“We hereby declare that the work represented in this report represents our combined individual effort. We understand that we are encouraged to seek advice and guidance on any course material from the professor, teaching assistants and fellow classmates, however, we are responsible, solely, for this written document and the oral presentation of components of this report.”

**Dominick DeMarsico & Sean Vieau 10/21/2024**

**(names) (date)**

**Project Breakdown Total: 50% Dominick 50% Sean**

**Intro: 70% Dominick, 30% Sean**

**Methods: 80% Dominick, 20% Sean**

**Results: 30% Dominick 70% Sean**

**Discussion: 50% Dominick 50% Sean**

**Presentation: 80% Dominick 20% Sean**

**Code: 20% Dominick 80% Sean**

**Introduction:**

Highly Active Antiretroviral Therapy (HAART) is the standard treatment for individuals infected with HIV. There is limited evidence that the use of hard drugs (e.g. heroin, cocaine) inhibits the immune system and facilitates HIV replication. However, there is a lack of clear results of this outcome in humans. We were interested in understanding how the treatment outcomes of HAART differs at two years between subjects who report using hard drugs and those who did not report using hard drugs. Secondarily, we are also interested in determining if this relationship differs based on hard drug use.

Data was collected as a part of the Multicenter AIDS Cohort Study, an ongoing cohort of HIV-1 infected men in 4 major U.S. cities. This is a secondary data analysis that includes 2 years of longitudinal health data starting from the most proximal visit prior to starting HAART treatment. We hypothesize that subjects who currently use hard drugs will have lessened treatment response over 2 years. Additionally, we also hypothesize that those who are in higher adherence groups will have better treatment outcomes regardless of drug use status

**Methods**

RStudio version 4.4.1 was used for both data cleaning and analysis. Data cleaning started by creating a working data frame for our project which only included year 0 through year 2. The data was in the “long form” (i.e. multiple within-patient timepoints reported as separate observations), so our data was transposed such that there was one row per patient. Potential covariates were dropped if they were missing at a rate greater than 20%.

Data was assessed for entry errors and none were found. Outliers were examined using jackknife residuals to assess for points of high leverage and influence in each model. However, our team did not have the proper clinical knowledge to determine true outliers. As viral load and CD4+ T cell count were true biological measures, we decided to keep all observations for our primary outcome variables and confer with clinicians before removing potential outliers.

This study is an ongoing prospective cohort study to understand the effects of both treated and untreated HIV-1 infection in homosexual and bisexual males. However, we will be conducting a secondary analysis to specifically assess the effects of drug use on treatment outcomes in HIV infected men. To do this, a factor was created to allow for analysis of each of the drug-use groups: Current Use (drug use in the most recent study year), Past-Use (drug use anytime within the study except the most recent year), and never use (no drug use whatsoever). We also created variables from the difference of our 4 primary outcome variables (Viral Load, Leukocyte count, Mental Wellbeing, and Physical Wellbeing) between year 2 and baseline to assess changes over the course of the treatment.

Adherence was a covariate that the principal investigator was exceptionally interested in. Adherence was measured as a categorical variable with 4 levels: 100%, 99-95%, 95-75%, and >75%. Due to the extreme similarity between the two highest adherence groups as determined by plotting outcome responses over time, we thought it prudent to split adherence into a binary variable for high (100% - 95%) and low adherence (<95%). This new adherence dummy variable was used to create an interaction term with hard drug use. Effect modification of this potential relationship between adherence and drug use on our primary outcome variables was assessed by including this interaction term into our models.

The first step of analysis was to generate descriptive statistics for our cohort. Table 1 shows descriptive statistics for our primary variables and our covariates of interest. Additionally, the amount of missing data is reported in Table 1. This would help us to determine if there was a need for a prognostic shift in our analysis. However, all primary outcome variables were missing <1% of values, well below the critical level for missingness that requires large scale changes in study interpretation.

Assessment of significance was done initially with a correlation matrix. If the correlation coefficient was approximately moderate (r = ~30% - 60%), we would consider these variables for model selection. We then ran a series of SLRs for each selected covariate to assess possible relationships with our primary outcome variables. If they had a significance of p < 0.10, they would be used in a more complex multiple linear regression model. We would then use both backwards elimination.

