Evaluating Baseline Variables as Moderators and Predictors for Smoking Cessation

Maia Lindner-Liaw

2024-11-07

Abstract

Background: More than 30% of people with major depressive disorder (MDD) are reported as daily smokers, yet smoking cessation treatments for this group are not well studied. The drug varenicline improved cessation rates in one study, but was less effective for MDD smokers, suggesting additional behavioral treatment is needed to help MDD patients quit. A follow up 2x2 factorial design study of both behavioral and drug treatment found no significant improvement in abstinence rates for the combined treatment. The goal of this study is to expand on the analysis of the 2x2 study to determine potential moderators for the behavioral treatment effect among baseline variables, as well as explore possible predictors of abstinence among the baseline variables.

Methods: Missing values were imputed using multiple imputation, and we fit both lasso and L0+L2 regression models for variable selection and regularization on a 75%/25% train-test data split for each imputed data set. 5-fold cross-validation was used to determine the optimal penalty for each lasso and L0+L2 model, and the model coefficients were pooled. The models were evaluated using ROC curves, AUC, and calibration plots.

Results: Menthol cigarette use was a possible moderators for BA, and race, education, income, nicotine dependence, anhedosis, other mental health diagnoses, current MDD, and nicotine metabolism rate were baseline predictors of abstinence. Menthol cigarette use had a moderate effect as a moderator, and the baseline predicters had weak effects on abstinence.

Conclusion: Some baseline variables were found to be possible moderators for behavioral therapy treatment and predictors of abstinence, but the small sample size and limitations of the methods may affect these conclusions.

Introduction

Major depressive disorder (MDD) is one of the most common mental health disorders in the world, and more than 30% of people with MDD have been reported as daily smokers [1]. This group often show greater dependence on tobacco products and experience stronger withdrawal symptoms, smoke more heavily, and experience a greater reward from smoking than people without MDD [2-5]. These smoking trends highlight the importance of effective smoking cessation methods for people with MDD. However, most existing smoking cessation studies have excluded this group, leading to a lack of literature and evidence. Of the five trials where MDD patients were included, two had fewer than 50 participants, and only one assessed the effect of the drug varenicline on smoking cessation [1]. Varenicline works by reducing the amount of reward people get from smoking and by stimulating dopamine release to relieve cravings and other withdrawal symptoms [6]. The study found that the drug did improve smoking cessation among those with MDD [7], but cessation rates are higher under treatment for those without mental health conditions [8]. This suggests that the addition of behavioral treatment may is necessary to obtain comparable cessation results for MDD sufferers [1].

In this project, we will analyze data from a recent study investigation the combination of varenicline and behavioral treatment on smoking cessation in patients with MDD [1]. The study by Hitsman et. al. assessed the effect of varenicline in combination with behavioral activation (BA), which aims to increase the the reward and enjoyment from daily activities. A pilot study found that BA in combination with nicotine patches improved smoking abstinance compared to standard smoking cessation treatment including nicotine patches [9]. The study enrolled 300 daily smokers interested in quitting that also had an MDD diagnosis either currently or in the past. A 2x2 factorial design to compare abstinance in groups receiving standard behavioral treatment (ST) and placebo, BA and placebo, ST and varenicline, and BA and varenicline. Abstinance at 27 weeks was measured after treatment. The study found that BA did not yield greater abstinence compared to ST for both placebo and varenicline, but varenicline did imrove abstinence compared to placebo. In this report, we will use the same data and extend the analysis. In particular, we are interested in determining if baseline covariates in the data are moderators for the behavioral treatment effects. Additionally, we will be assessing whether these baseline covariates are predictors of abstinence, controlling for the drug and behavioral treatments.

EDA

We will begin by conducting exploratory data analysis to inform data pre-processing and modeling. There are 300 observations in the data set, with 25 variables. The covariates are a mix of binary, continuous, and categorical variables. The binary variables are the outcome abstinenct (`abst`), varenicline (`var`), behavioral activation (`BA`), non-hispanic white (`NHW`), Black, Hispanic, smoking within five minutes of waking up (`ftcd.5.mins`), other mental health diagnoses (`otherdiag`), antidepressant medications (`antidepmed`), only menthol cigarette use (`Only.Menthol`), and current MDD diagnosis (`mde_curr`).

Table 1: Demographics by Treatment Type

Characteristic	**BA+Placebo** N = 68	**BA+Var** N = 83	**ST+Placebo** N = 68	**ST+Var** N = 81
Age	51 (14)	50 (13)	50 (11)	49 (13)
Sex (Female)				
0	30 (44%)	39 (47%)	29 (43%)	37 (46%)
1	38 (56%)	44 (53%)	39 (57%)	44 (54%)
Non-hispanic White				
0	44 (65%)	49 (59%)	46 (68%)	56 (69%)
1	24 (35%)	34 (41%)	22 (32%)	25 (31%)
Black				
0	31 (46%)	46 (55%)	28 (41%)	38 (47%)
1	37 (54%)	37 (45%)	40 (59%)	43 (53%)
Hispanic				
0	63 (93%)	79 (95%)	64 (94%)	76 (94%)
1	5 (7.4%)	4 (4.8%)	4 (5.9%)	5 (6.2%)
Income				
1	25 (37%)	30 (37%)	26 (38%)	29 (36%)
2	16 (24%)	17 (21%)	14 (21%)	21 (26%)
3	8 (12%)	13 (16%)	14 (21%)	11 (14%)
4	12 (18%)	12 (15%)	8 (12%)	6 (7.5%)
5	6 (9.0%)	10 (12%)	6 (8.8%)	13 (16%)
Education				
1	1 (1.5%)	0 (0%)	0 (0%)	0 (0%)
2	3 (4.4%)	7 (8.4%)	2 (2.9%)	4 (4.9%)
3	23 (34%)	15 (18%)	11 (16%)	27 (33%)
4	22 (32%)	32 (39%)	38 (56%)	24 (30%)

5	19 (28%)	29 (35%)	17 (25%)	26 (32%)

[1] Mean (SD); n (%)

The continuous variables are age (`age_ps`), cigarettes per day (`cpd_ps`), and nicotine metablism ratio (`NMR`). The the ordinal score variables are Beck Depression Inventory (`bdi_score_w00`), a measure of depression severity, cigarette reward value (`crv_total_pq1`), income (`inc`), education (`edu`), Fagerstrom Test for Cigarette Dependence (`FTCD`) score (`ftcd_score`), substitute (`hedonsum_n_pq1`) and complementary (`hedonsum_y_pq1`) reinforcer scores related to pleasure received from activities, anhedonia (`shaps_score_pq1`), and quit readiness (`readiness`). Because these variable have many levels and the number of total observations is small, it is likely that we have categories with very small n, which can affect modeling. Therefore, we will treat these variables as continuous, excluding income and education. Income and education only have 5 levels each, so we can reasonably keep those categorical.

