Substance Use & Misuse, 46:492–510, 2011

ISSN: 1082-6084 print / 1532-2491 online
DOI: 10.3109/10826081003682222

ORIGINAL ARTICLE

Using Decision Tree Analysis to Identify Risk Factors for Relapse to Smoking

Megan E. Piper[1,3], Wei-Yin Loh[2], Stevens S. Smith[1,3], Sandra J. Japuntich[4] and Timothy B. Baker[1,3]

[1]Center for Tobacco Research and Intervention, School of Medicine and Public Health, University of Wisconsin, Madison, Wisconsin, USA; [2]Department of Statistics, University of Wisconsin, Madison, Wisconsin, USA; [3]Department of Medicine, School of Medicine and Public Health, University of Wisconsin, Madison, Wisconsin, USA; [4]Mongan Institute for Health Policy, Tobacco Research and Treatment Center, Massachusetts General Hospital, Harvard University Medical School, Boston, Massachusetts, USA

This research used classification tree analysis and logistic regression models to identify risk factors related to short- and long-term abstinence. Baseline and cessation outcome data from two smoking cessation trials, conducted from 2001 to 2002 in two Midwestern urban areas, were analyzed. There were 928 participants (53.1% women, 81.8% White) with complete data. Both analyses suggest that relapse risk is produced by interactions of risk factors and that early and late cessation outcomes reflect different vulnerability factors. The results illustrate the dynamic nature of relapse risk and suggest the importance of efficient modeling of interactions in relapse prediction.

Keywords relapse, prediction, classification tree, regression, smoking

INTRODUCTION

Relapse or cessation failure is the modal outcome of smoking cessation attempts (Fiore et al., 2008). Despite the frequency of cessation failure, the accurate prediction of cessation success might yield important benefits. First, accurate identification of high-risk individuals might permit the allocation of intensive treatment on an empirically sound basis. Second, if we can identify those factors or situations that precipitate cessation failure or relapse, treatments might be designed or applied to mitigate such risk factors. Finally, knowing who is at risk, and the factors that index risk, could provide insight into mechanisms of dependence. Toward these goals, this research used a classification tree modeling approach to efficiently screen large numbers of variables to detect ordered relations that provide easily interpretable and accurate predictions of cessation outcomes. In other words, using variables from several diverse domains that have been theoretically linked to relapse (e.g., treatment/cessation methods, demographic, life context, and dependence variables), this research identified subgroups of individuals at higher risk for early and late relapse and specific variables that index this risk.

Theoretical and empirical work has implicated baseline person factors, environmental features, dependence, and smoking history variables in increasing risk for relapse. Such variables include the following: gender (Perkins, 2001; Wetter et al., 1999); living with a smoker (Derby, Lasater, Vass, Gonzalez, & Carleton, 1994; Garvey et al., 2000; Homish & Leonard, 2005; Osler & Prescott, 1998); smokers in the environment (Lu, Tong, & Oldenburg, 2001; Mermelstein, Cohen, Lichtenstein, Baer, & Kamarck, 1986; Morgan, Ashenberg, & Fisher, 1988); tobacco dependence (Alterman, Gariti, Cook, & Cnaan, 1999; Campbell, Prescott, & Tjeder-Burton, 1996; Harris et al., 2004; Hurt et al., 2002; Killen, Fortmann, Kraemer, Varady, & Newman, 1992; Patten, Martin, Calfas, Lento, & Wolter, 2001; Westman, Behm, Simel, & Rose, 1997; see also Fagerström & Schneider, 1989); length of abstinence in previous quit attempts (Garvey, Bliss,

This research was conducted at the University of Wisconsin–Madison and was supported by the Transdisciplinary Tobacco Use Research Center Award P50DA19706 from the National Institutes on Drug Abuse. Dr. Smith was partially supported by grant 1R03CA126406 from the National Cancer Institute. Dr. Piper was supported by an Institutional Clinical and Translational Science Award (University of Wisconsin–Madison; KL2 grant no. 1KL2RR025012-01). Dr. Loh was partially supported by the US Army Research Office under grants W911NF-05-1-0047 and W911NF-09-1-0205. Dr. Baker was partially supported by grant 1K05CA139871 from the National Cancer Institute.

Address correspondence to Megan E. Piper, Center for Tobacco Research and Intervention, 1930 Monroe St., Suite 200, Madison, WI 53711; E-mail: mep@ctri.medicine.wisc.edu

Hitchcock, Heinold, & Rosner, 1992; Ockene et al., 2000); alcohol consumption (Garvey et al., 1992; Hyland et al., 2004; McClure, Wetter, de Moor, Cinciripini, & Gritz, 2002; McKee, Maciejewski, Falba, & Mazure, 2003); age (Harris et al., 2004; Hurt et al., 2002; Miller, Ratner, & Johnson, 2003; Nides et al., 1995; Osler & Prescott, 1998); marital status (Derby et al., 1994); and educational attainment/socioeconomic status (Barbeau, Krieger, & Soobader, 2004; Eisinger, 1971; Fernandez et al., 2006; Hyland et al., 2004; Hymowitz, 1997; Hymowitz, Sexton, Ockene, & Grandits, 1991; Levy, Romano, & Mumford, 2005; Miller et al., 2003; Nollen et al., 2006; Osler, Prescott, Godtfredsen, Hein, & Schnohr, 1999; Shields, 2005; Siahpush, Heller, & Singh, 2005; Wetter et al., 2005). These studies identified a number of main effect predictors, suggesting that most predictors apply to all subjects. However, few studied interactions; so there is little information about whether the variables list above interact with one another to produce multiplicative patterns of risk or identify subgroups of smokers who might be at higher risk for relapse.

The majority of these studies relied on linear or logistic regression techniques to link risk factors with cessation outcome. This use of multivariate regression permits researchers to identify those variables that both individually and collectively predict outcomes across a group of smokers. With regard to a dichotomous outcome such as abstinence status, there is considerable evidence that logistic regression yields very good solutions when used as a "black box" (Lim, Loh, & Shih, 2000; Perlich, Provost, & Simonoff, 2003). However, regression approaches may have intrinsic biases that may affect the sorts of predictors that are identified. First, regression analysis is somewhat insensitive to variables that permit accurate prediction for a relatively small subgroup of smokers. It tests predictors on their ability to predict outcomes across an entire sample and may, therefore, identify variables that weakly predict outcomes for many individuals versus identifying variables that strongly predict outcomes for subgroups of individuals. Second, a regression model can be difficult to interpret, especially if it contains many predictor variables. The interpretation of interaction effects or cross products can be complicated, made even more so by the joint effect of more than one cross product in a model. In addition, it is very difficult to test interaction effects in a manner that is both comprehensive and methodologically principled because the number of interaction terms grows precipitously with increases in the number of predictors.[1] For example, if there are 70 predictors in a model, then there are $(70 \times 69)/2 = 2{,}415$ two-factor interaction terms associated with the 70 linear terms. In order to test for all interactions, one would need more than 2,485 observations. Although researchers usually only test a much smaller number of theoretically predicted interactions, this limits researchers' ability to explore the data completely. Finally, coefficients in a regression model measure only the residual effect of that variable, after all the other variables in the model are accounted for. This has two disadvantages. First, the values of the coefficients often change if other variables are added to or deleted from the model. Thus, each coefficient is conditional on other variables in a given model and cannot be evaluated on its merit alone. Second, if interaction terms are not thoroughly screened in a regression model, then the main-effect coefficients may be misleading because they may not apply to subjects in a uniform manner. There are no easy solutions to these challenges. For instance, stepwise variable addition and deletion, a strategy that restricts the size of regression models, exaggerates the statistical significance of the coefficient estimates (Miller, 2002; Zhang, 1992). Testing only a priori models may result in model misspecification. As a result of these factors, limitations in analytic strategies have limited our ability to identify which of the many candidate variables are the most important measures of risk.

Classification tree analysis is a complementary approach to logistic regression. A classification tree is a statistical model for predicting an outcome variable from the values of one or more predictor variables. The goal of a classification tree is to optimize prediction by iteratively dividing individuals into high- and low-risk groups. It is similar to polytomous logistic regression, in that the outcome variable takes a small number of values (e.g., smoking vs. nonsmoking). But unlike logistic regression, which models the log odds as a linear function of the predictor variables, a classification tree recursively partitions a data set into two or more subgroups such that the observations within a subgroup are more homogeneous than those across subgroups. Each partition is based on one predictor at a time with different classification trees using different algorithms to partition the data at each step. When numerous variables are tested at each recursive step, a classification tree selects the variable that most efficiently divides subjects on the basis of outcome likelihood. This allows researchers to examine potential interactions of a variable in one group versus another group (e.g., treatment effects on men vs. those on women). Therefore, a classification tree model is readily interpretable. In other words, decision trees offer a new way to look at complex data sets. They are not meant to replace traditional methods.

Classification or decision trees and logistic regression analyses are complementary approaches. In some situations, one approach is better than the other; see Perlich

[1] This comparison of regression and classification tree models with regard to interaction effects should not be taken to mean that subnode branching within classification tree models is mathematically equivalent to multiplicative interactions in regression models. The classification tree model illustrates interactions by showing that a variable predicts the dependent variable only among individuals who meet a certain threshold on a different variable. For instance, Figure 1 shows that treatment condition is related to abstinence status only if individuals smoke within 30 min of waking (FTND1). Thus, we can say that there is an interaction between FTND1 and treatment condition. Note that the interaction effect involves identifying the interaction variables (FTND1 and treatment condition) and identifying a threshold value for FTND1. Traditional regression approaches model interactions only through cross-product terms. However, both models indicate a predictive relation that differs as a function of another variable in the model.

et al. (2003) for a study that shows decision trees having superior prediction accuracy compared with logistic regression when the sample size is large. Classification trees can offer advantages over logistic regression in that they capture optimal, sequential decision rules that may possess clinical utility or theoretical significance, as they apply to specific subgroups of individuals. Classification tree analysis is more likely to detect variables that powerfully predict outcomes for just a subgroup of individuals. In essence, regression analysis relatively weighs pervasiveness, while classification tree analysis relatively weighs specificity.

Swan and his colleagues (Swan et al., 2003; Swan, Jack, Javitz, McAfee, & McClure, 2008; Swan, Javitz, Jack, Curry, & McAfee, 2004) have written that decision tree models are now widely accepted as providing a good complement to traditional methods. Swan & colleagues (2004) were the first to use such classification trees in the field of tobacco research. They used Classification and Regression Trees (CARTs; Breiman, Friedman, Olshen, & Stone, 1984) to explore gender differences in 12-month cessation outcome from a bupropion smoking cessation trial. The results revealed six subgroups for women based on the following: longest quit attempt, body-mass index (BMI), education, family history of depression, and number of previous quit attempts. Six subgroups were also found for men, which were based on the following: the Fagerström Test of Nicotine Dependence (FTND) score, longest quit attempt, previous use of nicotine replacement therapy, depression history, and years smoked. These findings differed from logistic regression findings previously reported by the same research group using the same sample (Swan et al., 2003). This supports the premise that there are predictors that are particularly relevant to subgroups of smokers and that regression analysis may not identify them. These researchers have also used CART to identify subgroups of at-risk smokers in two different treatment conditions (Swan et al., 2008).