Four final models were selected, one for each outcome variable. Our null hypothesis for our primary question was that there would not be a statistically significant difference in any of the 4 primary outcome variables’ change scores between our 3 drug use groups while accounting for adherence (95% CI) . Our alternative hypothesis was that there would be a statistically significant difference in any of the primary outcome variables’ change scores between our 3 drug use groups based on different levels of adherence (95% CI).

**Results**

The total population of the study was 550 participants. Of that total, all were male, HIV+, and taking HAART treatment after baseline. Never drug use had 444 subjects, past drug use had 46 subjects, and current drug use had 60 subjects. While there are large population size discrepancies, all of the groups were sufficiently large (>30) (See Table 1). Primary outcome data was missing in 18 patients for Viral Load, 18 for Leukocyte count, 7 for Mental Wellbeing, and 7 for Physical wellbeing. Income, Triglycerides, LDL, and BMI were all dropped from analysis due to missingness being greater than 20%. We then decided to move forward with a complete case analysis. Figure 1 reveals that there were moderate correlations (r = ~30% - 60%) between at least one of our primary outcome variables and education, frailty phenotype, age, smoking status, depression, and adherence.

Backward elimination using BIC determined the final covariates to be included for viral load were education, hard drug use, adherence, and the interaction between drug use and adherence. There were significant differences in change in log viral load based on hard drug usage and adherence to the treatment regimen, while controlling for education at baseline (F(6, 513)= 9.51), p < 0.0001). Low Adherence led to 3.48 increased log viral load in Current users (p = .04) and 5.81 increased log viral load in past users (p <.001). However, there were no significant differences in viral load for drug users with High Adherence (See Figure 2)

Backwards elimination using BIC determined the covariates to be included for CD4+ T Cell count were frailty phenotype along with adherence, drug use category, and their interaction. There were significant differences in change in CD4+ T Cell count over 2 years based on hard drug usage and adherence to the treatment regimen, while controlling for Frailty Related Phenotype (F(6,512) = 8.42, p < 0.00001). Low Adherence led to a 197.90 cell increase in CD4 in previous users compared to never users (p = .04). Never users and current users did not differ significantly from each other. Current users had a 342.0 cell decrease compared to past users (p = 01). High Adherence led to a 99.23 cell decrease in past users (p = .003) and a 111.16 cell decrease in current users (p <.001) when compared to never users (See Figure 3)

For Mental Wellbeing, backward elimination using BIC determined that the final covariates to include were drug use, adherence, and their interaction. There were significant differences in mental QOL over 2 years based on hard drug usage and adherence to the treatment regimen (F(5,526) = 4.78, p = 0.000278). Low Adherence led to a 15.95 point decrease in mental QoL in previous drug users compared to never drug users (p =.009), and a 23.91 decrease compared to never hard drug users (p = 0.007). There was no difference in mental QOL between current and never hard drug users at low adherence (p = 0.64). At high adherence, there was no difference in mental QOL between previous and never hard drug users (p = 0.48), or between current and never hard drug users (p = 0.20). The difference in mental QOL for current and previous users was borderline significant (p = 0.050) (See Figure 3). While depression was highly correlated to mental QOL, it was determined to be a confounder and a separate analysis run, showing heightened risk of depression for previous users (See Figure 3a).

For Physical wellbeing, backward elimination using BIC selected the final model to include frailty related phenotype along with drug use, adherence, and their interaction. There were significant differences in change in physical QOL over 2 years based on hard drug usage and adherence to the treatment regimen, while controlling for Frailty Related Phenotype (F(6, 525) = 20.67, p < 0.0001). Low Adherence led to a 10.89 point decrease in physical QoL in past drug users compared to never drug users (p =.004). Current users did not differ significantly from never users. High Adherence led to current users having a 3.94 point decrease in physical QoL compared to never users (p = .001), and a 5.35 point decrease when compared to past users (p = .002). There was no difference between previous and never users (p = 0.8). See Figure 5.