The `sex_ps` covariates is coded as 1 for male and 2 for female, so we subtract 1 to get a binary indicator for female. All binary and categorical variables are factored to ensure they are not read as continuous variables. The four treatment types used in the original paper were BA and placebo, BA and varenecline, ST and placebo, and ST and varenicline. To that end, we create a new variable called `trt` that summarizes the treatment group for each observation from the existing indicators for behavioral treatment and drug treatment. Table 1 shows the basic demographic information from the study participants, grouped by this new `trt` variable.

The average age across all groups is about 50 and females make up about 55% of the sample as indicated by the `sex_ps` variable, where 1 indicates female. Black subjects make up roughly 45%-60% of the sample, followed by non-hispanic whites (NHW) at 30%-40%, and finally hispanics with less than 8% in each group. Most subjects fall into the lowest income category earning less than $20,000 a year, and the majority have some college or technical school education. Overall, the baseline covariates are balanced across the treatment groups.

Table 2: Measured Variables by Abstinance

Characteristic	**0**	**1**
	N = 236	N = 64
FTCD Score	5 (2)	4 (2)
FTCD 5 Minutes		
0	123 (52%)	39 (61%)
1	113 (48%)	25 (39%)
BDI Score	19 (12)	17 (11)
Cigarettes Per Day	16 (8)	14 (8)
Cigarette Reward Value	7 (4)	7 (4)
Substiture Reinforcer Score	22 (19)	25 (22)
Complementary Reinforcer Score	26 (19)	23 (20)
Anhedonia Score	2 (3)	2 (2)
Other Diagnoses		
0	127 (54%)	40 (63%)
1	109 (46%)	24 (38%)
Antidepressant Meds		
0	171 (72%)	47 (73%)
1	65 (28%)	17 (27%)
Current MDD		
0	114 (48%)	39 (61%)
1	122 (52%)	25 (39%)
NMR	0.35 (0.22)	0.42 (0.26)

Menthol Only		
0	91 (39%)	29 (45%)
1	143 (61%)	35 (55%)
Quit Readiness	7 (1)	7 (1)

[1] Mean (SD); n (%)

Additional variables available in the data set are shown in Table 2, stratified by abstinence status. These include recorded measurements such as level of nicotine dependence (FTCD score), number of cigarettes per day (`cpd_ps`), depression severity score (`bdi_score_w00`), and antidepressant medication status (`antidepmed`). In this table, we do see some differences in these measures between the abstinent and non-abstinent group. In particular, baseline nicotine dependence as measured by FTCD score is higher in the non-abstinent group with small standard error, which makes sense as higher dependence will make quitting more difficult. The `mde_curr` variable, which is the number of subjects currently vs previously suffering from MDD, is also much higher in the non-abstinent group, at 52% compared to 39%. This agrees with existing findings that MDD makes smoking cessation more challenging. In the baseline readiness to quit, scores from 5 to 8 were the most common in both outcome groups.

We can also look at the correlation between covariates to get an idea of which variables are related to each other. Figure 1 visualizes these correlations. We can see that there are some covariates that are highly correlated. The very strong negative correlation between `Black` and `NHW` is because subjects are either one or the other, and there are not many Hispanic subjects. Education and income have a strong positive correlation, which is expected, as higher education generally leads to higher pay. The indicator for smoking only menthol cigarettes is notably correlated with race, showing Blacks are more likely to smoke only menthols compared to non-hispanic whites. FTCD score, smoking within 5 minutes of waking up, and cigarettes per day all have strong positive correlations. Finally, the indicator for currently having MDD vs previously having it (`mde_curr`) also has a strong positive correlation with depression severity (`bdi_score_w00`). Most of these strong correlations are not unexpected, as the covariates has similarities in what they are measuring.

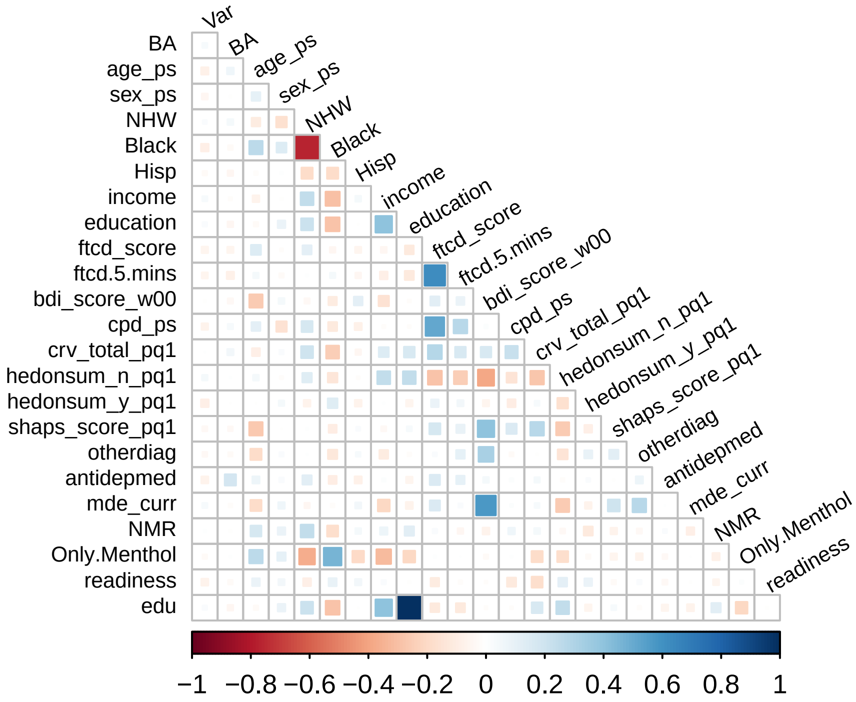


Figure 1: Correlation Between Covariates

Because we have a binary outcome, we can use boxplots to look at the effects of continuous variables on the outcome.