In addition to illustrating the different results produced by a logistic regression versus a classification approach, the CART results of Swan et al. (2004) are interesting in that they show that patterns of risk factors differ across the sexes and that outcomes are not strongly related to treatment or traditional measures of nicotine dependence. This raises important questions regarding the relative importance of nicotine dependence versus variables that reflect person factors such as socioeconomic status (e.g., education), mental or physical health, or situational/contextual factors.

Swan et al. (2004) produced important and intriguing findings, but there are several reasons to conduct additional classification tree research relevant to cessation outcome. First, Swan & colleagues forced an initial sex-based subgroup separation based on their findings of gender differences in outcome. Further decision tree analyses that use gender as a predictor variable, rather than creating different trees for the two genders, are needed. In this way the status of sex as a predictor can be contrasted with other variables. Thus, we would be able to determine if sex, relative to other variables, is effective at identifying subgroups of smokers who differ meaningfully in cessation outcome. It might be the case that if all factors were used in the classification models, sex would not provide a basis for classification because its predictive validity would be accounted for by other variables that are correlated with it. In addition, in Swan et al. (2008) different trees were created for the different treatment groups, which did not allow the researchers to analyze the role of treatment in predicting risk for relapse in the sample when all of the other relapse prediction variables were included.

A second reason to conduct additional decision tree analyses is that the research of Swan et al. (2004, 2008) used only a single follow-up time point, 12 months postquit, as an outcome variable. It is possible that very different solutions would result if models were built using outcomes at different time points (e.g., as has been found in multivariate logistic regression approaches; Garvey et al., 1992; Hurt et al., 2002). For instance, Swan et al. found that treatment did not predict outcome at 12 months postquit. However, it may be that treatment exerts its effects on outcome at earlier time points (e.g., shortly after the end of treatment). Other predictors may differ in their ability to organize data as a function of follow-up latency. For instance, dependence might be more highly determinant of outcomes early (vs. late) in the postcessation period, as it might influence the severity of the withdrawal syndrome. However, factors such as the presence of smoking cues, living with a smoker, or social support might exert stronger effects over longer time periods (e.g., Mermelstein et al., 1986).

Finally, it is important to develop classification models with decision tree analytic methods other than CART. CART is biased toward choosing predictors with many values (e.g., BMI or years smoked) over predictors with few values (e.g., binary variables such as male vs. female or treatment vs. placebo) because the former allow more chances to divide the data into homogeneous subgroups (Loh & Shih, 1997). Other decision tree methods, including the Generalized, Unbiased Interaction Detection and Estimation (GUIDE; www.stat.wisc.edu/∼loh/guide.html) method, we will be using in this research have been shown to improve upon the average prediction accuracy of CART (Loh, 2002, in press).

The goals of this research are twofold: (1) to shed new light on relapse predictors and (2) to illustrate the value of a classification approach to relapse prediction and analysis. To accomplish these goals we identified predictors of cessation success among smokers at both early and late time points postquit (1 week, 8 weeks/end-of-treatment, and 6 months) using baseline variables assessed prequit. Second we compared the results yielded by classification tree analyses with those produced by a logistic regression algorithm to obtain complementary evidence on important predictors of cessation success. The two analytic procedures can be compared on the basis of overlap in predictors and the size and interpretability of the prediction models they yield.

METHODS

The data presented here were collected from two randomized placebo-controlled smoking cessation trials. Trial methods are discussed in more detail in Piper et al. (2007) and McCarthy, Bolt, & Baker (2007). In Study 1 (N = 608: Piper et al., 2007) participants were randomly assigned to one of the three treatment groups: active bupropion + active 4-mg nicotine gum (AA, n = 228), active bupropion SR + placebo nicotine gum (AP, n = 224), or placebo bupropion SR + placebo gum (PP, n = 156). All participants also received three brief (10-min) counseling sessions (1 week prequit, on the quit day, and 1 week postquit) designed to provide the most effective elements recommended by the 2000 Public Health Service Guideline: intratreatment social support, information and problem solving, and aid in seeking extratreatment social support (Fiore, Bailey, & Cohen, 2000). In Study 2 (N = 463: McCarthy et al., 2007) participants were randomly assigned to receive active bupropion + counseling (AC, n = 113), active bupropion + no counseling (ANc, n = 116), placebo + counseling (PC, n = 121), or placebo + no counseling (PNc, n = 113) in a 2 (active bupropion SR vs. placebo) × 2 (counseling vs. no counseling) factorial design. Counseling comprised eight sessions of brief (10-min) individual cessation counseling.

Participants

In both studies, participants were recruited through TV, radio, and newspaper advertisements and community flyers in Madison and Milwaukee, Wisconsin. Participants were eligible to participate if they smoked 10 or more cigarettes per day, were motivated to quit smoking, did not have any physical or mental health issues that would prevent them from participating in or completing the study, and were not pregnant or breast-feeding and took steps to prevent pregnancy during treatment.

Procedure

In both studies, eligible participants were invited to an orientation session at which they learned about the study and provided written informed consent, demographic information, and smoking history information, including a carbon monoxide (CO) breath test (participants were excluded if exhaled CO was <10 ppm). Participants also completed health-screening questionnaires—i.e., Michigan Alcoholism Screening Test (MAST; Selzer, Vinokur, & van Rooijen, 1975), Primary Care Evaluation of Mental Disorders (PRIME-MD; Spitzer et al., 1994), and Center for Epidemiologic Studies Depression Scale (CES-D; Radloff, 1977)—to assess for medical or psychological exclusion criteria, in addition to the multiple tobacco dependence measures listed below. At 6 months postquit, all participants who reported abstinence for the previous 7 days were scheduled to return to the clinic to provide a breath sample for CO analysis.

Predictors

Carbon Monoxide Assessment

Participants provided a breath sample to permit alveolar CO analysis to verify their smoking status and estimate their smoking heaviness. A Bedfont Smokerlyzer was used to measure the CO in the breath samples. Results were recorded as parts per million of CO.

Demographics and Smoking History

A demographics questionnaire assessed characteristics such as gender, ethnicity, age, marital status, education level, and employment. The Smoking History Questionnaire included items such as the number of cigarettes smoked per day, age of smoking initiation, smoking status (e.g., daily smoker, occasional smoker), number of quit attempts, longest time abstinent, and other smokers in the household. The Smoking History Questionnaire yielded a total of 25 variables.

Direct Assay of Dependence Criteria

The Direct Assay of Dependence Criteria (DADC) comprises 14 items designed to assess 3 dependence-related constructs: relapse likelihood, withdrawal symptoms, and self-administration. Each item is answered on a 7-point Likert scale. This measure was developed by the study authors for use in the clinical trials described above.

Fagerström Test of Nicotine Dependence

The FTND (Heatherton, Kozlowski, Frecker, & Fagerström, 1991) is a 6-item scale designed to measure tobacco dependence. Each item has its own individual response scale that varies by item. Previous research indicates that it has fair internal consistency (α = .61; Heatherton et al., 1991).

Nicotine Dependence Syndrome Scale

The Nicotine Dependence Syndrome Scale (NDSS; Shiffman, Waters, & Hickcox, 2004) is a 19-item self-report measure, comprising five theoretically derived subscales: Drive, Priority, Tolerance, Continuity, and Stereotypy. Each item is rated on a 5-point Likert scale from 1 = "Not at all true" to 5 = "Extremely true."

Tobacco Dependence Screener

The Tobacco Dependence Screener (TDS; Kawakami, Takatsuka, Inaba, & Shimizu, 1999) is a self-report measure designed to assess 10 of the criteria for tobacco dependence of the *Diagnostic and Statistical Manual of Mental Disorders–Fourth Edition* (*DSM-IV*; American Psychiatric Association, 1994), each on a dichotomous scale with 0 indicating lack of the symptom and 1 indicating endorsement of the symptom. Research has shown that the TDS has good internal consistency (α ranging from .76 to .81 across three studies; Kawakami et al., 1999).

Wisconsin Inventory of Smoking Dependence Motives

The Wisconsin Inventory of Smoking Dependence Motives (WISDM; Piper et al., 2004) comprises 68 items designed to assess 13 different theoretically derived motivational domains: Affiliative Attachment, Automaticity, Behavioral Choice/Melioration, Cognitive Enhancement, Craving, Cue Exposure/Associative Processes, Loss of Control, Negative Reinforcement, Positive Reinforcement, Social and Environmental Goads, Taste and Sensory Properties, Tolerance, and Weight Control. Each item is

answered on a 7-point Likert scale ranging from 1 = "Not true of me at all" to 7 = "Extremely true of me." A subscale is scored by taking the average of all of the answers relevant to that subscale.

Outcomes

Seven-day point-prevalent abstinence was assessed at each study visit. For this research, the important outcomes were smoking at 1 week postquit (early cessation failure), end-of-treatment (8 weeks postquit), and 6 months postquit. We also examined early (end-of-treatment) and late (6 months postquit) relapse by removing individuals who were smoking in the first week (cessation failures) from the analyses. Smoking status was verified using CO ratings (CO < 10 = abstinent). Using the intent-to-treat principle, individuals who could not be reached for a specific follow-up were considered to be smoking at that follow-up.

Analytic Strategy

Using a listwise deletion procedure, we identified 928 participants who had complete data for the 70 baseline variables of interest (see the Appendix). We deliberately included only participants with complete data so that the type of the analysis would not be confounded by approach to missing data. The predictor variables were selected based on previous research and theory as to their relations with relapse. The variables were analyzed as predictors of 1-week, end-of-treatment, and 6-month postquit abstinence using both the GUIDE classification tree program and a stepwise logistic regression algorithm (Loh, 2002, 2010). The GUIDE method, which does not have the selection bias of the CART method, uses chi-square tests to measure the degree of association between the dependent variable and each predictor variable (the range of each continuous predictor variable is divided into four groups at its sample quartiles for this purpose). The most significant predictor variable is selected to form the partition. If the selected variable X is continuous, the method searches for a split of the form, "X < c," with c chosen to make both of the resulting data subsets as homogeneous as possible. If X is categorical (e.g., marital status), the best split of the form, "X in S," is found, where S is a subset of the values taken by X. This step is applied recursively to each partition, and the whole process can be described by a tree structure. Partitioning stops when the sample size is less than 20. Since the resulting tree model probably overfits the data, a sequence of smaller tree models is obtained by sequentially pruning the tree structure until only one node is left, using the same internal-validation method as CART. Finally, the tree model with the lowest estimate of prediction error is chosen. In sum, the GUIDE classification tree program analyzes pairs of variables recursively, and because it examines variable effects in subgroups determined by prior variable cut scores, it is able to accommodate interactions when choosing the best variable to split the sample (its subsequent cuts account for the levels of previously entered variables).

The stepwise logistic regression algorithm was conducted using R and the Akaike Information Criterion (AIC). The AIC algorithm is designed to include as many terms as needed to minimize the AIC criterion function and may include terms that are not statistically significant. As such, the AIC model is known to overfit the data (Simonoff, 2003). We did not use a usual stepwise procedure based on p-values because it is inapplicable when the number of potential terms (main effects plus all possible interactions) exceeds the sample size. Some of the regression models included interactions, while others included only main effect terms.

