance index and classification accuracy out of 100 randomly selected subsamples (n = 188 each).

Variable	K	P	D	Importance Index			
				Mean	Min	Max	SD
SHAQ Number of types of self-help activities	90	39	>	0.863	0.501	1.000	0.147
DTCQ-8 Global Confidence score (relapse risk)	79	17	>	0.796	0.435	1.000	0.156
SUSS Disease model subscale score	51	6	>	0.786	0.416	1.000	0.160
ASI Longest fulltime job in years	31	4	>	0.831	0.651	1.000	0.109
ASI Alcohol composite score	25	1	<	0.722	0.435	1.000	0.185
SHAQ Days attend AA,NA,CA,CMA meetings	24	0	>	0.612	0.293	0.884	0.138
ASI Drug composite score	22	8	<	0.774	0.292	1.000	0.230
<i>UDS positive urine screen for cocaine</i>	22	7	n	0.881	0.543	1.000	0.121
<i>UDS positive urine screen for marijuana</i>	22	8	n	0.841	0.530	1.000	0.162
SEQ-P Drug-taking confidence item	21	2	>	0.648	0.382	1.000	0.186
SOCRATES Ambivalence subscale score	20	0	>	0.648	0.378	1.000	0.144
SEQ-P Self-efficacy composite score	20	0	>	0.657	0.434	1.000	0.184
ASI Employment composite score	18	0	<	0.662	0.407	0.879	0.135
SOCRATES Taking Steps subscale score	18	0	<	0.660	0.448	0.983	0.157
ASI Medical composite score	14	0	<	0.719	0.412	0.998	0.179
SIBS-R Core Spirituality subscale score	13	0	<	0.648	0.369	0.758	0.102
<i>ASI Spent money on drugs in past 30 days</i>	13	1	n	0.697	0.503	1.000	0.146
ASI Legal composite score	12	0	<	0.761	0.373	1.000	0.207
SYRAAP Perceived Benefit subscale score	12	0	>	0.710	0.400	0.955	0.177
SYRAAP Perceived Severity subscale score	10	0	>	0.637	0.368	1.000	0.175
<i>DEM Gender (female)</i>	10	0	y	0.710	0.515	0.974	0.132
<i>ASI Ever arrested & charged for weapons offense</i>	10	4	n	0.788	0.640	1.000	0.126
<i>ASI Ever arrested & charged for contempt of court</i>	9	0	n	0.733	0.505	0.923	0.111
SIBS-R Acceptance/Insight subscale score	8	0	>	0.604	0.372	0.796	0.123
SYRAAP Perceived Barriers subscale score	8	0	<	0.696	0.476	0.885	0.119
<i>ASI Spend most free time alone vs. family/friends</i>	8	1	n	0.750	0.478	1.000	0.144
<i>ASI Received meds for psychological problem</i>	8	0	y	0.732	0.533	0.949	0.179
<i>DSM-IV Abuse of opiates</i>	7	0	n	0.768	0.557	0.972	0.126
SIBS-R Spiritual Involvement & Beliefs total score	7	0	<	0.671	0.526	0.853	0.107
<i>ASI Ever arrested & charged for drug charges</i>	7	0	n	0.777	0.452	1.000	0.216
SIBS-R Perspective Existential subscale score	6	0	<	0.597	0.329	0.771	0.171
<i>ASI Ever arrested & charged for other offenses</i>	6	0	n	0.669	0.402	0.870	0.170
<i>ASI Treated in outpatient setting for alcohol/drugs</i>	6	0	y	0.700	0.584	0.778	0.076
DEM Age in years	5	0	<	0.629	0.431	0.794	0.131
<i>DSM-IV Abuse of alcohol</i>	5	0	n	0.602	0.502	0.776	0.111
SOCRATES Readiness to Change total score	5	0	>	0.653	0.556	0.747	0.087

Indices of Classification Accuracy with 95% Confidence Intervals

	Mean	95% CI	Min	Max	SD
Misclassification Rate	0.209	0.174, 0.244	0.165	0.255	0.018
Sensitivity	0.865	0.785, 0.946	0.784	0.960	0.041
Specificity	0.641	0.458, 0.824	0.435	0.855	0.093
False Negative Rate	0.135	0.054, 0.215	0.040	0.216	0.041
False Positive Rate	0.359	0.176, 0.542	0.145	0.565	0.093
Positive Predictive Value	0.833	0.772, 0.894	0.775	0.917	0.031
Negative Predictive Value	0.706	0.615, 0.797	0.623	0.844	0.046
Likelihood Ratio Positive	2.569	1.568, 3.980	1.700	5.413	0.665
Likelihood Ratio Negative	0.207	0.126, 0.322	0.091	0.305	0.045
Odds Ratio	12.775	7.419, 19.402	6.741	22.525	3.345

K = number of times the variable was selected for a model

P = number of times the variable was the primary splitter

D = direction of effect (> and < indicate higher scores and lower scores, respectively, are associated with completion; whereas, "n" indicates no and "y" indicates yes to binary items in relation to completion)

Note: Descriptive indices of the importance index calculated are based on the number of times (K) the variable was selected. Dichotomous items are *italicized*.

Table 3

Classification accuracy and 95% confidence intervals for the total sample majority vote (n = 234) across 100 ensemble models.

Misclassification Rate = 0.111 (0.071, 0.151)	Likelihood Ratio Positive = 3.127 (2.245, 4.355)
Sensitivity = 0.987 (0.955, 0.997)	Likelihood Ratio Negative = 0.019 (0.005, 0.074)
Specificity = 0.684 (0.573, 0.778)	Odds Ratio = 169.002 (38.615, 739.643)
False Negative Rate = 0.013 (0.000, 0.030)	False Negatives (n = 2)
False Positive Rate = 0.316 (0.211, 0.420)	True Negatives (n = 52)
Positive Predictive Value = 0.867 (0.817, 0.916)	False Positives (n = 24)
Negative Predictive Value = 0.963 (0.913, 1.000)	True Positives (n = 156)