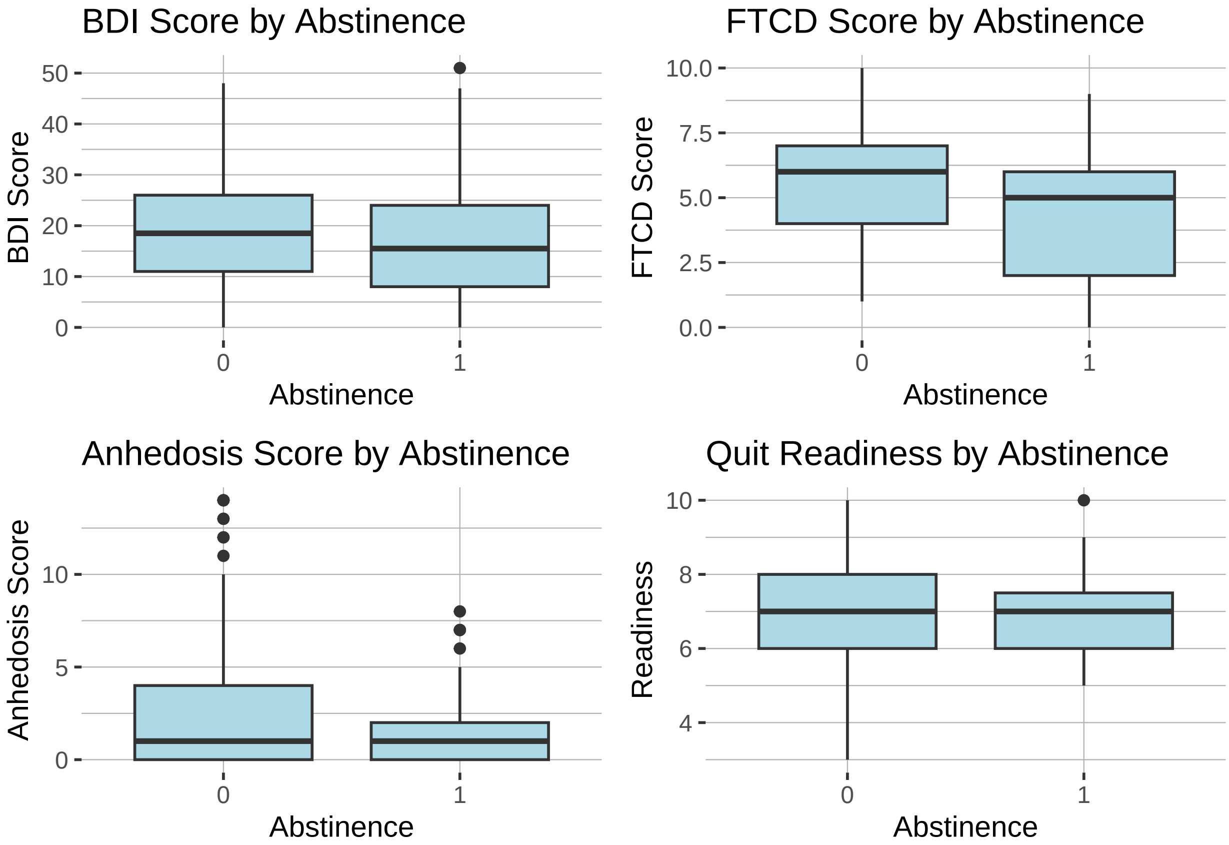


Figure 2: Distribution of Continuous Variables By Abstinence

In Figure 2 we see boxplots for BDI score, Anhedosis, FTCD score, and quit readiness. In the BDI plot, we see that BDI scores are slightly lower among the abstinent group. This suggests that more severe depression makes it harder to quit smoking. We see an even larger difference in FTCD between the outcome groups, indicating higher nicotine dependence is related to a failure to quit. Anhedosis scores are also higher in the non-abstinent group, meaning lack of enjoyment in previously enjoyable activities is also related to more difficulty quitting. While the average quit readiness is the same for the outcome groups, we see that the spread of values is larger in the non-abstinent group, indicating a larger range of commitment to quitting. The other continuous variables had no clear differences between the outcome groups.

Checking missingness in the data is another important consideration in data pre-processing. In this data set, there is low missingness, with only 7 of the 25 variables containing missing values. For each of these variables, the missingness is less than 6%, as seen in Table 3. There do not appear to be any patterns in the missing values either with each other or with other variables. Therefore, it is not unreasonable to assume MCAR missingness. In a large data set, this level of missingness would likely be inconsequential, and complete-case analysis would be unlikely to bias the results. However, given the sample size of 300 in this case, it would be advantageous to retain those observations to make use of every observation in the sample. Therefore, we will perform multiple imputation using the `mice` package to create 5 complete data sets for analysis. Multiple imputation has been shown to have lower bias than other imputation methods, which is why we choose it for this analysis. After imputation, we note there are no longer missing values in any of the imputed data sets.

Table 3: Variables with Missingness

Variable	Proportion
income	0.0100
ftcd_score	0.0033
crv_total_pq1	0.0600
shaps_score_pq1	0.0100
NMR	0.0700
Only.Menthol	0.0067
readiness	0.0567

Modeling Methods

Before we begin model fitting, some considerations are testing the model and properly using our imputation. A common way to ensure data for model validation is to split the data set into a training and test set. We will do so using 75% of the data in the training set, and 25% in the test set, randomly sampling to obtain this split. However, there are only 64 observations with abstinence out of the 300 observations, which is a class imbalance. When sampling randomly, we may obtain a sample where the test or train data has almost all of the events, which would affect model fit. Therefore, we will use stratified sampling so that the proportion of events is similar for the test and train data. We then fit the model on the training set and validate it on the test set. However, after multiple imputation, we actually have 5 full data sets to analysis instead of just 1. Therefore, each of the imputed data sets gets split in the same way, so that the training set for each data set contains the same observations. We fit the model on each of these training sets and then pool the coefficient estimates at the end by taking the mean. We then validate the pooled coefficients on the test set.

For our analysis, the outcome of interest is the binary indicator for smoking abstinence `abst`. This means that we need to use a logistic regression model to fit this outcome. More specifically, we are interested in moderators of the effects of the behavioral treatment and baseline predictors of abstinance. This means that we are interested in any significant main effects and interactions between baseline covariates and the behavioral treatment variable. The upper limit of possible variables in the model is then quite large, which makes variable selection important if we are to obtain a parsimonious model. Regularized regression, such as best subset, lasso, and ridge regression are all methods that can assist in variable selection without the need for an additional algorithm, step or otherwise.