RESULTS

Participant Characteristics

The combined sample from Studies 1 and 2 comprised 1,071 smokers, 928 of whom had complete data for the 70 predictor variables used in the analyses. See Table 1 for demographic information. While the two study samples were comparable, there were statistically significant differences in the number of women (χ^2 (1, N = 928) = 5.01, p = .03), racial composition (χ^2 (1, N = 928) = 39.21, $p < .01$), educational attainment (χ^2 (1, N = 928) = 23.37, $p < .01$), and age (t (926) = −5.12, $p < .01$), although some of these differences were relatively modest.

One Week Postquit

At the end of the first week postquit, approximately 46% of participants reported no smoking after their quit day.

TABLE 1. Demographics

	Combination of Studies 1 and 2 (N = 928)	Study 1 (N = 554)	Study 2 (N = 374)
Women[a] (%)	53.1	56.1	48.7
White[a] (%)	81.8	75.3	91.4
High school or greater education[a] (%)	92.7	91.0	95.2
Married or living with a partner (%)	54.8	56.9	51.9
Age (*SD*)[a]	40.32 (11.68)	41.93 (11.37)	37.98 (11.74)
Mean cigarettes per day (*SD*)	22.22 (9.91)	22.64 (9.87)	21.60 (9.94)
Baseline CO in ppm (*SD*), n = 896	25.12 (11.29)	25.70 (11.13)	24.28 (11.49)

[a]Statistically significant differences between Study 1 and Study 2.

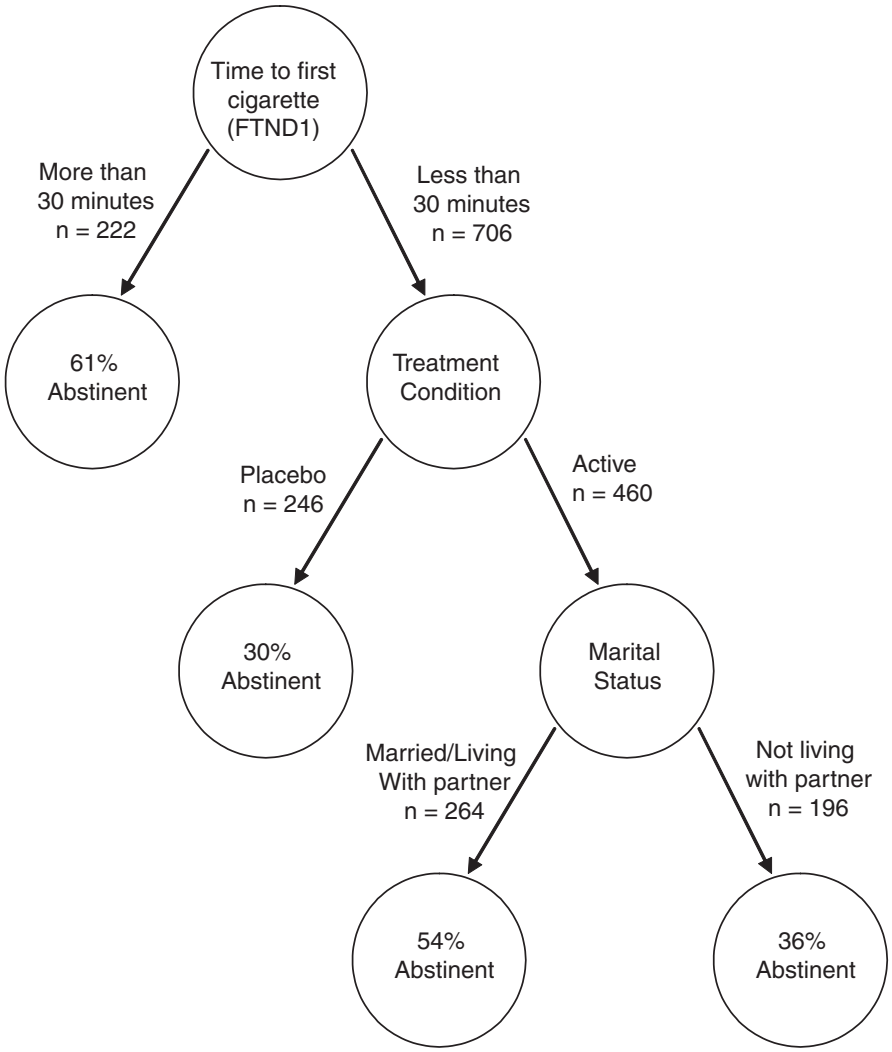

FIGURE 1. GUIDE decision tree predicting abstinence at 1 week postquit ($N = 928$; average abstinent rate = 45.7%).

Figure 1 shows a pruned classification tree model for predicting abstinence at the end of the first week of treatment. The first variable chosen by GUIDE to split the data was FTND Item 1, "How soon after you wake do you smoke your first cigarette?" Participants who reported smoking their first cigarette at least 30 min after waking went to the left node, while those who smoked their first cigarette within 30 min went to the right node. There are four leaf nodes and, hence, four subgroups. The subgroup consisting of the participants who waited at least 30 min to smoke their first cigarette ($n = 222$) had the highest abstinence rate of 61%. Participants who smoked within 30 min after waking and received placebo medication had the lowest abstinence rates ($n = 246$; 30%). Of those individuals who smoked within 30 min of waking and received active medication, those who were married or living with a partner had higher abstinence rates ($n = 264$; 54%) than those who were not married or living with a partner ($n = 196$; 36%).

End of Treatment

At the end of treatment (8 weeks postquit), 29.7% of the sample reported 7-day point-prevalence abstinence. The classification tree identified one predictor consistent with the 1-week tree, namely, treatment condition, and two new predictors, namely, income and self-reported health status (Figure 2). Individuals who received placebo medication had the lowest abstinence rates ($n = 335$; 20%). However, their abstinence rate was similar to that of individuals who received active medication but reported that their household income was less than $35,000 and that their health was good/fair/poor/don't know ($n = 172$; 21%). Individuals with the highest abstinence rates at the end of treatment were those who received active medication and had a household income of $35,000 or greater ($n = 327$; 43%), and those who received active medication had a household income of less than $35,000 but had excellent/very good health ($n = 94$; 36%).

The data were then reanalyzed excluding individuals who were smoking in the first week postquit (i.e., cessation failures) to assess predictors of ability to establish and maintain abstinence through treatment. The average end-of-treatment abstinence rate among smokers who were able to quit during the first week was 54.7%. Marital status, gender, and age when the individual began daily smoking were the significant predictors of being able to

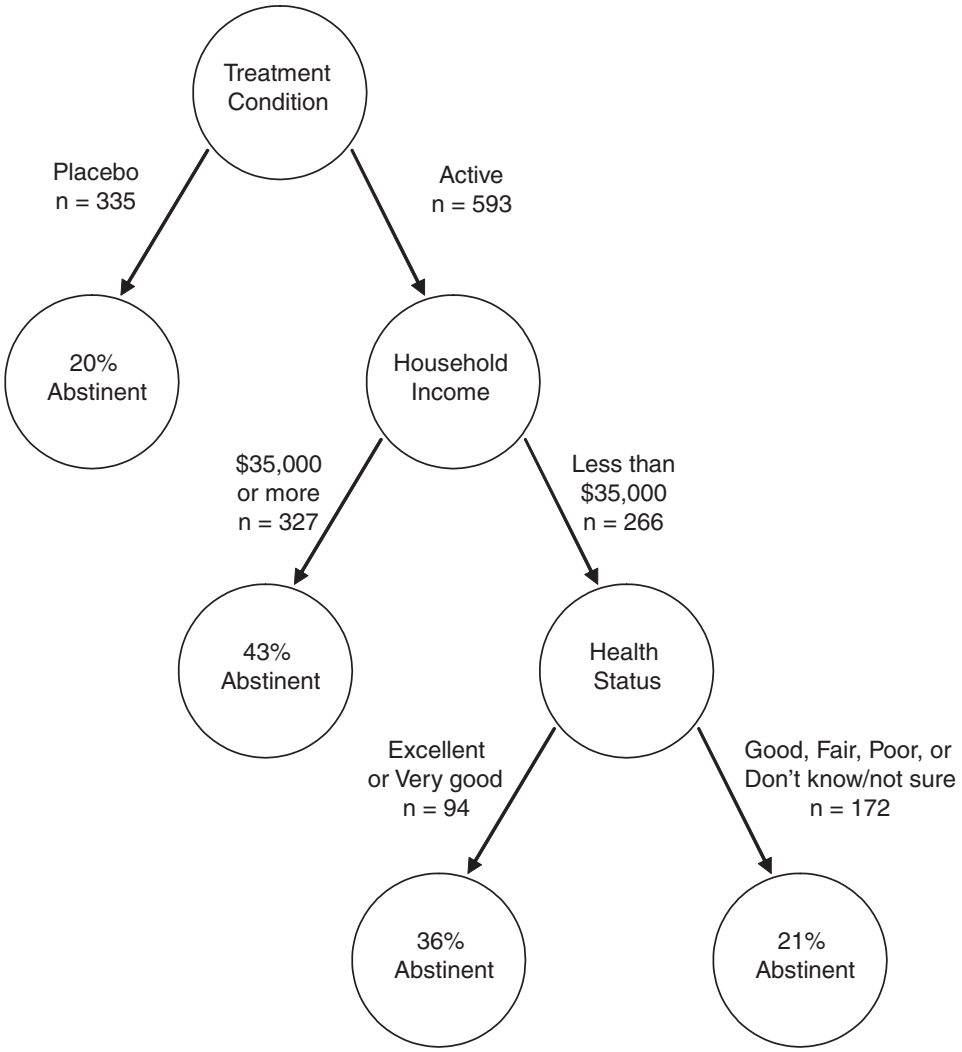

FIGURE 2. GUIDE decision tree predicting abstinence at the end of treatment ($N = 928$; average abstinent rate = 29.7%).

establish and maintain abstinence through treatment (Figure 3). Results indicate that men who are married or living with a partner and individuals who started daily smoking after age 16 and were not married or living with a partner were the most likely to be abstinent at the end of treatment ($n = 141$; 68% and $n = 81$; 60%, respectively), if they were able to achieve initial abstinence in the first week postquit. Approximately half of the women who were married or living with a partner ($n = 112$) maintained abstinence through the end of treatment. Only one third of individuals who were not married or living with a partner and started smoking daily by age 16 ($n = 90$) were able to maintain abstinence through the end of treatment.

Six Months Postquit

The overall abstinence rate at 6 months postquit was 18.5%. The classification tree model had only two predictors—health status and longest previous quit attempt (see Figure 4). Individuals who were in good, fair, or poor health but had a history of being able to quit for more than 5 months had the highest abstinence rates ($n = 282$; 25%). Individuals who reported very good or excellent health had the next highest abstinence rates ($n = 109$; 20%). Individuals in good, fair, or poor health who had been unable to quit for more than 5 months had the lowest abstinence rates at 6 months postquit ($n = 537$; 15%).

