
title: "Analysis of Smoking Cessation Strategies and Psychological Influences in Major Depressive Disorder"
format: pdf
editor: Wenyu Yang

Abstract

****Background and Aim:**** Individuals with Major Depressive Disorder (MDD) often struggle more to quit smoking because of depression-related factors. Our study aims to explore how baseline psychological and behavioral characteristics influence the effectiveness of cessation interventions, with the goal of enhancing treatment personalization and success rates among smokers with MDD.

****Methods:**** In our study, we analyzed data from a randomized, placebo-controlled 2x2 factorial trial involving 300 adult smokers with current or past Major Depressive Disorder (MDD). Participants were divided between Behavioral Activation for Smoking Cessation (BASC) or Standard Treatment (ST) and received either varenicline or placebo. We used logistic regression to evaluate the interactions between baseline variables and treatment outcomes at the end of treatment, controlling for both behavioral interventions and pharmacotherapy. Data preprocessing steps included imputation for missing values and categorical variable encoding, ensuring accurate and robust statistical.

****Results:**** Our analysis highlights the key predictors of smoking cessation success. We identify Anhedonia as a significant moderator, with higher scores correlating with better outcomes in the BASC group. This underscores the effectiveness of tailored behavioral interventions. Additionally, lower nicotine dependence and higher readiness to quit were predictive of successful cessation, emphasizing the importance of baseline characteristics in treatment planning.

****Conclusions:**** Our findings suggest that personalized interventions, which consider individual psychological profiles and readiness to quit, can enhance cessation rates among smokers with MDD. Identifying baseline moderators and predictors enhances the personalization of smoking cessation interventions. The logistic regression analysis indicates that the critical role of nicotine dependence, as measured by FTCD scores, significantly predicts smoking cessation success. However, the lack of significant effects from pharmacotherapy and the minimal impacts of other baseline variables on moderating treatment outcomes suggest that individual treatment plans focusing on reducing nicotine dependence might be more effective than broader treatment approaches. Future research should focus on integrating these baseline variables into clinical practice to optimize treatment strategies for smokers with MDD.

Introduction

Smoking remains one of the most significant causes of preventable illness and death worldwide. Individuals diagnosed with Major Depressive Disorder (MDD) are particularly susceptible to nicotine addiction, exhibiting higher rates of smoking and greater difficulty in quitting compared to the general population. This increased vulnerability is often attributed to more severe nicotine dependence, intensified withdrawal symptoms, and the interplay between depressive symptoms and smoking behaviors.

Varenicline is a pharmacological agent known to aid smoking cessation by reducing cravings and withdrawal symptoms through partial agonist activity at nicotinic receptors. While its effectiveness is well-established in the general population, its safety and efficacy in individuals with MDD warrant further investigation due to potential neuropsychiatric side effects. Behavioral interventions offer additional support for smoking cessation, especially when tailored to address underlying psychological factors associated with MDD. Behavioral Activation for Smoking Cessation (BASC) is one such intervention that encourages engagement in positive activities to counteract the avoidance and inactivity linked with depression. By targeting the psychological aspects of nicotine dependence, BASC may enhance the likelihood of successful cessation.

A prior randomized controlled trial explored the combined and individual effects of BASC and varenicline in adult smokers with current or past MDD. The results indicated that while varenicline significantly improved abstinence rates, BASC did not demonstrate a superior effect over Standard Behavioral Treatment (ST).

Building on these findings, the present study aims to:

- Examine which baseline variables influence the effectiveness of behavioral treatments (BASC vs. ST) on end-of-treatment smoking abstinence.
- Assess baseline factors that predict successful smoking cessation while controlling for the types of behavioral treatment and pharmacotherapy administered.

By understanding these moderators and predictors, we hope to enhance the personalization of smoking cessation interventions for individuals with MDD, potentially improving outcomes in this high-risk group.

DATA

In this study, we conduct a comprehensive exploratory data analysis (EDA) to examine baseline characteristics and distribution patterns related to smoking cessation treatments among adults diagnosed with Major Depressive Disorder (MDD). To prepare the dataset, missing values were imputed with mean values, and categorical variables were encoded to ensure compatibility with analytical methods. This data preprocessing allowed for a more accurate examination of patterns within the data and laid the groundwork for informed interpretation of each visualization generated.

- ****Summary Table****

{width="352"}

The summary table details the sample's distribution across characteristics like race, age, gender, income, education, nicotine dependence, and motivation to quit. Nearly half of participants are Black or African American, which may influence treatment outcomes due to cultural and socioeconomic factors. Low income and varied education levels highlight potential impacts on adherence and success. Moderate levels of nicotine dependence and depression could complicate cessation, while high motivation to quit (scores mainly between 6 and 8) suggests strong readiness, potentially supporting better treatment outcomes across groups.