Best subset regression uses the L0 penalty, which puts a constraint on the number of non-zero coefficients. Lasso uses the the L1 penalty, which puts a constraint on the sum of the absolute values of the coefficients. Ridge uses the L2 penalty, which puts a constraint on the sum of the squared coefficients. In each of these cases, the penalty term is controlled by a lambda value, which increases the penalty as it gets bigger. As the penalty increases, best subset regression will remove variables from the model, lasso will snap coefficient values to zero, and ridge will force coefficients closer to zero, but they will never reach zero. Because a parsimonious model will be most helpful in assessing significant predictors and moderators, we want to use method that will actually reduce the number of coefficients, which rules out ridge regression. However, regularization could be important in this case because it prevents overfitting and reduces the variance of the model. Therefore, we can also use a model using the L0+L2 penalty that combines the effects of best subset and ridge regression. When using L0+L2 regression, we get the variable selection of of best subset but also the regularization of ridge regression. We fill fit both a lasso and an L0+L2 regression to the data and compare the models.

An important step in fitting the models is selecting the optimal lambda values to use. To do so, we will use 5-fold cross-validation to determine the lambda value that minimizes the error of the model. We then refit the models to the training set using these optimal lambdas. We will include interactions between behavioral treatment and all other covariates, which will allow the model to force all non-significant interactions to zero. As previously stated, we will perform this procedure on each of the 5 imputed data sets and then average the coefficients from each of the 5 models.

To evaluate the model, we will assess its performance on both the training data and the test data. We will assess and visualize its accuracy and discrimination using AUC and a ROC curve. We will also evaluate the model's calibration by plotting the predicted and observed probabilities for the test data. The calibration plot will also include the ideal line where the observed and predicted values are equal to better compare the performance of the model.

Modeling Results

Tables 4 and 5 show the pooled coefficients and corresponding odds ratios from each of the 5 models for lasso and L0+L2 regression. Counting the levels from the categorical covariates and their interactions, there were 56 possible coefficients for the model, which the lasso regression shrunk to 10, excluding the intercept term. There is 1 interaction remaining in the model: the interaction between behavioral treatment and consuming only menthol cigarettes. This means that smoking only menthol cigarettes is a potential moderator for the effects of the behavioral treatment. The remaining main effects in the model are predictors of abstinence. The L0+L2 regression actually dropped all of the coefficients leaving only the intercept. While this is unexpected, we know that the penalty in best subset regression is stronger than in lasso, which often results in a more parsimonious model. Based on the L0+L2 model, there are no moderators or predictors of abstinence.

Table 4: Pooled Coefficient Estimates From Lasso

	coef	OR
(Intercept)	-1.3551	0.2579
Var1	0.5908	1.8055
NHW1	0.0617	1.0637
income2	-0.1061	0.8993
edu4	-0.0088	0.9912
ftcd_score	-0.0512	0.9501
shaps_score_pq1	-0.0016	0.9984
otherdiag1	-0.0272	0.9732
mde_curr1	-0.0166	0.9835
NMR	0.1084	1.1144
BA1:Only.Menthol1	-0.1886	0.8281

Table 5: Pooled Coefficient Estimates From L0+L2

	coef	OR
Intercept	1.2133	3.3647

Figures 3 and 4 show the ROC curves for the fitted values on both the test and the training data for each model. For lasso, we see that the model performs better than 45 degree representing random prediction for both the test and train data. The AUC on the training data is 0.7302 and on the test data is 0.8443. For L0+L2, the curve falls exactly on the 45 degree line with an AUC of 0.5. This is what we would expect of a model with only an intercept; its discrimination is no better than random chance.

Figure 5 show the calibration of the lasso model on the test data. The red line is the ideal line where the observed and predicted values are the same. The blue line shows calibration with linear smoothing, and the black line shows the calibration with `loess` smoothing, which allows more flexibility. The grey area represents the confidence interval for the `loess` smoothing. Both of the smoothed lines are pretty far off from the ideal, indicating poor calibration of the model. We do not need to create a calibration plot for the L0+L2 regression because it will only ever give one probability estimate.