Assumptions for a regression were assessed using Q-Q plots, histograms of the residuals, scale-location plots, and scatterplots of the residuals against the fitted-values. All assumptions were met for Leukocyte levels. However, Viral Load, Mental QoL, and Physical QoL have slight violations of normality and homogeneity of variances.

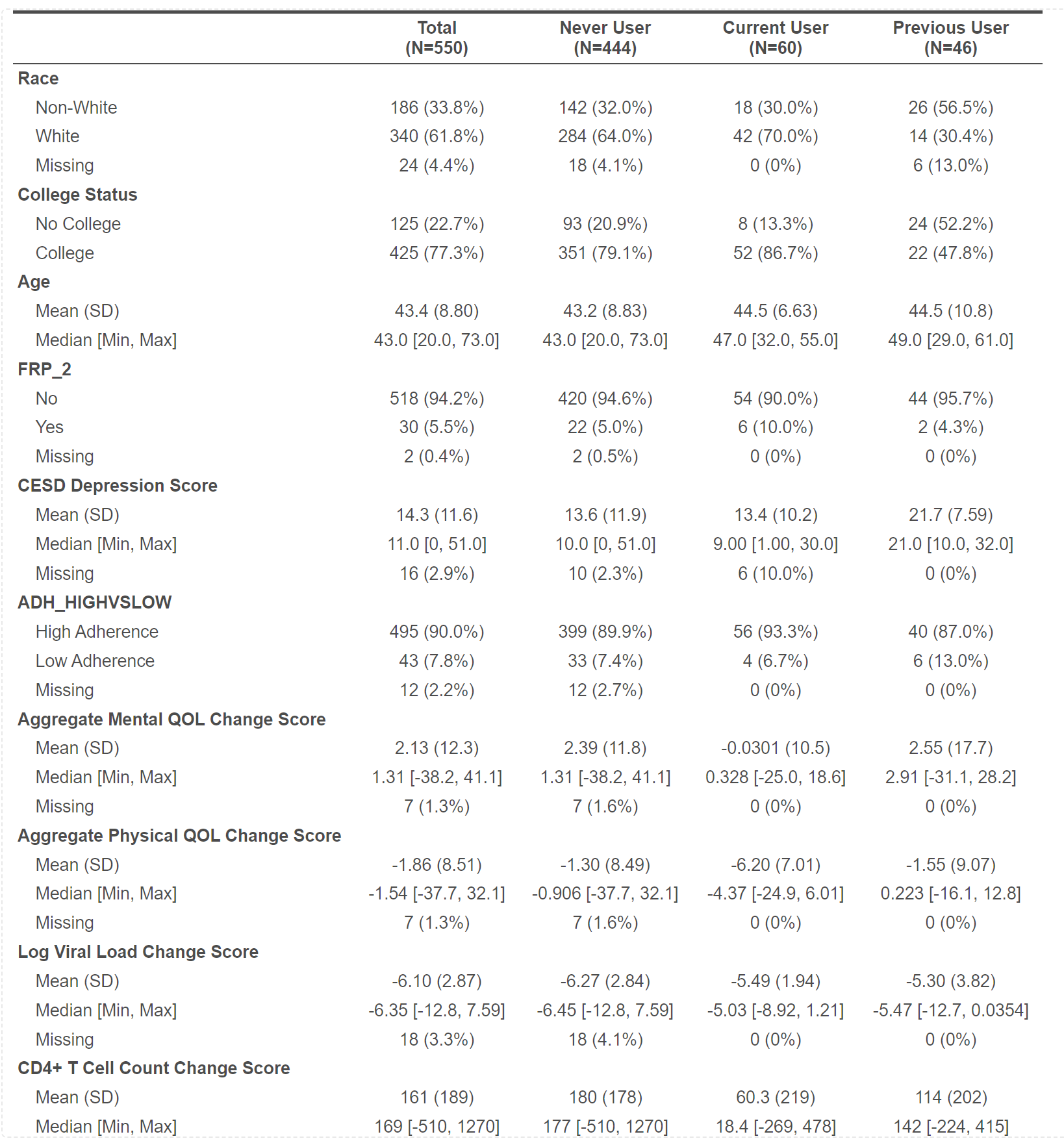
**Discussion**

With regard to our research question, adherence played a major role in affecting the reliability of the models. We have evidence to suggest that the relationship between treatment outcomes and drug use is modified by the interaction of adherence and drug use. Therefore, we have evidence to *reject the null hypothesis*. From our perspective, adherence would act as a confounder if it were not already accounted for in our model. Specifically, low adherence leads to worsened outcomes in current users for viral load, and in previous users for all outcome variables including mental and physical well being. Thus, previous hard drug users are a vulnerable population and appear to be struggling with quitting, and more aggressive treatment strategies should be considered to support them.