The GUIDE model predicting abstinence only among individuals who were abstinent at 1 week postquit was completely different from the model predicting abstinence using the entire sample (see Figure 5). The average abstinence rate at 6 months postquit for those who achieved initial abstinence during the first week was 32.5%. Among this group, self-reported feelings of dependence and marital status predicted abstinence at 6 months postquit. Individuals who reported that they had never felt dependent on tobacco had the highest abstinence rates ($n = 21$; 57%). However, it should be noted that only 21 of the 424 individuals in the sample reported never having felt dependent. Of those who reported that they had ever felt dependent on tobacco, individuals who were married or living with a partner had higher abstinence rates ($n = 242$; 36%) than did individuals who were not married or living with a partner ($n = 161$; 24%).

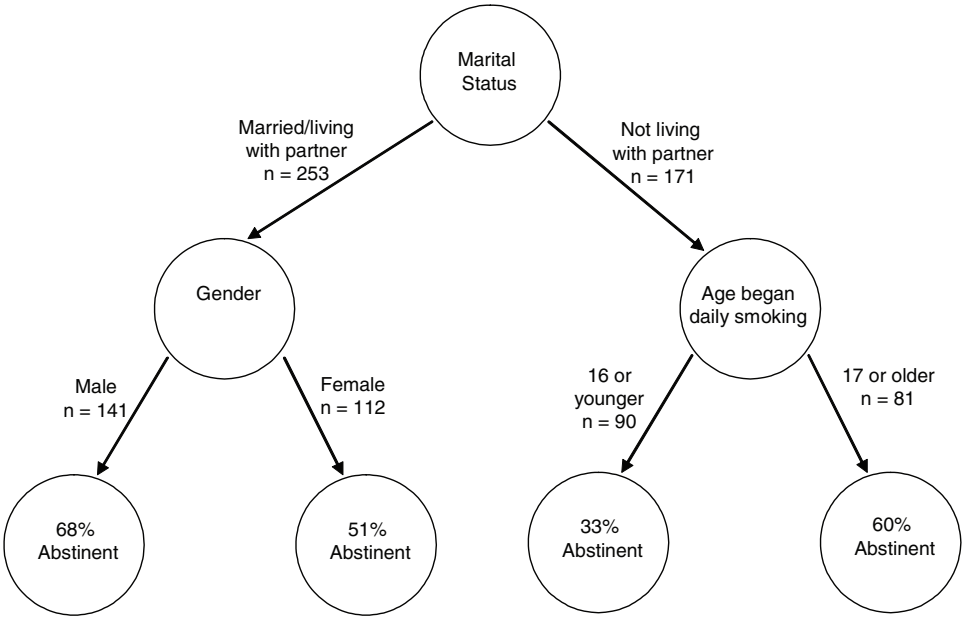

FIGURE 3. GUIDE decision tree predicting abstinence at the end of treatment excluding individuals who relapsed in the first week ($N = 424$; average abstinent rate = 54.3%).

Logistic Regression Analyses

The results for the AIC stepwise logistic regression analyses predicting 6-month abstinence are presented in Tables 2–5. When all smokers and only main effects were included in the analysis, the AIC regression yielded 14 predictors, 7 of which were statistically significant ($p < .05$) and 1 of which was also included in the GUIDE model—longest previous quit attempt (see Table 2). When data from all smokers were used and main effects and all possible interactions were analyzed, the regression yielded 52 predictors: 18 main effects (8 statistically significant) and 36 interaction effects (23 statistically significant; see Table 3). Both of the GUIDE predictors, health status and longest previous quit attempt, were included in this AIC model. When the analyses were done using only those individuals who achieved initial abstinence and only

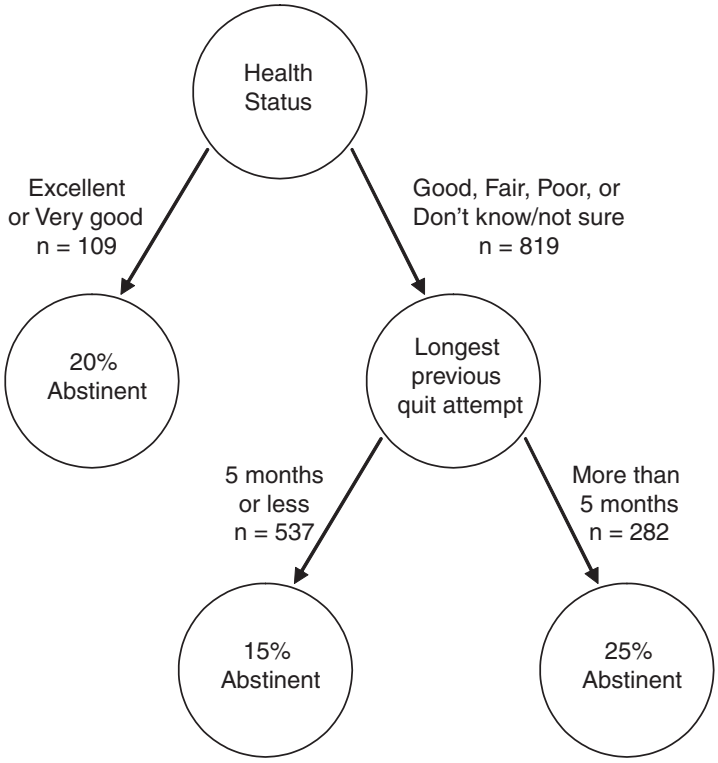

FIGURE 4. GUIDE decision tree predicting abstinence at 6 months postquit ($N = 928$; average abstinent rate = 18.5%).

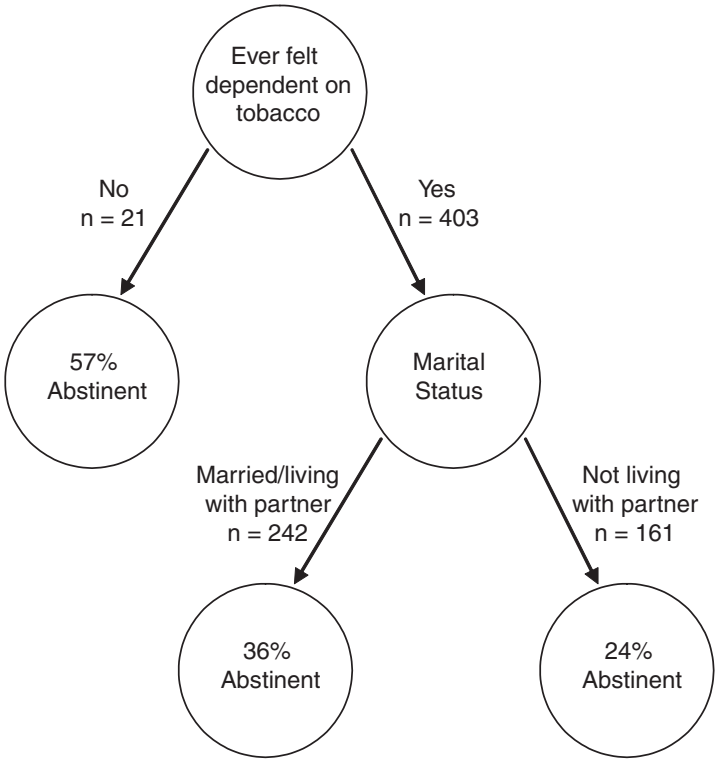

FIGURE 5. GUIDE decision tree predicting abstinence at 6 months postquit excluding individuals who relapsed in the first week ($N = 424$; average abstinent rate = 32.5%).

main effects were analyzed, the AIC model had 11 predictors, 6 of which were statistically significant (see Table 4). Both of the GUIDE predictors, marital status and ever felt dependent on tobacco, were included in this AIC model. When the data set was restricted to those who achieved initial abstinence and main effects and interactions were included in the model, there were 21 predictors: 12 main effects (5 statistically significant) and 9 interaction effects (4 statistically significant; see Table 5). As with the main-effects-only model, both the GUIDE predictors were included in this AIC model.

DISCUSSION

The first goal of this paper was to use a somewhat novel methodology, the GUIDE decision tree method, to provide insight into the factors that predict abstinence following a quit attempt and how these factors might interact

TABLE 2. Predicting abstinence at 6 months postquit using the AIC logistic regression model without interactions

Variable	Estimate	*SE*	*p*-value
Intercept	−2.50	0.84	.003
Income	0.13	0.06	.04
Time to first cigarette in the morning (FTND1)	−0.28	0.11	.01
Treatment condition	0.57	0.20	.003
Longest previous quit attempt[a]	0.07	0.04	.07
BMI	0.03	0.01	.04
Age first at which smoked a cigarette	0.05	0.02	.02
Positive affect (PPANAS)	−0.02	0.01	.07
Gender	−0.55	0.20	.01
WISDM Weight Control	0.10	0.06	.08
Number of years smoked daily	0.01	0.01	.14
Number of cigarettes smoked per day (FTND4)	−0.26	0.13	.04
Race	−0.45	0.27	.09
Household smoking restriction	−0.32	0.20	.10
Marital status	−0.30	0.21	.16

[a]Variable present in the GUIDE model.
Note: PPANAS = Positive Affect Scale of the Positive Affect Negative Affect Schedule.

TABLE 3. Predicting abstinence at 6 months postquit using the AIC logistic regression model including all interactions

Variable	Estimate	*SE*	*p*-value
Intercept	2.81	3.43	.41
Income	−0.84	0.42	.04
Time to first cigarette in the morning (FTND1)	−0.34	0.64	.59
Treatment condition	−0.65	0.93	.49
Longest previous quit attempt[a]	0.14	0.29	.63
BMI	−0.07	0.05	.16
Age at which first smoked a cigarette	0.12	0.08	.13
Positive affect (PPANAS)	−0.01	0.06	.90
Gender	1.10	1.66	.51
WISDM Behavioral Choice/Melioration	1.47	0.46	.002
WISDM Positive Reinforcement	−0.90	0.46	.05
Number of cigarettes smoked per day (FTND4)	−1.47	0.75	.05
Ever given up work or social activities to use tobacco	−7.17	2.30	.002
Household smoking restriction	−1.64	0.85	.05
Race	0.16	0.78	.84
WISDM Automaticity	0.38	0.24	.11
WISDM Weight Control	−0.59	0.25	.02
Time since previous quit attempt	−0.29	0.28	.30
Health status[a]	−1.64	0.49	.001
Longest previous quit attempt × Positive affect	−0.01	0.006	.03
Time to first cigarette × Positive affect	−0.03	0.02	.07
BMI × WISDM Positive Reinforcement	0.04	0.01	.001
Gender × Ever given up work or social activities to use tobacco	−2.02	0.69	.004
Number of cigarettes smoked per day × Ever given up work or social activities to use tobacco	1.12	0.40	.005
Positive affect × Number of cigarettes smoked per day	−0.26	0.11	.02
Time to first cigarette × Treatment condition	0.55	0.26	.03
BMI × Gender	−0.08	0.03	.02
Income × Gender	−0.29	0.13	.03
Ever given up work or social activities to use tobacco × Race	−18.99	651.66	.98
Longest previous quit attempt × Race	−0.36	0.13	.004
Positive affect × Ever given up work or social activities to use tobacco	0.12	0.05	.02
Longest previous quit attempt × Gender	0.25	0.09	.01
Household smoking restriction × WISDM Automaticity	0.26	0.14	.05
WISDM Positive Reinforcement × WISDM Automaticity	−0.11	0.06	.04
Treatment condition × WISDM Automaticity	−0.37	0.15	.01
Race × WISDM Weight Control	0.52	0.18	.004
Number of cigarettes smoked per day × WISDM Weight Control	0.16	0.08	.05
Positive affect × Gender	0.06	0.03	.06
WISDM Positive Reinforcement × WISDM Weight Control	0.10	0.05	.03
Age first at which smoked a cigarette × WISDM Behavioral Choice/Melioration	−0.05	0.02	.03
Age at which first smoked a cigarette × Cigarettes smoked per day	0.06	0.04	.08
Time to first cigarette × Cigarettes smoked per day	0.33	0.16	.04
WISDM Behavioral Choice/ Melioration × Ever given up work or social activities to use tobacco	0.46	0.25	.07
Household smoking restriction × Time since previous quit attempt	−0.46	0.14	.002
WISDM Positive Reinforcement × Time since previous quit attempt	0.12	0.05	.01
Gender × Time since previous quit attempt	−0.25	0.15	.08
Gender × Race	−1.07	0.64	.09
Time to first cigarette × Household smoking restriction	0.47	0.24	.06
Treatment condition × Health status	0.78	0.28	.01
Income × Health status	0.18	0.07	.01