Figure 2. Lasso Model ROC

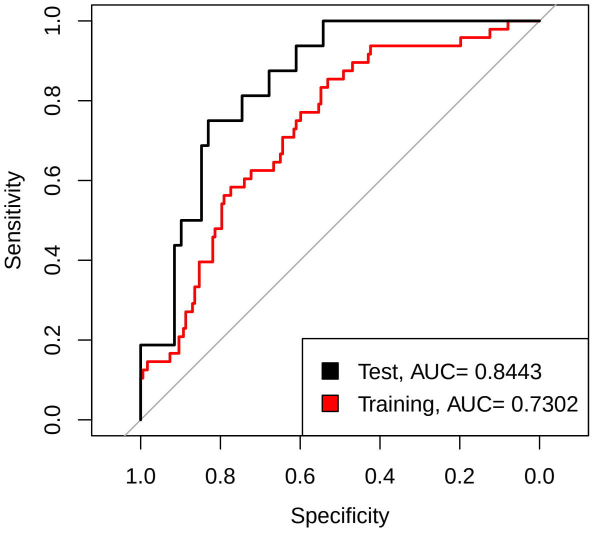


Figure 3. L0+L2 Model ROC

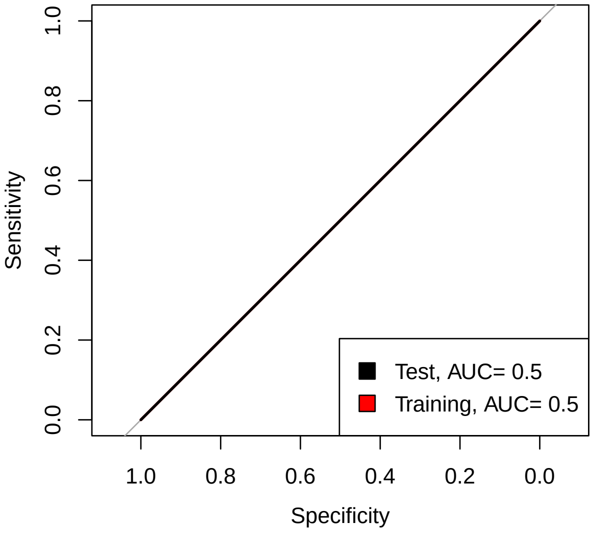
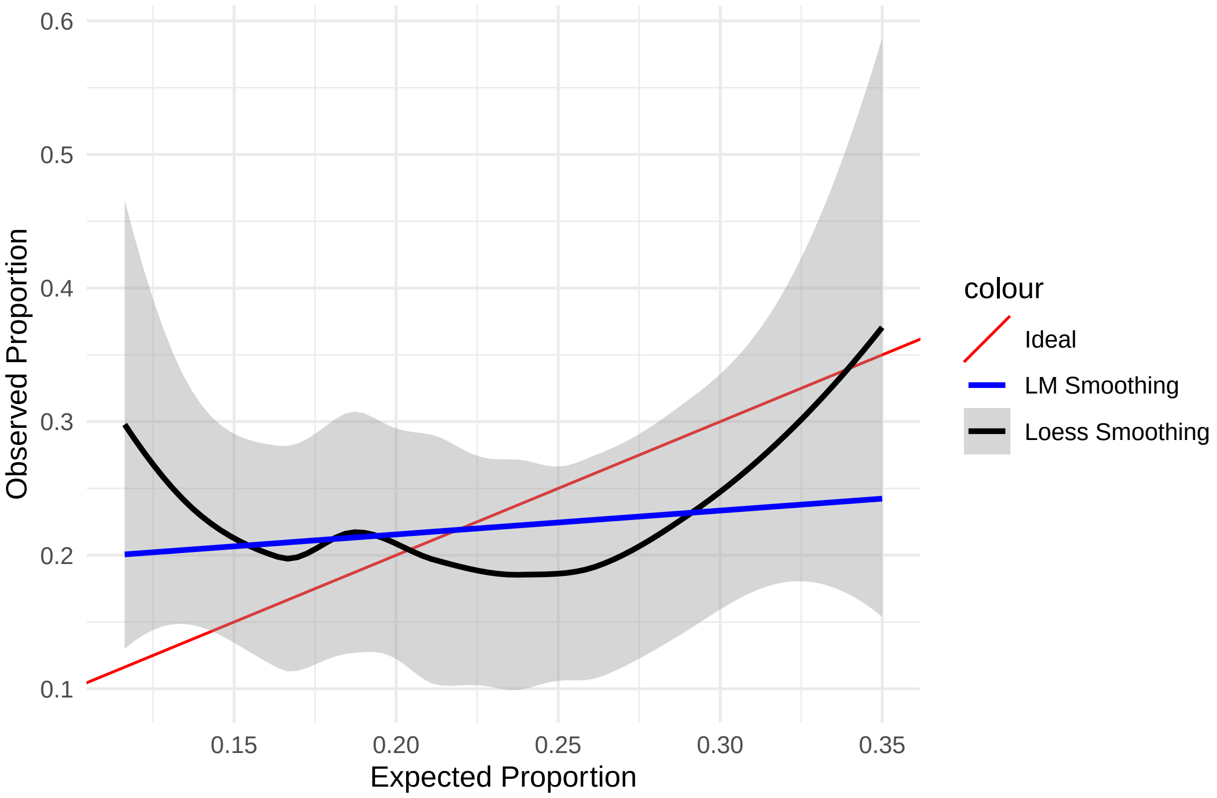


Figure 4. Lasso Calibration Plot on Test Data



Discussion

In our results, we saw that the L0+L2 regression model removed all of the variables from the model, leaving only the intercept term. As previously mentioned, best subset has the strongest variable selection penalty, which is likely what caused this to happen. Therefore, an L0+L2 regression is not appropriate to answer the aims of this project. The goal of using L0+L2 regression was to combine variable selection with regularization because regularization can prevent model overfitting and reduce model variance. Because we were not able to really see the effect of regularization, future work could compare the performance of the lasso model to a relaxed lasso model, which also incorporates regularization.

We will continue with the lasso model to draw our conclusions. In the lasso model, we saw that the penalty shrunk the model from a possible 56 terms down to 11 total terms, indicating a successful variable selection procedure for a parsimonious model. We found that smoking only menthol cigarettes is a potential moderator for the behavioral treatment effect. The interaction term had a negative coefficient, indicating that the behavioral treatment was less effective for only menthol smokers. This aligns with existing evidence that menthol cigarettes are more addictive than normal cigarettes. Although this is a moderators according to the model, it is also important to note the magnitude of the coefficient, which is small. Because the interaction coefficient is so small, although it was significant enough to include in the model, it may not be large enough to be clinically significant in practice. However, it provides a starting point for further investigation into moderators for the treatment effect.

We can interpret the other coefficients as odds ratios after exponentiation. The baseline covariates remaining in the model are non-hispanic white, income at level 2, education at level 4, FTCD score, anhedosis, current MDD, NMR, other mental health diagnoses, as well as the varenicline treatment indicator. The odds of abstinance are 1.8 times higher for a person on varenicline compared to a person not on varenicline, controlling for the other variables left in the model. Being non-hispanic white increases the odds of quitting by 1.06 times, controlling for the other variables. Having an income between 20 and 35 thousand decreases the odds of quitting by 0.90 times, controlling for the other variables. Having some college or technical school education decreases the odds of quitting by 0.99 times, controlling for the other variables. For a unit increase in FTCD score, the odds of abstinence are 0.95 times lower, controlling for the other variables. The other coefficients all have similar interpretations.

Based on this model, varenicline, being white, and higher NMR results in improved odds of quitting, whereas having the second lowest income level, some college or technical school education, higher FTCD score, higher anhedosis score, other mental health diagnoses, and currently having MDD, decrease the odds of abstinence. Most of these are not surprising, although we may have expected something different for education. Education at level 4 is the second highest education level, and it is reasonable to think that more education would make people understand more strongly why they should quit. On the other hand, level 4 indicates incomplete higher education, which means that these people dropped out or failed to graduate. It is not unlikely that MDD is a factor in why a person didn't graduate, and we have seen that MDD decreases the likelihood of quitting smoking. As with the interaction term, some of the main effect coefficients are very small resulting in odds ratios very close to one. Although they were included in the model, they aren't necessarily clinically significant.

Note that the main effect for the behavior treatment was dropped from the model. We typically always want to include main effects in a model when we include their interactions, which is not the case here. Because we don't know the effect of the behavioral treatment outside of the interactions, our conclusions on potential moderators may not be true. Additionally, this means that our coefficient interpretations did not control for the behavior treatment, which could affect their significance as predictors. Therefore, a clear next step would be to refit the model excluding the behavioral treatment term from the penalty, which will always keep it in the model.