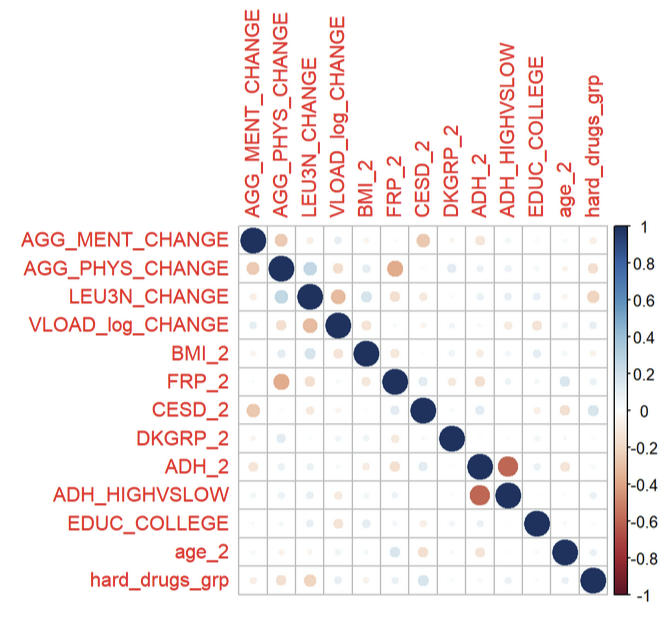
Additionally, only 7.8% of subjects fell into the category of low adherence, with only 4 previous and 6 current drug users. The lack of subjects in this category severely hampers our ability to use an interaction term at all for certain groups. A larger N in this category would allow for more information on the true interaction within these groups.

Another consideration should be physical and mental changes that occur during pathology but can not be explained by treatment. Access to a list of Adverse Events suffered by the subjects would be very helpful to better understand what is causing changes. Additionally, the slight deviation from normality with some of our outcome variables can be attributed to using change scores rather than each individual time point as a unique observation. This method effectively cut our N by two-thirds. We expect that longitudinal analyses allowing for an N closer to the original value of 1600 would alleviate these symptoms for normality.