{width="324"}

Figure1: Heatmap illustrates the correlation coefficients among variables relevant to smoking cessation, highlighting key relationships like the inverse correlation between nicotine dependence and treatment type.

The heatmap of the correlation matrix highlights key relationships among variables related to smoking cessation, such as age, FTCD scores, daily cigarette consumption, depression scores, and treatment types. A moderate positive correlation between FTCD scores and daily cigarette consumption aligns with expectations, as higher dependence often means more intake. Depression scores, however, show minimal correlation with other variables, suggesting that depression may independently impact cessation outcomes. This independence underscores the need to examine depression as a distinct predictor, as it could influence outcomes in ways unrelated to age or nicotine dependence.

{width="274"}{width="276"}

Figure2: Histogram displaying the distribution of participant ages across training, validation, and test sets, ensuring representativeness across data splits.

Figure3: Histogram of standardized age distributions across data splits, essential for models sensitive to the scale of inputs.

The two age distribution plot shows an even age distribution across training, validation, and test sets, ensuring representativeness and reducing age-related bias. Most participants are middle-aged, relevant since age affects smoking behavior and treatment response. Additionally, scaling age for the analysis confirmed that standardization maintains representativeness, crucial for algorithms requiring normalized inputs like regularized regression. This consistency supports the reliability of model outcomes when applied to new data.

{width="256"}

Figure4: Boxplot comparing baseline depression scores between treatment groups, showing similar distributions to ensure comparability.

The examination of depression scores using the Beck Depression Inventory across two treatment groups, Behavioral Activation for Smoking Cessation (BASC) and Standard Treatment (ST), showed a wide range in severity with outliers in the BASC group. Despite this, median scores were similar between the groups, indicating comparable baseline depression levels. This suggests that any differences in treatment outcomes are unlikely to be influenced by initial depression severity, allowing for a clearer analysis of each treatment's effectiveness.

{width="247"}

Figure5: Density plot of nicotine dependence levels for participants with and without varenicline, demonstrating similar distributions across groups.

The analysis of nicotine dependence using the Fagerström Test for Cigarette Dependence (FTCD) across groups showed that individuals on Varenicline generally had lower to mid-range FTCD scores compared to the placebo group, suggesting different dependence levels might influence response to pharmacotherapy. This highlights the need to further explore how nicotine dependence interacts with treatment, providing key insights for model development and feature selection in predicting smoking cessation outcomes for individuals with MDD.

Methods

In our study, we employed a rigorous model comparison and selection process. We assessed various models using criteria such as Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), ROC curves, AUC values, and interpretability. We also performed k-fold cross-validation (k=5) to ensure generalizability and incorporated regularization techniques like Lasso and Ridge to prevent overfitting and multicollinearity. The selected model balanced statistical significance, goodness-of-fit, predictive accuracy, and clinical interpretability, aligning with our objectives to identify baseline predictors and moderators of cessation success. This approach enhanced the robustness and clinical relevance of our findings.

- **Data Preprocessing AND Variable Selection**

{width="272"}{width="393"}

In analyzing smoking cessation among individuals with MDD, we use multiple logistic regression methods (Relaxed Lasso, Best Subset, Forward Stepwise, and Backward Stepwise) to identify consistent predictors of cessation outcomes. We found that Pharmacotherapy, especially varenicline, emerged as a key factor, with odds ratios around 8–9 across models. This highlights its strong impact on cessation success. Nicotine dependence, measured by baseline FTCD scores and the habit of smoking within 5 minutes of waking, also significantly predicted outcomes, with higher dependence and early smoking behavior linked to lower cessation success. These results emphasize the need for targeted approaches addressing nicotine dependence in cessation programs for MDD populations. The Nicotine Metabolism Ratio (NMR) emerged as a significant predictor in some models, with high odds ratios indicating that individual metabolic differences may impact cessation success, suggesting benefits from tailoring aids based on metabolic profiling. Model evaluation metrics highlighted varied performance; the Relaxed Lasso model showed high sensitivity

but low specificity, leading to more false positives, while Best Subset and Stepwise methods had poor specificity and negative kappa values, indicating limited reliability in distinguishing non-quitters. This restricts their clinical utility, especially in contexts where minimizing false negatives is essential.