In terms of model performance, the ROC curves and AUC values are relatively large, indicating decent discrimination and accuracy on both the training data and the test data. The calibration plot also shows that the smoothed curves roughly follow the ideal line, indicating no glaring lack of model calibration.

Other than the aforementioned exclusion of the behavioral treatment term from the model, another limitation is the sample size of the data. Because we only have 300 total observations, and only 75% are used for model training, the results are likely to be dependent on the seed choice for the random sampling. Not only could this affect the coefficients included in the model and their magnitude, it could also affect the model performance and evaluation measures, for better or worse. To address this issue, future work could include additional cross-validation or bootstrap sampling methods in the model fitting to get a better idea of the model's true performance and generalizability.

The pooling of the coefficients across the 5 imputed data sets may also be a limitation. The models fit to each of the 5 data sets did not always include the same variables, so some zeros were included in the mean calculations. This could be a source of bias, and it is unclear whether this was the best way to handle this. Another possibility would be to have excluded those zero values, which reduces the denominator of the mean by 1, although that may upweight the pooled coefficient too much. This is an area requiring further investigation and literature review to determine what the proper pooling method would be.

Conclusion

Our modeling results found that smoking only menthol cigarettes is a possible moderator for the effect of behavioral treatment for smoking cessation in patients with depression. Additionally, varenicline, race, income, education, nicotine dependence, anhedosis, NMR, other mental health diagnoses, and current MDD were predictors of abstinence. However, there are several limitations that could affect these conclusions and require further research.

References

[1] Hitsman B, Papandonatos GD, Gollan JK, Huffman MD, Niaura R, Mohr DC, Veluz-Wilkins AK, Lubitz SF, Hole A, Leone FT, Khan SS, Fox EN, Bauer AM, Wileyto EP, Bastian J, Schnoll RA. Efficacy and safety of combination behavioral activation for smoking cessation and varenicline for treating tobacco dependence among individuals with current or past major depressive disorder: A 2×2 factorial, randomized, placebo-controlled trial. Addiction. 2023 Sep;118(9):1710-1725. doi: 10.1111/add.16209. Epub 2023 May 3. Erratum in: Addiction. 2024 Sep;119(9):1669. doi: 10.1111/add.16609. PMID: 37069490.

[2] Breslau N, Kilbey MM, Andreski P. Nicotine withdrawal symptoms and psychiatric disorders: findings from an epidemiologic study of young adults. Am J Psychiatry. 1992 Apr;149(4):464-9. doi: 10.1176/ajp.149.4.464. PMID: 1554030.

[3] Lyons M, Hitsman B, Xian H, Panizzon MS, Jerskey BA, Santangelo S, Grant MD, Rende R, Eisen S, Eaves L, Tsuang MT. A twin study of smoking, nicotine dependence, and major depression in men. Nicotine Tob Res. 2008 Jan;10(1):97-108. doi: 10.1080/14622200701705332. PMID: 18188750.

[4] Weinberger AH, Desai RA, McKee SA. Nicotine withdrawal in U.S. smokers with current mood, anxiety, alcohol use, and substance use disorders. Drug Alcohol Depend. 2010 Apr 1;108(1-2):7-12. doi: 10.1016/j.drugalcdep.2009.11.004. Epub 2009 Dec 16. PMID: 20006451; PMCID: PMC2835820.

[5] Spring B, Pingitore R, McChargue DE. Reward value of cigarette smoking for comparably heavy smoking schizophrenic, depressed, and nonpatient smokers. Am J Psychiatry. 2003 Feb;160(2):316-22. doi: 10.1176/appi.ajp.160.2.316. PMID: 12562579.

[6] Anthenelli RM. Varenicline: novel agent to help smokers quit. Current Psychiatry. 2007;6:91-6.

[7] Anthenelli RM, Morris C, Ramey TS, Dubrava SJ, Tsilkos K, Russ C, Yunis C. Effects of varenicline on smoking cessation in adults with stably treated current or past major depression: a randomized trial. Ann Intern Med. 2013 Sep 17;159(6):390-400. doi: 10.7326/0003-4819-159-6-201309170-00005. Erratum in: Ann Intern Med. 2013 Oct 15;159(8):576. PMID: 24042367.

[8] Evins AE, Benowitz NL, West R, Russ C, McRae T, Lawrence D, Krishen A, St Aubin L, Maravic MC, Anthenelli RM. Neuropsychiatric Safety and Efficacy of Varenicline, Bupropion, and Nicotine Patch in Smokers With Psychotic, Anxiety, and Mood Disorders in the EAGLES Trial. J Clin Psychopharmacol. 2019 Mar/Apr;39(2):108-116. doi: 10.1097/JCP.0000000000001015. PMID: 30811371; PMCID: PMC6488024.

[9] MacPherson L, Tull MT, Matusiewicz AK, Rodman S, Strong DR, Kahler CW, Hopko DR, Zvolensky MJ, Brown RA, Lejuez CW. Randomized controlled trial of behavioral activation smoking cessation treatment for smokers with elevated depressive symptoms. J Consult Clin Psychol. 2010 Feb;78(1):55-61. doi: 10.1037/a0017939. PMID: 20099950; PMCID: PMC3108050.