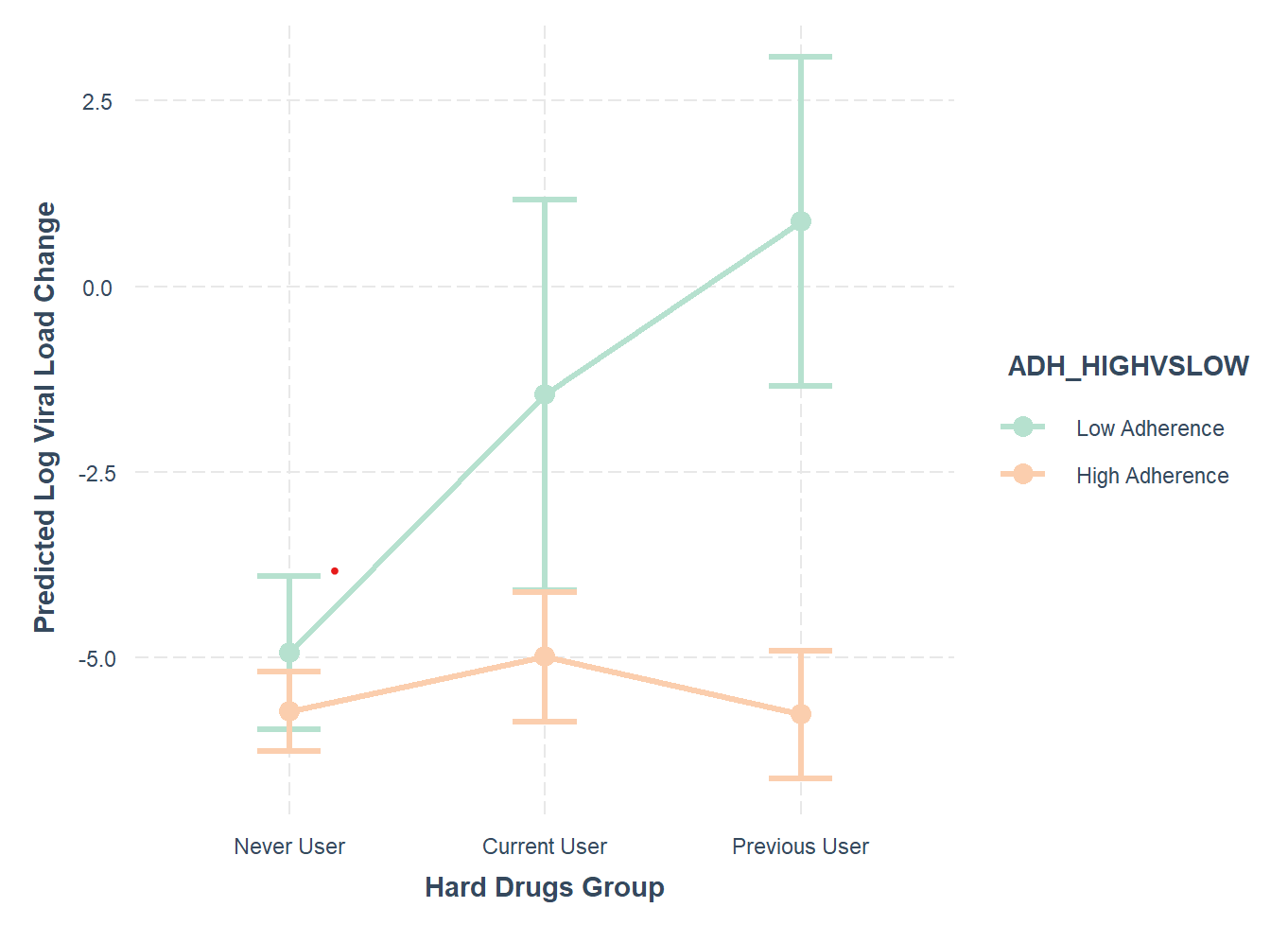
***Table 1.***

******

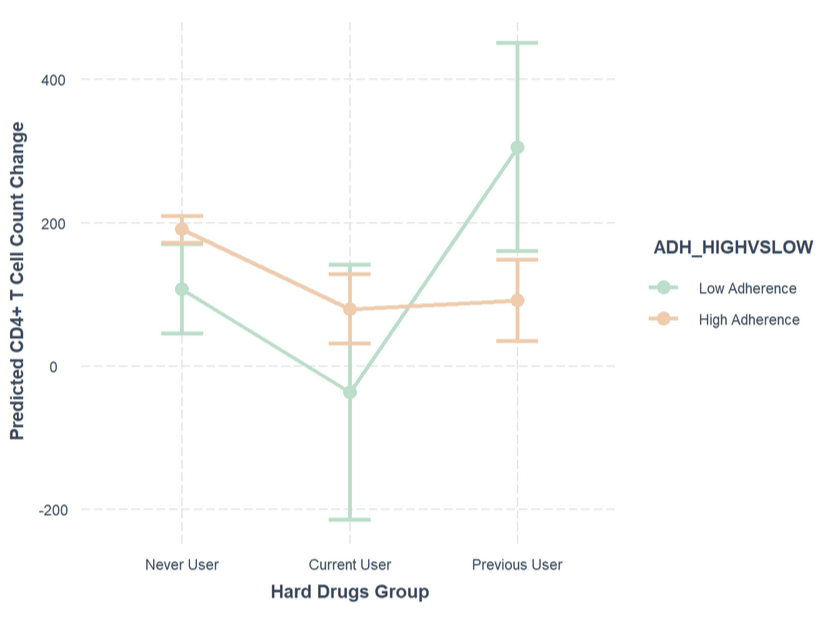