(Continued on next page)

TABLE 3. Predicting abstinence at 6 months postquit using the AIC logistic regression model including all interactions *(Continued)*

Variable	Estimate	*SE*	*p*-value
Income × Positive Affect	0.02	0.01	.06
Longest previous quit attempt × Health status	0.10	0.06	.08
Longest previous quit attempt × Household smoking restriction	0.15	0.09	.09
WISDM Behavioral Choice/Melioration × WISDM Positive Reinforcement	−0.10	0.07	.14
Income × Time since previous quit attempt	0.06	0.04	.14

[a]Variable present in the GUIDE model.
Note: PPANAS = Positive Affect Scale of the Positive Affect Negative Affect Schedule.

to reveal predictors for specific subgroups of smokers. In addition, this study was designed to examine whether different predictors are relevant at different points during the quit attempt (e.g., whether treatment predicts late as well as early success).

Results of the GUIDE analyses revealed that individuals who delay smoking their first cigarette of the day by 30 min or more have the greatest likelihood of achieving initial abstinence in a quit attempt. But for those who smoke their first cigarette within 30 min after waking, which is a sign of strong nicotine dependence, receiving pharmacotherapy and being married/living with a partner predict initial success. These results also demonstrate the importance of providing pharmacotherapy to individuals who smoke their first cigarette within 30 min of waking.

While treatment continues to be an important predictor of cessation success at the end of treatment, income and health status are also important. Individuals with household incomes of at least $35,000 were more likely to be able to quit. If an individual had a household income of less than $35,000, then health status was a significant predictor of abstinence. It may be that good/fair/poor health predicts cessation failure because only highly relapse-vulnerable individuals have continued to smoke in the face of significant health risk. When the analyses of end-of-treatment abstinence focused on those who had achieved initial abstinence, marital status was once again a significant predictor. It may be that being married or living with a partner provides an extra motivation or extra support in dealing with the challenges of quitting. Or this variable may be a more general indicant of social functioning. Interestingly, marital/partner status prognosticated greater success for males than females. Thus, this finding adds to the findings by Swan and colleagues (2004), showing differential abstinence prediction as a function of gender. Results also showed that for those who did not live with a partner, the age of initiating daily smoking was an important predictor. These findings resonate with the recent genetics findings that individuals with a certain genetic loading are more likely to develop strong nicotine dependence if they begin daily smoking prior to age 17 (Weiss et al., 2008).[2]

[2]It should be noted that the data presented here were used in the analyses of Weiss et al. (2008).

At 6 months postquit, health status and longest previous quit attempt are the most important predictors of cessation success. Good/fair/poor health and an inability to quit for more than 5 months in the past combined to produce the lowest abstinence rates. As in the end-of-treatment analyses, the impact of health status may be an index of motivation (e.g., wanting to quit because of negative health) but may also index dependence as those who have tried to quit before because of health problems but have been unsuccessful, possibly owing to strong nicotine dependence, may continue to have difficulty quitting. When the results focused on smokers who achieved initial abstinence in the first week of the quit attempt, outcome was predicted by self-reported dependence for a small number of participants (only 21 of the 424 reported that they had never felt dependent). However, marital status predicted outcome for the majority of the individuals.

It is also important to keep in mind that, as is the case with all approaches to prediction, variables may reflect the effects of correlated variables. It is unknown, therefore, why individuals who are married or living with a domestic partner are less likely to relapse than are other individuals. It may be, for instance, that the social support that occurs in marriage helps smokers maintain abstinence. Or marital status might index a variety of person factors such as neuroticism, social skills, socioeconomic status, and intelligence, and cessation medication is most helpful in persons with these factors. With respect to household income being predictive of outcome, research has shown that socioeconomic status is related to many factors that may affect cessation success, including the following: availability for treatment scheduling (Macken, Wilder, Mersy, & Madlon-Kay, 1991), financial and other stress (De Vogli & Santinello, 2005; McKee et al., 2003; Siahpush et al., 2005), living with smokers or having a partner who smokes (Chandola, Head, & Bartley, 2004; Graham, Francis, Inskip, & Harman, 2006; Honjo, Tsutsumi, Kawachi, & Kawakami, 2006), presence of smoking at work and in one's peer group, (Honjo et al., 2006), and having a blue-collar job (Sorensen, Gupta, & Pednekar, 2005). These factors suggest that a low-income smoker might have more stress and less support for quitting and may live in an environment richer in smoking cues. Household income may also implicate mental illness

TABLE 4. Predicting abstinence at 6 months postquit using the AIC logistic regression model without interactions, excluding 1-week relapsers

Variable	Estimate	*SE*	*p*-value
Intercept	−0.80	0.89	.37
Ever felt dependent on tobacco[a]	−1.26	0.49	.01
Marital status[a]	−0.55	0.23	.02
Ever diagnosed with or treated for a drug use disorder	−1.19	0.66	.08
Health status	0.30	0.14	.03
Time since last quit attempt	0.15	0.08	.06
Age at which first smoked a cigarette	0.05	0.03	.05
Negative affect (NPANAS)	−0.04	0.02	.05
Time to first cigarette (FTND1)	−0.47	0.18	.01
Household smoking restriction	−0.38	0.23	.10
WISDM Tolerance	0.18	0.12	.14
Number of years smoked daily	0.02	0.01	.15

[a]Variable present in the GUIDE model.
Note: NPANAS = Negative Affect Scale of the Positive Affect Negative Affect Schedule.

as a mechanism of relapse, since severe mental illness is associated with lower employment status (McIntyre et al., 2008; Ridgeway & Rapp, 1998). Future research should explore the mechanisms via which life-context variables are related to abstinence outcomes. This is especially true for such variables as marital status and income, which appear to be especially prognostic of outcome. It is possible that their influence is mediated by factors that can be affected by intervention (e.g., social and material resources, knowledge).

Another important facet of the GUIDE algorithm is its ability to determine cut scores. If the predictor variable selected to split a node in the tree takes ordered values (e.g., age), GUIDE finds the cut score for that variable that makes the observations in the resulting subnodes as homogeneous as possible in terms of the binomial variance

TABLE 5. Predicting abstinence at 6 months postquit using the AIC logistic regression model including all interactions, excluding 1-week relapsers

Variable	Estimate	*SE*	*p*-value
Intercept	−0.91	5.37	.86
Ever felt dependent on tobacco[a]	2.68	5.16	.60
Marital status[a]	0.64	0.25	.01
Ever diagnosed with or treated for a drug use disorder	−35.93	14.97	.02
Health status	4.67	2.81	.10
Time since last quit attempt	−0.21	0.28	.46
Age at which first smoked a cigarette	0.09	0.10	.38
Negative affect (NPANAS)	−0.21	0.08	.01
Number of years smoked daily	−0.14	0.08	.07
Time to first cigarette (FTND1)	−2.31	1.16	.05
Household smoking restriction	1.71	0.74	.02
WISDM Tolerance	0.21	0.13	.09
Use tobacco despite mental problems	0.91	0.56	.10
Ever diagnosed with or treated for a drug use disorder × Negative affect	1.19	0.53	.03
Ever felt dependent on tobacco × Health status	−4.97	2.80	.08
Ever felt dependent on tobacco × Number of years smoked daily	0.17	0.08	.03
Ever diagnosed with or treated for a drug use disorder × Number of years smoked daily	0.39	0.17	.03
Age at which first smoked a cigarette × Negative affect	0.01	0.01	.12
Ever felt dependent on tobacco × Time to first cigarette	1.96	1.15	.09
Negative affect × Household smoking restriction	0.08	0.04	.08
Time to first cigarette × Use tobacco despite mental problems	−0.79	0.30	.01
Health status × Time since last quit attempt	0.15	0.11	.16

[a]Variable present in the GUIDE model.
Note: NPANAS = Negative Affect Scale of the Positive Affect Negative Affect Schedule.

of the outcome variable. If the predictor variable takes unordered values (e.g., marital status), GUIDE finds the best grouping of values to minimize the binomial variances in the subnodes. However, it should be noted that the cut scores, and indeed the overall predictions, will be highly affected by the populations involved, item response scales, and the other items submitted to analysis.

In sum, at all three time points (1 week, end-of-treatment, and 6 months postquit), both tobacco-related factors (e.g., treatment, dependence) and life-context factors (e.g., health status, marital status, income) are important in determining the likelihood of cessation success, including ability to achieve and maintain abstinence. These data reinforce the notion that only a portion of the variation in relapse is related to dependence and that covariates should be used to distill the portion of relapse that is reflective of dependence (Piper, McCarthy, & Baker, 2006; Uhl et al., 2007). However, it is important to note that the predictive power of the specific tobacco-related or life-context variable appears to change based on the time period being predicted, although marital status did predict outcome at all three time periods. Marital status appears to be the only variable that predicted not only the ability to establish initial abstinence but also the ability to maintain it over time. The results agree with theoretical models that hold that the determinants of relapse change with time (Piasecki, Fiore, McCarthy, & Baker, 2002; Shiffman, 1993).

These results share some similarity with those reported by Swan and colleagues (2004). These researchers found that for women, longest previous quit attempt, education, number of previous quit attempts, BMI, and family history of depression were significant predictors of 12-month abstinence. For men, FTND total score, longest previous quit attempt, previous use of nicotine replacement therapy, depression history, and years smoked were the significant predictors. There is consistency in both analyses regarding the ability of past behavior (longest previous quit attempt) to predict future abstinence. In addition, both sets of analyses reveal the importance of dependence variables as well as life-context/demographic variables in predicting cessation success. The predictive validity of contextual/demographic variables (e.g., marital status, income, education) may have important clinical implications. Uncovering the reasons that these variables predict treatment failure may suggest new interventions designed to buffer their mechanisms of risk (Shiffman, 1993).