Code Appendix

```r
knitr::opts_chunk$set(echo = FALSE)
knitr::opts_chunk$set(message = FALSE)
knitr::opts_chunk$set(warning = FALSE,
                      fig.pos = "H" ,
                      fig.align = 'center')
library(tidyverse)
library(glmnet)
library(gtsummary)
library(gt)
library(knitr)
library(kableExtra)
library(mice)
library(pROC)
library(corrplot)
library(gridExtra)
library(L0Learn)

# read data
dat<-read.csv("/Users/mlindnerliaw/Desktop/project2.csv")

# rename income variable
names(dat)[10]<-"income"
names(dat)[11]<-"education"

#make sex binary
dat$sex_ps<-dat$sex_ps-1

# create treatment status variable
dat<-dat%>%mutate(trt=case_when(Var==1 & BA==0~"ST+Var",
                                Var==0 & BA==1~"BA+Placebo",
                                Var==1 & BA==1~"BA+Var",
                                TRUE~"ST+Placebo"))

# factor categoricals
dat$income<-factor(dat$income)
dat$edu<-factor(dat$education)
dat$sex_ps<-factor(dat$sex_ps)
dat$abst<-factor(dat$abst)
dat$Var<-factor(dat$Var)
dat$BA<-factor(dat$BA)
dat$NHW<-factor(dat$NHW)
dat$Black<-factor(dat$Black)
dat$Hisp<-factor(dat$Hisp)
dat$ftcd.5.mins<-factor(dat$ftcd.5.mins)
dat$otherdiag<-factor(dat$otherdiag)
dat$antidepmed<-factor(dat$antidepmed)
dat$mde_curr<-factor(dat$mde_curr)
dat$Only.Menthol<-factor(dat$Only.Menthol)


# group columns by demographic and medical details for separate tables
demographic<-names(dat)[5:11]
```

```r
med<-names(dat)[12:25]

# demographic table
t1<-tbl_summary(select(dat, c(all_of(demographic), trt)), by=trt, missing="no",
            statistic = list(all_continuous()~ "{mean} ({sd})",
                             all_categorical()~ "{n} ({p}%)"),
            label= list(age_ps~"Age",
                        sex_ps~"Sex (Female)",
                        income~"Income",
                        education~"Education",
                        Hisp~"Hispanic",
                        NHW~"Non-hispanic White"))%>%
  modify_header(all_stat_cols() ~ "**{level}**  \nN = {n}")%>%
  modify_caption("Demographics by Treatment Type")%>%
  as_kable_extra(booktabs=TRUE, longtable=TRUE)

t1
# measurements table
t2<-tbl_summary(select(dat, c(all_of(med), abst)), by=abst, missing="no",
            statistic = list(all_continuous()~ "{mean} ({sd})",
                             all_categorical()~ "{n} ({p}%)"),
            type = list(readiness ~ "continuous"),
            label = list(ftcd_score~"FTCD Score",
                         ftcd.5.mins~"FTCD 5 Minutes",
                         bdi_score_w00~"BDI Score",
                         cpd_ps~"Cigarettes Per Day",
                         crv_total_pq1~"Cigarette Reward Value",
                         hedonsum_n_pq1~"Substiture Reinforcer Score",
                         hedonsum_y_pq1~"Complementary Reinforcer Score",
                         shaps_score_pq1~"Anhedonia Score",
                         otherdiag~"Other Diagnoses",
                         antidepmed~"Antidepressant Meds",
                         mde_curr~"Current MDD",
                         Only.Menthol="Menthol Only",
                         readiness="Quit Readiness"))%>%
  modify_caption("Measured Variables by Abstinance")%>%
  as_kable_extra(booktabs=TRUE, longtable=TRUE)

t2
# Create correlation matrix
cor_mat <- dat%>%
  select(-c(id, abst, trt)) %>% apply(2, as.numeric)%>%
  cor(use = "complete.obs")

# Plot correlation matrix
corrplot(cor_mat, method = "square", type = "lower", diag = FALSE,mar=c(0,0,2,0),
         tl.cex = 0.7, tl.col = "black", tl.srt = 30)

# create boxplots for continous variables
bdi<-ggplot(data=dat, aes(x=abst, y=bdi_score_w00))+
  geom_boxplot(fill="lightblue")+
  theme(legend.position = "top",
        panel.background = element_rect(fill="white"),
        strip.background =element_rect(fill="grey90", linewidth = .2,
                                       colour = "grey70"),
        panel.grid.major = element_line(color="grey70", linewidth=.2),
        panel.grid.minor = element_line(color="grey70", linewidth=.2),
        legend.key = element_rect(fill = "grey90")
        )+
  labs(title="BDI Score by Abstinence",
       y="BDI Score",
       x="Abstinence")


ftcd<-ggplot(data=dat, aes(x=abst, y=ftcd_score))+
  geom_boxplot(fill="lightblue")+
  theme(legend.position = "top",
        panel.background = element_rect(fill="white"),
        strip.background =element_rect(fill="grey90", linewidth = .2,
                                       colour = "grey70"),
        panel.grid.major = element_line(color="grey70", linewidth=.2),
        panel.grid.minor = element_line(color="grey70", linewidth=.2),
        legend.key = element_rect(fill = "grey90")
        )+
  labs(title="FTCD Score by Abstinence",
       y="FTCD Score",
       x="Abstinence")

anh<-ggplot(data=dat, aes(x=abst, y=shaps_score_pq1))+
  geom_boxplot(fill="lightblue")+
  theme(legend.position = "top",
        panel.background = element_rect(fill="white"),
        strip.background =element_rect(fill="grey90", linewidth = .2,
                                       colour = "grey70"),
        panel.grid.major = element_line(color="grey70", linewidth=.2),
        panel.grid.minor = element_line(color="grey70", linewidth=.2),
        legend.key = element_rect(fill = "grey90")
        )+
  labs(title="Anhedosis Score by Abstinence",
       y="Anhedosis Score",
       x="Abstinence")

read<-ggplot(data=dat, aes(x=abst, y=readiness))+
  geom_boxplot(fill="lightblue")+
  theme(legend.position = "top",
        panel.background = element_rect(fill="white"),
        strip.background =element_rect(fill="grey90", linewidth = .2,
                                       colour = "grey70"),
        panel.grid.major = element_line(color="grey70", linewidth=.2),
        panel.grid.minor = element_line(color="grey70", linewidth=.2),
        legend.key = element_rect(fill = "grey90")
        )+
  labs(title="Quit Readiness by Abstinence",
       y="Readiness",
       x="Abstinence")
```

```r
grid.arrange(bdi, ftcd, anh, read, ncol=2)
# deal with missing data
num_missing<-function(data){