***Figure 1.***

******

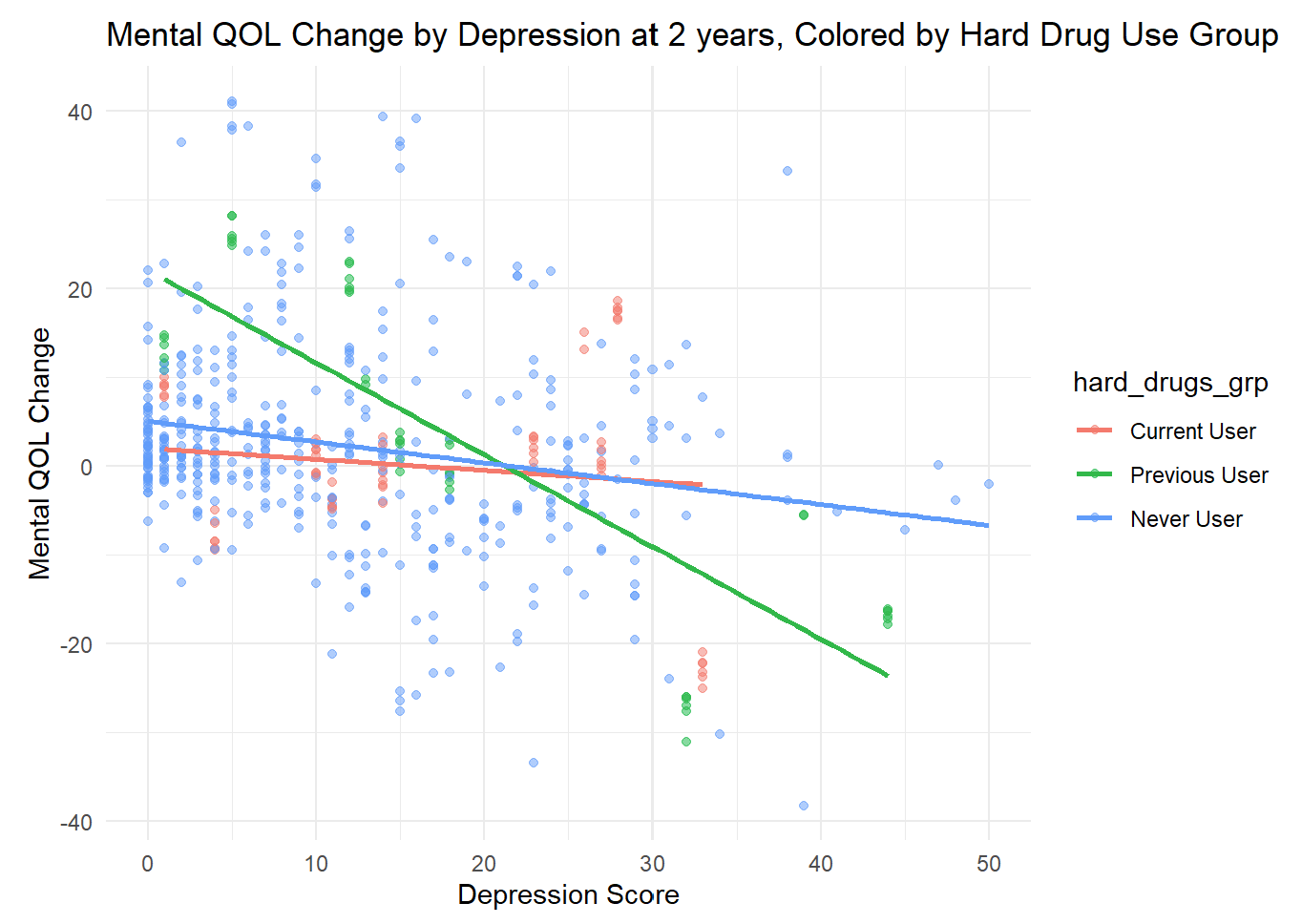