To address class imbalance, we applied oversampling to the minority class in the training dataset using the ROSE (Random Over-Sampling Examples) method. This approach generated synthetic samples to balance class proportions, which helps the model better detect individuals who successfully quit smoking—especially relevant given the study’s focus on the unique cessation challenges faced by individuals with MDD. Additionally, we handled missing values by imputing numerical data with the mean, preserving dataset size and variance without introducing bias. Categorical variables were one-hot encoded to ensure compatibility with logistic regression and other machine learning models, enhancing model accuracy and robustness by minimizing biases from missing data and class imbalance.

Result

– **Models Selection Metrics Analysis**

{width="403"}

We applied Logistic Regression, Lasso, Ridge, and Elastic Net models to identify predictors of successful cessation, evaluating each based on accuracy, recall, precision, and F1 score. Logistic Regression outperformed the others, with an accuracy of 0.638 and an F1 score of 0.761, reflecting a strong balance of precision (0.818) and recall (0.711), making it effective in identifying true positives with minimal false positives. Both Lasso and Elastic Net showed similar metrics (accuracy of 0.553, F1 of 0.656), likely due to regularization, which helped avoid overfitting but limited sensitivity. Ridge regression performed slightly lower, with an accuracy of 0.532 and recall of 0.500, possibly due to its less flexible approach to feature collinearity.

– **Analysis of ROC and AUC**

{width="264"}

Figure6: ROC curves comparing the predictive accuracy of Logistic Regression, Lasso, Ridge, and Elastic Net models, with Logistic Regression performing best.

The ROC plot showed that Logistic Regression had the highest AUC at 0.65, indicating its superior ability to distinguish between successful and unsuccessful quitters. Lasso and Ridge models followed with an AUC of 0.61, while Elastic Net underperformed with an AUC of 0.4, suggesting classification challenges. Logistic Regression's higher AUC highlights its suitability for managing the complex behavioral and pharmacological predictors in smoking cessation for individuals with MDD. This model’s performance underscores its potential for tailoring cessation strategies, making it valuable for clinical settings where individual variations in nicotine dependence and pharmacotherapy responses are critical.

– **Analysis of Cross-Validation for Model Tuning**

{width="266"}

{width="265"}

Figure7: Plot showing Lasso model’s cross-validation error across different lambda values, identifying the optimal balance of bias and variance. Figure8: Error plot for Ridge regression indicating the lambda that minimizes prediction error, essential for model optimization.

{width="272"}

{width="263"}

Figure9: Illustration of Elastic Net’s performance across varying lambda values, guiding parameter tuning for model improvement.

Figure10: Plot showing variability in Logistic Regression's performance across cross-validation folds, reflecting sensitivity to data partitions.

The cross-validation error plots for Lasso and Ridge showed clear minima, indicating optimal lambda values that minimize prediction errors. These models effectively managed the bias-variance tradeoff through regularization, enhancing generalizability by shrinking coefficients and preventing overfitting. Elastic Net, combining L1 and L2 penalties, handled correlated predictors well but showed less clear error minima, suggesting a need for more precise parameter tuning. Logistic Regression's AUC values varied across folds, indicating sensitivity to data partitioning and potential issues with stability, which could lead to overfitting or underfitting.

- Final Model Selection

In summary, Logistic Regression is the most effective model among those considered, including Lasso, Ridge, and Elastic Net. It demonstrated strong performance with an accuracy of 0.638, an F1 score of 0.761, and an AUC of 0.65, indicating its ability to accurately distinguish between individuals who successfully quit smoking and those who did not. The model's high recall (0.711) and precision (0.818) rates underscore its suitability for clinical settings where identifying true positive outcomes is critical. Despite its lower stability in cross-validation, where AUC values varied across folds, suggesting sensitivity to data partitioning and potential stability issues, Logistic Regression was chosen due to its overall effectiveness in real-world applications. Its variability across different data partitions indicated potential issues with stability that could lead to overfitting or underfitting in certain scenarios. However, its strengths in model interpretability and alignment with clinical outcomes guided the decision to prioritize it over the other models. While Lasso, Ridge, and Elastic Net helped reduce overfitting, they fell short in predictive accuracy and AUC compared to Logistic Regression. Consequently, Logistic Regression was selected for its accuracy, interpretability, and clinical relevance, supporting personalized smoking cessation strategies tailored to the unique challenges faced by individuals with MDD.