The second goal of this study was to compare the more traditional regression approach with the classification tree approach. The former contrasts variables on the basis of prediction across a whole population. The latter recursively partitions individuals on the basis of the information value of predictors in providing optimal classification of subgroups of individuals. The current results illustrate the very different results that can be obtained from the two different methods, despite the fact that both of them used the same subjects, the same independent variables, and the same dependent variables. As such, regression and classification tree approaches should be viewed as complementary analytic approaches, and researchers can gain additional perspective on a phenomenon by using both approaches.[3]

Consistent with previous research, the AIC regression models identified gender (Perkins, 2001; Wetter et al., 1999), tobacco dependence (Alterman et al., 1999; Campbell et al., 1996; Harris et al., 2004; Hurt et al., 2002; Killen et al., 1992; Patten et al., 2001; Westman et al., 1997), length of abstinence in previous quit attempts (Garvey et al., 1992; Ockene et al., 2000), marital status (Derby et al., 1994), and educational attainment/socioeconomic status (Barbeau et al., 2004; Eisinger, 1971; Fernandez et al., 2006; Hyland et al., 2004; Hymowitz et al., 1991, 1997; Levy et al., 2005; Miller et al., 2003; Nollen et al., 2006; Osler et al., 1999; Shields, 2005; Siahpush et al., 2005; Wetter et al., 2005) as predictors of outcome. They did not identify alcohol use or age as being related to outcome.

However, the use of the AIC regressions illustrated several complications that arise when using a logistic regression approach to sift and winnow through a multitude of predictors. First, we were unable to use a more traditional *p*-value-based logistic regression because the number of predictors, including both main effects and all possible two-way interactions, exceeded the sample size. In exploratory analyses, when all possible predictors and combinations need to be assessed, logistic regression approaches may not be suitable unless there are a priori predictions or interactions that are intrinsically interesting or theoretically relevant and the model building is designed to test these. This strategy, though, can result in model misspecification if the a priori model does not include important predictors. The decision tree methodology is specifically designed to handle a large number of predictors, perhaps making it better suited to large exploratory analyses, relative to the logistic regression analyses.

Second, when there are numerous predictors, the AIC regression model will overfit the data, producing rather large models. Compared with the GUIDE models that included only two or three predictors, the AIC model that included 52 predictors seems somewhat excessive. It should be noted that using the model with only main effects is not sufficient in that the models with the interactions included revealed that the interaction terms were the best predictors. However, with large models like this, interpretation becomes an important concern. Even if the number of predictors is narrowed to only those demonstrating statistical significance, interpretation becomes a substantial problem. One of the AIC models had 8 significant main effects and 23 interaction effects. Interpreting interactions or cross products is challenging enough; but when there are several interactions in a single model, forming an understanding of the joint interpretation of all of the interactions becomes even more complex. Comparatively, the GUIDE model produces a parsimonious solution that is readily interpretable in terms of dividing the sample into

[3] It should be noted that we conducted analyses using RPART (Atkinson & Therneau 2000), an R implementation of CART. The CART trees were much bigger than the GUIDE trees.

groups that are at higher or lower risk for relapse. The GUIDE results also provide information on which variables are the most important or best predictors, based on which variables are chosen at the initial nodes as well as information on the order in which the interactions become operative.

While much previous research on relapse prediction has not focused on interaction terms, the present results suggest that interactions may be highly informative. This suggests that we may be able to predict the fate of smokers better and identify who needs more or different treatments if we take interactions into account. However, we must replicate these findings, since interaction effects can be highly affected by sampling error. The use of numerous highly correlated variables in prediction models increases the likelihood that different solutions will be obtained across samples. This is one of the limitations of classification tree analyses. By design, a classification tree model gives a stepwise description of a data set. At each step, the most predictive explanatory variable is selected to split a node of the tree. After splitting on the variable, its explanatory power is depleted, and other variables may be selected in subsequent steps. However, if two or more highly correlated explanatory variables are equally predictive, at most one is selected. Thus the main limitation of such a model is that the variables in the model may not be the only important ones. Of course, multicollinearity affects other statistical methods too, including logistic regression where it makes the estimated regression coefficients more difficult to interpret and increases their variability. One way to see if multicollinearity is a serious concern in a classification tree model is to reconstruct it without the selected variables and see which other variables are selected in their place.

These results suggest a potential use for decision tree approaches, i.e., to identify potentially important interactions that might be replicated in validation samples. In turn, the classification tree approach can identify specific subpopulations for which a specific treatment is especially important. Therefore it may be useful for creating treatment algorithms. For instance, the decision tree results showed that treatment predicted outcome only among individuals who smoke their first cigarette within 30 min of waking (the highly dependent; Figure 1). This suggests that it is particularly important for more dependent smokers to receive bupropion pharmacotherapy, compared with less dependent smokers. However, it is important to note that treatment might have benefited all smokers somewhat, while still producing effects that differed greatly for one type of smoker versus another (the more and less dependent smoker). Similarly, Figure 2 might suggest that it does not make sense to give pharmacotherapy to persons who are in good/fair/poor health, and who have a household income of less than $35,000. This is because when these individuals are given active medication they are no more successful than placebo controls. However, it is the case that such individuals might have done worse than placebo controls had they not been given medication (although they could not have done much worse). In short, decision tree results may suggest a treatment algorithm, but they by themselves cannot validate such an algorithm.

Finally, several measures from the GUIDE models might serve as important measures in the development of new relapse risk instruments, and such measures might be administered via adaptive testing strategies. That is, predictors might be administered as a function of the respondents' status on previously administered variables (dependence level, marital status, and so on). This might result in efficient earmarking of risk as it manifests in subgroups of smokers, e.g., subgroups based upon life-context or treatment factors, rather than across smokers as a whole.

LIMITATIONS

Some design and analysis factors limit the interpretation of this research. First, while these results are based on two separate clinical trials, these findings have not been replicated. It may be that the identified predictors might not be selected in models in which there are other, similar predictors that better account for the variance in a specific population. In addition, while different predictors were identified at different time points, we did not formally test interactions between predictors and time. Therefore, we cannot conclude that a specific predictor becomes less important over time, just that it no longer is the optimal predictor in a specific model. A second limitation is that both clinical trials were conducted in the Midwest, and there was little representation of any racial/ethnic group beyond White and African American smokers. This may limit the generalizability of these results to other racial/ethnic populations. One analysis factor that limits the generalizability of these findings is that other decision tree models may produce different optimal selection variables. Another limitation is that this research uses the same samples as previous research (Transdisciplinary Tobacco Use Research Center Tobacco Dependence Phenotype Workgroup, 2007; Weiss et al., 2008). While the use of point-prevalence outcome is accepted and quite common (Fiore et al., 2008), the Society for Research on Nicotine and Tobacco workgroup recommended that prolonged abstinence be used as the primary outcome for cessation trials (Hughes et al., 2003). However, use of the dichotomous point-prevalence abstinence outcome is more appropriate with classification tree analysis. Finally, while we have shown that the classification tree analyses yielded different prediction models than did the regression analyses, we did not demonstrate that they produce more accurate predictions. This is because accuracy measures gain meaning only with cross-validation; otherwise, relative accuracy largely reflects the extent to which the models are overfit.[4]

[4] While acknowledging the limitations of accuracy determination in derivation samples, it is of interest to note that in the current research the smaller classification tree models were similar in accuracy to the larger regression models. This is consistent with other research, suggesting the comparable accuracy of the two approaches (Lim et al., 2000; Perlich et al., 2004).

CONCLUSIONS

In summary, the GUIDE classification tree analyses were able to provide efficiently an ordered summary of variables that predict smoking cessation at 1 week, end-of-treatment, and 6 months postquit. These results, in concert with the stepwise logistic regression algorithm, revealed that both tobacco-related and context-related variables are important in predicting initial cessation, as well as long-term cessation and maintenance of abstinence. These results also highlight the complementary nature of decision tree and traditional regression approaches in informing researchers about prediction for small subgroups as well as prediction for the overall population, respectively.

Declaration of Interest

Megan E. Piper, Wei-Yin Loh, Stevens S. Smith, and Sandra J. Japuntich have no potential conflicts of interest to disclose. Timothy B. Baker has served as an investigator on research projects sponsored by pharmaceutical companies including Pfizer, Glaxo Wellcome, Sanofi, and Nabi.

RÉSUMÉ

Esta investigación utilizó el análisis llamado "classification tree analysis" y modelos de regresión logística para identificar factores de riesgo relacionados con la abstinencia a corto y a largo plazo. Se analizaron datos de línea base (baseline) y de resultado (outcome) de dos estudios de investigación clínica para dejar de fumar llevados a cabo desde el 2001 hasta el 2002 en dos áreas urbanas del Medio Oeste. Hubo 928 participantes (53.1% mujeres, 81.8% blancos) con datos completos. Ambos análisis sugieren que el riesgo de recaída es producido por la interacción de los factores de riesgo y que los resultados de cesación temprana y tardía reflejan diferentes factores de vulnerabilidad. Los resultados ilustran el carácter dinámico del riesgo de recaída y sugieren la importancia del modelado eficiente de las interacciones en la predicción de las recaídas.

RESUMEN

Cette recherche a utilisé des analyses de classifications par arbre décisionnel et des modèles de régression logistique pour identifier les facteurs de risques liés à l'abstinence tabagique à court et long terme. Les caractéristiques de base et les résultats sur l'arrêt tabagique provenant de deux études portant sur la désaccoutumance au tabac ont été analysés. Ces études, menées de 2001 à 2002, dans deux regions urbaines du Midwest des Etats-Unis comprenaient 928 participants (53.1% de femmes, 81.8% de blancs) avec des données complètes. Les deux analyses suggèrent que les risques de rechute sont dus à l'interaction de facteurs de risques et que les résultats sur l'arrêt tabagique à court et long terme reflètent différents facteurs de vulnérabilité. Ces résultats illustrent la nature dynamique du risque de rechute et suggèrent l'importance d'une modélisation efficace des interactions dans la prédiction de la rechute.

THE AUTHORS

Megan E. Piper, Ph.D., is an Assistant Professor in the Department of Medicine at the University of Wisconsin School of Medicine and Public Health and a lead researcher at the University of Wisconsin Center for Tobacco Research and Intervention (UW-CTRI). Dr. Piper completed her Ph.D. in Clinical Psychology in 2006 at the University of Wisconsin-Madison. Dr. Piper's research interests include understanding and measuring tobacco dependence, identifying which smoking cessation treatments are most effective and how cessation treatments work. In addition, her research addresses tobacco dependence and cessation treatment effects in specific populations such as women and smokers with psychiatric comorbidities such as depression and anxiety.

Wei-Yin Loh, Ph.D., is Professor, Department of Statistics, University of Wisconsin, Madison. He received his Ph.D. from the University of California, Berkeley, in 1982, and was a visiting junior faculty research fellow at the IBM T.J. Watson Research Center in 1986-87. Dr. Loh is a Fellow of the American Statistical Association and the Institute of Mathematical Statistics. He currently serves as associate editor of the ACM Transactions on Knowledge Discovery in Data. Dr. Loh received the Benjamin Smith Reynolds Award for excellence in teaching in 1999 and the U.S. Army Wilks Award for excellence in statistics research and application in 2007. His main research interest is in decision tree algorithms for statistical modeling. He is the author of the GUIDE algorithm.