  #' Returns the number of missing values in the given data
  #' @param data, data vector or data frame of any type
  #' @return the number of NA values in the data

  return(length(which(is.na(data))))
}

# table of missingness proportion
prop<-apply(dat, 2, num_missing)/nrow(dat)
# non-zero missingness
prop<-prop[prop>0]%>%round(digits=4)
kable(prop, caption = "Variables with Missingness",
      col.names = c("Variable", "Proportion"))

# run and save multiple imputation
#dat_mice<-mice(select(dat, -c(id, trt)), seed=1)
#saveRDS(dat_mice, "PDA2Mice")

# read in imputed data
dat_mice<-readRDS("/Users/mlindnerliaw/Desktop/PDA2Mice.txt")

set.seed(2550)
# split data into train/validate set
# get events and non-events
event<-dat[dat$abst==1,]
no_event<-dat[dat$abst==0,]

# sample from the strata
train_e<-sample(event$id, .75*nrow(event))
train_no<-sample(no_event$id, .75*nrow(no_event))

train_id<-c(train_e, train_no)

# initialize objects to store lambdas and coefficients
lambda<-c()
lasso_coefficients<-list()
l0l2_coefficients<-list()

for (i in 1:5){
  # extract imputed dataset and split data
  dat<-mice::complete(dat_mice,i)
  train<-dat[train_id,]
  test<-dat[-train_id,]

  # create model matrix and outcome
  x <- model.matrix(abst~BA*(Var+age_ps+sex_ps+NHW+Black+Hisp+income+edu+
                              ftcd_score+ ftcd.5.mins+ bdi_score_w00+cpd_ps+
                              crv_total_pq1+hedonsum_n_pq1+shaps_score_pq1+
                              otherdiag+antidepmed+mde_curr+NMR+Only.Menthol+
                              readiness), data=train)[,-1]
  y <- train$abst

  # use cv to find optimal lambda for lasso
  lasso <- cv.glmnet(x, y, alpha=1, nfolds = 5, family="binomial")
  l<-lasso$lambda.min

    # fit lasso with optimal lambda
  lasso.fit <- glmnet(x, y, alpha=1, lambda=l, family="binomial")

    # extract coefficients and save
  coefs<-coef(lasso.fit)
  lambda<-c(lambda, l)
  lasso_coefficients[i]<-coefs

  # fit cv for optimal L0L2
  cv_l0l2 = L0Learn.cvfit(x, y, loss = "SquaredError",
                        nFolds = 5, penalty = "L0L2")
  cv_res <- sapply(cv_l0l2$cvMeans, as.numeric)
  # extract best penalties
  min_ind <- which(cv_res == min(cv_res), arr.ind = TRUE)
  gamma_min <- cv_l0l2$fit$gamma[[min_ind[2]]]
  lambda_min <- cv_l0l2$fit$lambda[[min_ind[2]]][min_ind[1]]
  # extract coefficients
  cv_coef_l0 <- coef(cv_l0l2, gamma = gamma_min, lambda = lambda_min)
  rownames(cv_coef_l0) <- c("Intercept", colnames(x))
  l0l2_coefficients[i]<-cv_coef_l0


}

# get data frame of coefficients
lasso_coefficients<-as.data.frame(as.matrix(cbind(lasso_coefficients[[1]], lasso_coefficients[[2]], lass
# get mean across imputations
pooled_lasso_coef<-rowMeans(lasso_coefficients)

## same for l0l2
l0l2_coefficients<-as.data.frame(as.matrix(cbind(l0l2_coefficients[[1]], l0l2_coefficients[[2]], l0l2_co
# get mean across imputations
pooled_l0l2_coef<-rowMeans(l0l2_coefficients)


# extract non-zero coefficients
no0coef_lasso<-pooled_lasso_coef[pooled_lasso_coef!=0]

no0coef_l0l2<-pooled_l0l2_coef[pooled_l0l2_coef!=0]

# return coefs and odds ratios
kable(data.frame(coef=no0coef_lasso, OR=exp(no0coef_lasso)), caption="Pooled Coefficient Estimates From 

kable(data.frame(coef=no0coef_l0l2, OR=exp(no0coef_l0l2)), caption="Pooled Coefficient Estimates From L


# predict on training data
```

```
pred_train<-plogis(as.numeric(x%*%pooled_lasso_coef[2:56]+pooled_lasso_coef[1]))

pred_train_l0l2<-plogis(as.numeric(x%*%pooled_l0l2_coef[2:56]+pooled_l0l2_coef[1]))

# create model matrix for test data
x_test <- model.matrix(abst~BA*(Var+age_ps+sex_ps+NHW+Black+Hisp+income+edu+
                                ftcd_score+ ftcd.5.mins+ bdi_score_w00+cpd_ps+
                                crv_total_pq1+hedonsum_n_pq1+shaps_score_pq1+
                                otherdiag+antidepmed+mde_curr+NMR+Only.Menthol+
                                readiness), data=test)[,-1]
# predict on test data
pred_test<-plogis(as.numeric(x_test%*%pooled_lasso_coef[2:56]+pooled_lasso_coef[1]))
pred_test_l0l2<-plogis(as.numeric(x_test%*%pooled_l0l2_coef[2:56]+pooled_l0l2_coef[1]))

# roc curve for training data
roc_train<-roc(dat$abst[train_id], pred_train)
roc_train_l0l2<-roc(dat$abst[train_id], pred_train_l0l2)
roc_test<-roc(dat$abst[-train_id], pred_test)
roc_test_l0l2<-roc(dat$abst[-train_id], pred_test_l0l2)

par(mfrow=(c(1,2)))
# plot roc curves
plot(roc_train, col="red", main="Figure 2. Lasso Model ROC")
plot(roc_test, add=TRUE)
legend("bottomright",
       legend = c(paste("Test, AUC=", as.character(round(roc_test$auc, 4))),
                  paste("Training, AUC=", as.character(round(roc_train$auc, 4)))),
       fill = c("black", "red"))

# plot roc curves
plot(roc_train_l0l2, col="red", main="Figure 3. L0+L2 Model ROC")
plot(roc_test_l0l2, add=TRUE)
legend("bottomright",
       legend = c(paste("Test, AUC=", as.character(round(roc_test_l0l2$auc, 4))),
                  paste("Training, AUC=", as.character(round(roc_train_l0l2$auc, 4)))),
       fill = c("black", "red"))
### calibration plot
num_cuts <- 50

test_calib<-data.frame(prob = pred_test,
                          bin = cut(pred_test, breaks = num_cuts),
                          # converting to numeric from factor makes it 1,2, so
                          # subtract 1 to get binary
                          class = as.numeric(train$abst)-1)




test_calib <- test_calib %>%
             group_by(bin) %>%
             summarize(observed = sum(class)/n(),
                       expected = sum(prob)/n(),
                       se = sqrt(observed * (1-observed) / n()))
```

```
cols <- c("Ideal"="red","Loess Smoothing"="black","LM Smoothing"="blue")
ggplot(test_calib) +
  geom_abline(aes(intercept = 0, slope = 1, color="Ideal")) +
  geom_smooth(aes(x = expected, y = observed, color="Loess Smoothing"), se=TRUE)+
  geom_smooth(aes(x = expected, y = observed, color="LM Smoothing"), se=FALSE, method="lm")+
  scale_color_manual(values=cols)+
  labs(x = "Expected Proportion", y = "Observed Proportion", title="Figure 4. Lasso Calibration Plot on
  theme_minimal()
```