Discussion

{width="593"}

Figure11: 1) Residuals vs. Fitted, confirming no clear patterns that indicate a good model fit; 2) Q-Q plot of Residuals, which assesses the normality of residuals; 3) Scale-Location, used to check for consistent variance across predicted values; and 4) Residuals vs. Leverage, identifying influential data points with Cook's distance contours. Together, these diagnostics verify the assumptions and robustness of the regression model

The goal of our study is to investigate the impact of baseline variables on the effectiveness of behavioral treatments in promoting end-of-treatment (EOT) abstinence among participants, controlling for behavioral treatment modalities and pharmacotherapy usage. Our logistic regression model included a variety of baseline variables as potential moderators and predictors of success in achieving abstinence.

We evaluated the diagnostics of the logistic regression model to ensure the robustness and validity of our findings. The residuals versus fitted values plot did not show any obvious patterns or systematic deviations, suggesting that the model was correctly specified without any apparent violation of the linearity assumption in the logit. This observation is crucial because it indicates that the linear combination of predictors used in the model adequately captures the relationship without transformations or additional non-linear terms. The Q-Q plot of the residuals, although showing slight deviations in the tails, generally supported the assumption that the model adequately captures the distribution of errors, a non-trivial consideration in logistic regression which does not assume normality of residuals but rather that the log-odds are linearly related to the predictors. The scale-location plot also showed that the residuals were spread evenly across the range of predicted values, suggesting homoscedasticity among the residuals. The observations identified in the residuals versus leverage plot, which did not appear to excessively influence the model according to Cook's distance. This suggests that while there are data points with higher leverage, they do not disproportionately impact the

model's coefficients, which supports the generalizability of our model to the population from which the sample was drawn.

We also found that certain characteristics acted as significant predictors and moderators of treatment success. For instance, higher baseline levels of readiness to quit smoking and lower levels of nicotine dependence were associated with a greater likelihood of achieving abstinence. These findings are consistent with previous research suggesting that motivation and lower dependence are key factors in successful cessation interventions. Furthermore, the interaction terms included in the model allowed us to explore the moderating effects of baseline characteristics on the effectiveness of specific behavioral treatments. It shows that certain baseline factors such as psychological profiles and previous smoking behaviors modified how participants responded to different types of behavioral interventions.

{width="258"}

Furthermore, the logistic regression model identifies two types of behavioral treatments: Type 0, which serves as the control group, and Type 1, the experimental group. The odds ratio (OR) for Treatment Type 1 is 0.39 (95% confidence interval [CI]: 0.00, 120), with a p-value of 0.7. This indicates that the effect of Treatment Type 1 on EOT abstinence compared to Type 0 is not statistically significant, and the wide confidence interval suggests a high degree of uncertainty in this estimate, potentially due to small sample sizes or outlier effects within those treatment groups. The model also explores the impact of demographic factors on EOT abstinence. Age has an OR of 0.97 per year increase (95% CI: 0.92–1.03, $p = 0.3$), suggesting a marginal, non-significant effect where older participants are slightly less likely to achieve abstinence.

Among psychological factors, the Fagerström Test for Cigarette Dependence (FTCD) scores significantly predict cessation success. An OR of 0.60 (95% CI: 0.42–0.81, $p = 0.002$) indicates that lower nicotine dependence significantly increases the likelihood of achieving abstinence, highlighting the importance of nicotine dependence as a predictor. However, baseline depression scores measured by the Beck Depression Inventory (BDI) have an OR of 0.97 (95% CI: 0.91–1.03, $p = 0.4$), indicating no significant impact on abstinence outcomes. Pharmacotherapy, represented by those receiving medication (labeled "1" under pharmacotherapy), shows an OR of 0.77 compared to those not receiving medication (95% CI: 0.42–1.41, $p = 0.4$). This result suggests that pharmacotherapy does not significantly affect abstinence rates, indicating that pharmacotherapy does not provide a significant advantage over behavioral treatments alone in this study's context. Interaction terms in the model provide additional insights. The interaction between treatment type and age (OR = 1.06, 95% CI: 0.99–1.13, $p = 0.087$) suggests a non-significant trend where older participants might benefit slightly more from Treatment Type 1. Other interaction terms, including those involving FTCD scores, gender, BDI scores, and readiness, do not show significant effects, with p-values ranging from 0.4 to 0.8.

In conclusion, this logistic regression analysis emphasizes the critical role of nicotine dependence, as measured by FTCD scores, in predicting smoking cessation success. The lack of significant effects from pharmacotherapy and the minimal impacts of other baseline variables on moderating treatment outcomes suggest that individual treatment plans focusing on reducing nicotine dependence might be more effective than broader treatment approaches. These findings support the growing body of evidence favoring personalized medicine approaches in smoking cessation programs, where treatments are tailored based on individual characteristics.