Stevens S. Smith, Ph.D., is Associate Professor, Department of Medicine, University of Wisconsin School of Medicine and Public Health. In addition, he has been a clinical investigator at the University of Wisconsin Center for Tobacco Research and Intervention (UW-CTRI) since 1992. After completing his Ph.D. in psychology in 1990 at the University of Wisconsin-Madison, Dr. Smith completed a research fellowship at the National Institute on Drug Abuse Addiction Research Center (ARC) in Baltimore, Maryland. At the ARC,

he conducted psychological and molecular genetic research on vulnerability to addiction. At UW-CTRI, Dr. Smith has been involved in numerous studies investigating behavioral and medication treatments for nicotine dependence, measurement of nicotine withdrawal and dependence, development of culturally-appropriate smoking cessation treatment for American Indian smokers, and efficacy of telephone tobacco quit lines. In addition to conducting clinical research, Dr. Smith maintains an outpatient clinical practice as a Licensed Psychologist.

Sandra J. Japuntich Dr. Japuntich is a post-doctoral fellow in the Massachusetts General Hospital Tobacco Research and Treatment Center and Mongan Institute for Health Policy, Harvard Medical School. She received her Ph.D. in Clinical Psychology in 2009 from the University of Wisconsin-Madison, working under the supervision of Timothy Baker, Ph.D. at the UW Center for Tobacco Research and Intervention. Her research interests include measurement of nicotine dependence and related processes (e.g., relapse, craving, withdrawal), clinical trials of behavioral and pharmacological interventions for smoking cessation, and broadly disseminable treatments and intervention models for smoking cessation.

Timothy B. Baker, Ph.D., is a Professor of Medicine in the University of Wisconsin School of Medicine and Public Health. His principal research goals are to increase understanding of the motivational bases of addictive disorders and to develop and evaluate treatments for such disorders. He is also highly interested in developing and using technological advances to deliver effective treatments for addictive disorders and cancer. For many years he was a faculty member and Director of Clinical Training in the Department of Psychology at the University of Wisconsin. Dr. Baker has served as the Editor of the *Journal of Abnormal Psychology*, is the Principal Investigator of the University of Wisconsin Transdisciplinary Tobacco Use Research Center award (NIDA), has contributed chapters to multiple *Reports of the Surgeon General*, and is the recipient of the James McKeen Cattell Award from the Association for Psychological Science.

REFERENCES

Alterman, A. I., Gariti, P., Cook, T. G., & Cnaan, A. (1999). Nicodermal patch adherence and its correlates. *Drug and Alcohol Dependence, 53*(2), 159–165.

American Psychiatric Association. (1994). *Diagnostice and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

Atkinson, E. J., & Therneau, T. M. (2000). An introduction to recursive partitioning using the RPART routines. Technical report #61, Division of Biomedical Statistics and Informatics, Mayo Clinic, Rochester, MN.

Barbeau, E. M., Krieger, N., & Soobader, M. J. (2004). Working class matters: Socioeconomic disadvantage, race/ethnicity, gender, and smoking in NHIS 2000. *American Journal of Public Health, 94*(2), 269–278.

Breiman, L., Friedman, J., Olshen, R., & Stone, C. (1984). *Classification and regression trees*. Pacific Groce, CA: Wadsworth.

Campbell, I. A., Prescott, R. J., & Tjeder-Burton, S. M. (1996). Transdermal nicotine plus support in patients attending hospital with smoking-related diseases: A placebo-controlled study. *Respiratory Medicine, 90*(1), 47–51.

Chandola, T., Head, J., & Bartley, M. (2004). Socio-demographic predictors of quitting smoking: How important are household factors? *Addiction, 99*(6), 770–777.

De Vogli, R., & Santinello, M. (2005). Unemployment and smoking: Does psychosocial stress matter? *Tobacco Control, 14*(6), 389–395.

Derby, C. A., Lasater, T. M., Vass, K., Gonzalez, S., & Carleton, R. A. (1994). Characteristics of smokers who attempt to quit and of those who recently succeeded. *American Journal of Preventive Medicine, 10*(6), 327–334.

Eisinger, R. A. (1971). Psychosocial predictors of smoking recidivism. *Journal of Health and Social Behavior, 12*(4), 355–362.

Fagerström, K. O., & Schneider, N. G. (1989). Measuring nicotine dependence: A review of the Fagerström Tolerance Questionnaire. *Journal of Behavioral Medicine, 12*(2), 159–182.

Fernandez, E., Schiaffino, A., Borrell, C., Benach, J., Ariza, C., Ramon, J. M., et al. (2006). Social class, education, and smoking cessation: Long-term follow-up of patients treated at a smoking cessation unit. *Nicotine & Tobacco Research, 8*(1), 29–36.

Fiore, M. C., Bailey, W. C., & Cohen, S. J. (2000). *Treating tobacco use and dependence: Clinical practice guideline*. Rockville, MD: US Department of Health and Human Services, US Public Health Service.

Fiore, M. C., Jaen, C. R., Baker, T. B., Bailey, W. C., Benowitz, N., Curry, S. J., et al. (2008). *Treating tobacco use and dependence: 2008 update*. Rockville, MD: US Department of Health and Human Services, US Public Health Service.

Garvey, A. J., Bliss, R. E., Hitchcock, J. L., Heinold, J. W., & Rosner, B. (1992). Predictors of smoking relapse among self-quitters: A report from the Normative Aging Study. *Addictive Behaviors, 17*(4), 367–377.

Garvey, A. J., Kinnunen, T., Nordstrom, B. L., Utman, C. H., Doherty, K., Rosner, B., et al. (2000). Effects of nicotine gum dose by level of nicotine dependence. *Nicotine & Tobacco Research, 2*(1), 53–63.

Graham, H., Francis, B., Inskip, H. M., & Harman, J. (2006). Socioeconomic lifecourse influences on women's smoking status in early adulthood. *Journal of Epidemiology and Community Health, 60*(3), 228–233.

Harris, K. J., Okuyemi, K. S., Catley, D., Mayo, M. S., Ge, B., & Ahluwalia, J. S. (2004). Predictors of smoking cessation among African-Americans enrolled in a randomized controlled trial of bupropion. *Preventive Medicine, 38*(4), 498–502.

Heatherton, T. F., Kozlowski, L. T., Frecker, R. C., & Fagerström, K. O. (1991). The Fagerström Test for Nicotine Dependence: A revision of the Fagerström Tolerance Questionnaire. *British Journal of Addiction, 86*(9), 1119–1127.

Homish, G. G., & Leonard, K. E. (2005). Spousal influence on smoking behaviors in a US community sample of newly married couples. *Social Science & Medicine, 61*(12), 2557–2567.

Honjo, K., Tsutsumi, A., Kawachi, I., & Kawakami, N. (2006). What accounts for the relationship between social class and smoking cessation? Results of a path analysis. *Social Science & Medicine, 62*(2), 317–328.

Hughes, J. R., Keely, J. P., Niaura, R. S., Ossip-Klein, D. J., Richmond, R. L., & Swan, G. E. (2003). Measures of abstinence in clinical trials: Issues and recommendations. *Nicotine & Tobacco Research, 5*, 13–25.

Hurt, R. D., Wolter, T. D., Rigotti, N., Hays, J. T., Niaura, R., Durcan, M. J., et al. (2002). Bupropion for pharmacologic relapse prevention to smoking: Predictors of outcome. *Addictive Behaviors, 27*(4), 493–507.

Hyland, A., Li, Q., Bauer, J. E., Giovino, G. A., Steger, C., & Cummings, K. M. (2004). Predictors of cessation in a cohort of current and former smokers followed over 13 years. *Nicotine & Tobacco Research, 6*(Suppl. 3), S363–S369.

Hymowitz, N., Cummings, K. M., Hyland, A., Lynn, W. R., Pechacek, T. F., & Hartwell, T. D. (1997). Predictors of smoking cessation in a cohort of adult smokers followed for five years. *Tobacco Control, 6*(Suppl. 2), S57–S62.

Hymowitz, N., Sexton, M., Ockene, J., & Grandits, G. (1991). Baseline factors associated with smoking cessation and relapse. MRFIT Research Group. *Preventive Medicine, 20*(5), 590–601.

Kawakami, N., Takatsuka, N., Inaba, S., & Shimizu, H. (1999). Development of a screening questionnaire for tobacco/nicotine dependence according to ICD-10, *DSM-III-R*, and *DSM-IV*. *Addictive Behaviors, 24*(2), 155–166.

Killen, J. D., Fortmann, S. P., Kraemer, H. C., Varady, A., & Newman, B. (1992). Who will relapse? Symptoms of nicotine dependence predict long-term relapse after smoking cessation. *Journal of Consulting and Clinical Psychology, 60*(5), 797–801.

Levy, D. T., Romano, E., & Mumford, E. (2005). The relationship of smoking cessation to sociodemographic characteristics, smoking intensity, and tobacco control policies. *Nicotine & Tobacco Research, 7*(3), 387–396.

Lim, T.-S., Loh, W.-Y., & Shih, Y.-S. (2000). A comparison of prediction accuracy, complexity, and training time of 33 old and new classification algorithms. *Machine Learning Journal, 40*, 203–228.

Loh, W.-Y. (2002). Regression trees with unbiased variable selection and interaction detection. *Statistica Sinica, 12*, 361–386.

Loh, W.-Y. (2010). Improving the precision of classification trees. *Annals of Applied Statistics, 3*, 1710–1737.

Loh, W. Y., & Shih, Y. S. (1997). Split selection methods for classification trees. *Statistica Sinica, 7*, 815–840.

Lu, Y., Tong, S., & Oldenburg, B. (2001). Determinants of smoking and cessation during and after pregnancy. *Health Promotion International, 16*(4), 355–365.

Macken, K. M., Wilder, D., Mersy, D. J., & Madlon-Kay, D. J. (1991). Smoking patterns in a low-income urban population. A challenge to smoking cessation efforts. *Journal of Family Practice, 32*(1):93, 95–96.

McCarthy, D. E., Bolt, D. M., & Baker, T. B. (2007). The importance of how: A call for mechanistic research in tobacco dependence treatment studies. In T. Treat, R. I. Bootzin, & T. B. Baker (Eds.), *Psychological clinical science: recent advances in theory and practicer. Integrative perspectives in honor of Richard M. McFall*. New York: Lawrence Erlbaum.

McClure, J. B., Wetter, D. W., de Moor, C., Cinciripini, P. M., & Gritz, E. R. (2002). The relation between alcohol consumption and smoking abstinence: Results from the Working Well Trial. *Addictive Behaviors, 27*(3), 367–379.

McIntyre, R. S., Wilkins, K., Gilmour, H., Soczynska, J. K., Konarksi, J. Z., Miranda, A., et al. (2008). The effect of bipolar I disorder and major depressive disorder on workforce function. *Chronic Diseases Canada, 28*(3), 84–91.

McKee, S. A., Maciejewski, P. K., Falba, T., & Mazure, C. M. (2003). Sex differences in the effects of stressful life events on changes in smoking status. *Addiction, 98*(6), 847–855.

Mermelstein, R., Cohen, S., Lichtenstein, E., Baer, J. S., & Kamarck, T. (1986). Social support and smoking cessation and maintenance. *Journal of Consulting & Clinical Psychology, 54*(4), 447–453.

Miller, A. (2002). *Subset selection in regression* (2nd ed.). Boca Raton, FL: Chapman & Hall/CRC.

Miller, C. E., Ratner, P. A., & Johnson, J. L. (2003). Reducing cardiovascular risk: Identifying predictors of smoking relapse. *Canadian Journal of Cardiovascular Nursing, 13*(3), 7–12.

Morgan, G. D., Ashenberg, Z. S., & Fisher, E. B., Jr. (1988). Abstinence from smoking and the social environment. *Journal of Consulting and Clinical Psychology, 56*(2), 298–301.

Nides, M. A., Rakos, R. F., Gonzales, D., Murray, R. P., Tashkin, D. P., & Bjornson-Benson, W. M. (1995). Predictors of initial smoking cessation and relapse through the first 2 years of the Lung Health Study. *Journal of Consulting and Clinical Psychology, 63*(1), 60–69.

Nollen, N. L., Mayo, M. S., Sanderson Cox, L., Okuyemi, K. S., Choi, W. S., Kaur, H., et al. (2006). Predictors of quitting among African American light smokers enrolled in a randomized, placebo-controlled trial. *Journal of General Internal Medicine, 21*(6), 590–595.

Ockene, J. K., Emmons, K. M., Mermelstein, R. J., Perkins, K. A., Bonollo, D. S., Voorhees, C. C., et al. (2000). Relapse and maintenance issues for smoking cessation. *Health Psychology, 19*(1 Suppl.), 17–31.

Osler, M., & Prescott, E. (1998). Psychosocial, behavioural, and health determinants of successful smoking cessation: A longitudinal study of Danish adults. *Tobacco Control, 7*(3), 262–267.

Osler, M., Prescott, E., Godtfredsen, N., Hein, H. O., & Schnohr, P. (1999). Gender and determinants of smoking cessation: A longitudinal study. *Preventive Medicine, 29*(1), 57–62.

Patten, C. A., Martin, J. E., Calfas, K. J., Lento, J., & Wolter, T. D. (2001). Behavioral treatment for smokers with a history of alcoholism: Predictors of successful outcome. *Journal of Consulting and Clinical Psychology, 69*(5), 796–801.

Perkins, K. A. (2001). Smoking cessation in women. Special considerations. *CNS Drugs, 15*(5), 391–411.

Perlich, C., Provost, F., & Simonoff, J. S. (2003). Tree induction vs. logistic regression: A learning curve analysis. *Journal of Machine Learning Research, 4*, 211–255.

Piasecki, T. M., Fiore, M. C., McCarthy, D. E., & Baker, T. B. (2002). Have we lost our way? The need for dynamic formulations of smoking relapse proneness. *Addiction, 97*(9), 1093–1108.

Piper, M. E., Federman, E. B., McCarthy, D. E., Bolt, D. M., Smith, S. S., Fiore, M. C., et al. (2007). Efficacy of bupropion alone and in combination with nicotine gum. *Nicotine & Tobacco Research, 9*(9), 947–954.

Piper, M. E., McCarthy, D. E., & Baker, T. B. (2006). Assessing tobacco dependence: A guide to measure evaluation and selection. *Nicotine & Tobacco Research, 8*(3), 339–351.

Piper, M. E., Piasecki, T. M., Federman, E. B., Bolt, D. M., Smith, S. S., Fiore, M. C., et al. (2004). A multiple motives approach to tobacco dependence: The Wisconsin Inventory of Smoking

Dependence Motives (WISDM-68). *Journal of Consulting and Clinical Psychology, 72*(2), 139–154.

Radloff, L. S. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement, 1*, 385–401.

Ridgeway, P., & Rapp, C. (1998). *The active ingredients in achieving competitive employment for people with psychiatric disabilities: A research synthesis*. Lawrence, KS: Commission on Mental Health and Developmental Disabilities.

Selzer, M. L., Vinokur, A., & van Rooijen, L. (1975). A self-administered Short Michigan Alcoholism Screening Test (SMAST). *Journal of Studies on Alcohol, 36*(1), 117–126.

Shields, M. (2005). The journey to quitting smoking. *Health Reports, 16*(3), 19–36.

Shiffman, S. (1993). Smoking cessation treatment: Any progress? *Journal of Consulting and Clinical Psychology, 61*(5), 718–722.

Shiffman, S., Waters, A., & Hickcox, M. (2004). The Nicotine Dependence Syndrome Scale: A multidimensional measure of nicotine dependence. *Nicotine & Tobacco Research, 6*(2), 327–348.

Siahpush, M., Heller, G., & Singh, G. (2005). Lower levels of occupation, income and education are strongly associated with a longer smoking duration: Multivariate results from the 2001 Australian National Drug Strategy Survey. *Public Health, 119*(12), 1105–1110.

Simonoff, J. S. (2003). *Analyzing categorical data*. New York: Springer.

Sorensen, G., Gupta, P. C., & Pednekar, M. S. (2005). Social disparities in tobacco use in Mumbai, India: The roles of occupation, education, and gender. *American Journal of Public Health, 95*(6), 1003–1008.

Spitzer, R. L., Williams, J. B., Kroenke, K., Linzer, M., deGruy, F. V., III, & Hahn, S. R., et al. (1994). Utility of a new procedure for diagnosing mental disorders in primary care. The PRIME-MD 1000 study. *Journal of the American Medical Association, 272*(22), 1749–1756.

Swan, G. E., Jack, L. M., Curry, S., Chorost, M., Javitz, H., McAfee, T., et al. (2003). Bupropion SR and counseling for smoking cessation in actual practice: Predictors of outcome. *Nicotine & Tobacco Research, 6*, 911–921.

Swan, G. E., Jack, L. M., Javitz, H. S., McAfee, T., & McClure, J. B. (2008). Predictors of 12-month outcome in smokers who received bupropion sustained-release for smoking cessation. *CNS Drugs, 22*, 239–256.

Swan, G. E., Javitz, H. S., Jack, L. M., Curry, S. J., & McAfee, T. (2004). Heterogeneity in 12-month outcome among female and male smokers. *Addiction, 99*(2), 237–250.

Transdisciplinary Tobacco Use Research Center (TTURC) Tobacco Dependence Phenotype Workgroup, Baker, T. B., Piper, M. E., McCarthy, D. E., Bolt, D. M., Smith, S. S., Kim, S.-Y., Colby, S., Conti, D., Viovino, G. A., Hatsukami, D., Hyland, A., Krishnan-Sarin, S., Niaura, R., Perkins, K. A., & Toll, B. A. (2007). Implications for nicotine dependence. *Nicotine & Tobacco Research*, 9, Supplement 4, S555–S570.

Uhl, G. R., Liu, Q. R., Drgon, T., Johnson, C., Walther, D., & Rose, J. E. (2007). Molecular genetics of nicotine dependence and abstinence: Whole genome association using 520,000 SNPs. *BMC Genetics, 8*, 10.

Weiss, R. B., Baker, T. B., Cannon, D. S., von Niederhausern, A., Dunn, D. M., Matsunami, N., et al. (2008). A candidate gene approach identifies the CHRNA5-A3-B4 region as a risk factor for age-dependent nicotine addiction. *PLoS Genetics, 4*(7), e1000125.

Westman, E. C., Behm, F. M., Simel, D. L., & Rose, J. E. (1997). Smoking behavior on the first day of a quit attempt predicts long-term abstinence. *Archives of Internal Medicine, 157*(3), 333–340.

Wetter, D. W., Cofta-Gunn, L., Irvin, J. E., Fouladi, R. T., Wright, K., Daza, P., et al. (2005). What accounts for the association of education and smoking cessation? *Preventive Medicine, 40*(4), 452–460.

Wetter, D. W., Kenford, S. L., Smith, S. S., Fiore, M. C., Jorenby, D. E., & Baker, T. B. (1999). Gender differences in smoking cessation. *Journal of Consulting and Clinical Psychology, 67*(4), 555–562.

Zhang, P. (1992). Inference after variable selection in linear regression models. *Biometrica, 79*, 741–746.

APPENDIX. ALL ITEMS CONSIDERED IN THE PREDICTION ANALYSES.

1. How old were you the first time you smoked a cigarette, even one or two puffs?
2. How old were you when you first started smoking daily/every day?
3. What is the total number of years you have smoked daily? Do not include any time you stayed off cigarettes for at least 6 months or longer.
4. Do you currently live with a spouse or partner who smokes cigarettes?
5. Do any of the other people who you currently live with, other than a partner or spouse, smoke cigarettes?
6. How many of your friends smoke or use tobacco?
7. If someone in your household wants to smoke, does he/she have to leave in order to smoke?
8. Which of these statements best describes your place of work's smoking policy for work areas?
9. How you ever tried to quit smoking?
10. How many times have you tried to quit smoking?
11. How long has it been since you last tried to quit smoking?
12. After you started smoking regularly, what is the longest time you ever went without smoking?
13. If you try to quit smoking within the next 30 days, how likely is it that you will be successful?
14. During the past 30 days, did you have at least one drink of any alcoholic beverages?
15. During the past 30 days, on how many different days did you have at least one drink of any alcoholic beverage?
16. On the days when you drank, about how many drinks did you have on average?
17. Considering all types of alcoholic beverages, on how many different days during the past 30 days did you have at least five drinks?
18. Have you ever been diagnosed with alcoholism, treated for alcoholism, or had significant problems with alcohol?
19. Have you ever been diagnosed with a drug use disorder, treated for drug use, or had significant problems related to your drug use?
20. Gender.
21. Are you Hispanic/Latino/Latina?

22. Race (White, non-White).
23. Marital status (married/living with a partner, not married).
24. Education.
25. Employment.
26. What is your annual household income from all sources?
27. Would you say that in general your health is—?
28. Have you ever been diagnosed with depression, treated for depression, or had significant problems with depression?
29. FTND (six items).
30. TDS (10 items).
31. CES-D total.
32. DADC Relapse Subscale.
33. DADC Withdrawal Subscale.
34. DADC Self-Administration Subscale.
35. Meets the *DSM* criteria for alcohol abuse.
36. BMI.
37. Active versus placebo bupropion.
38. Negative Affect Scale—Positive Affect Negative Affect Schedule (PANAS).
39. Positive Affect Scale—PANAS.
40. Total MAST score.
41. WISDM Affiliative Attachment Subscale (mean).
42. WISDM Automaticity Subscale (mean).
43. WISDM Control Subscale (mean).
44. WISDM Behavioral Choice/Melioration Subscale (mean).
45. WISDM Cognitive Enhancement Subscale (mean).
46. WISDM Craving Subscale (mean).
47. WISDM Cue Exposure/Associative Processes Subscale (mean).
48. WISDM Negative Reinforcement Subscale (mean).
49. WISDM Positive Reinforcement Subscale (mean).
50. WISDM Social/Environmental Goads Subscale (mean).
51. WISDM Taste/Sensory Properties Subscale (mean).
52. WISDM Tolerance Subscale (mean).
53. WISDM Weight Control Subscale (mean).
54. WISDM total (sum of means).
55. Study (bupropion gum vs. electronic diary).