***Figure 2.***

******

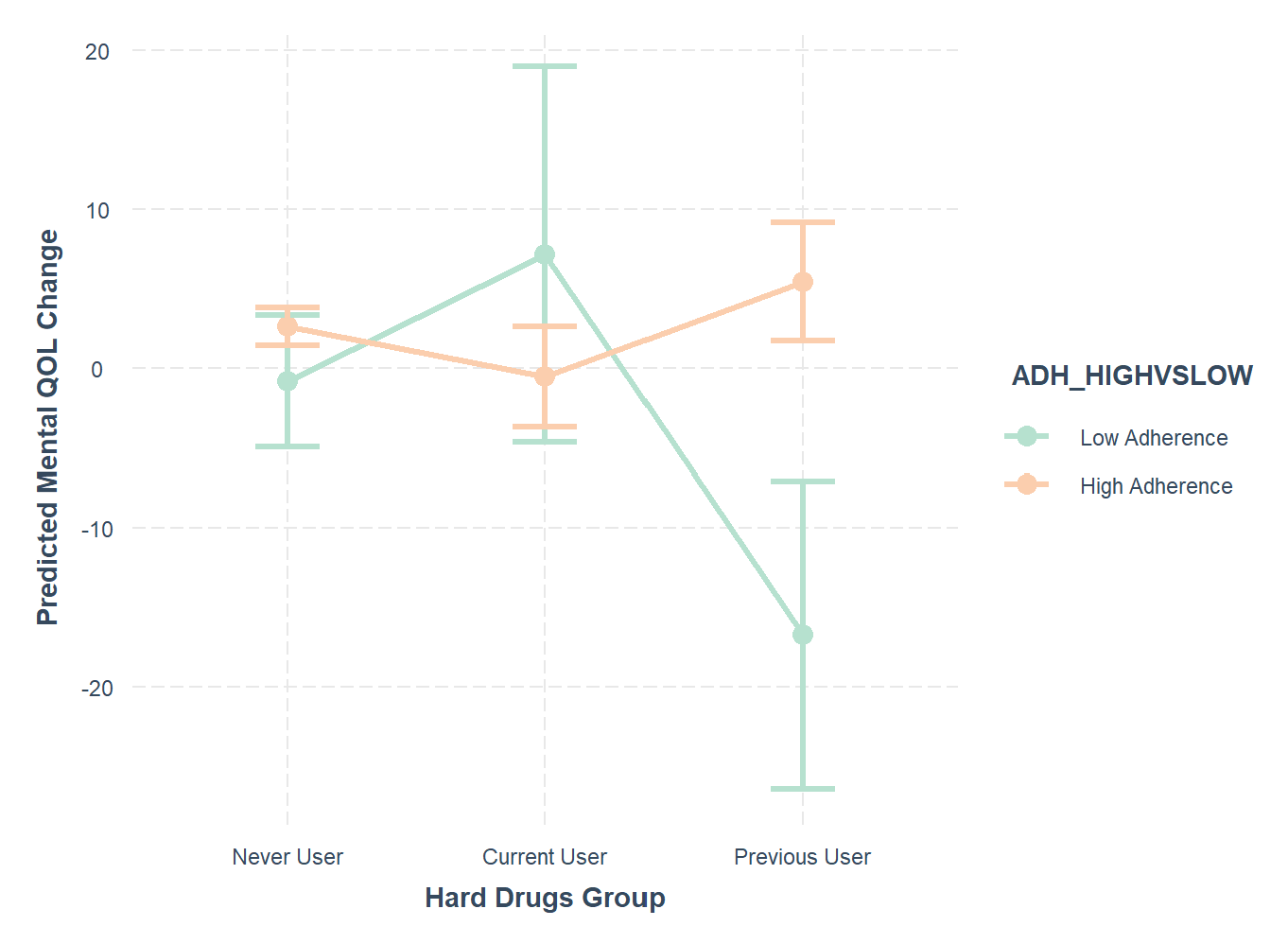
***Figure 3.***

******

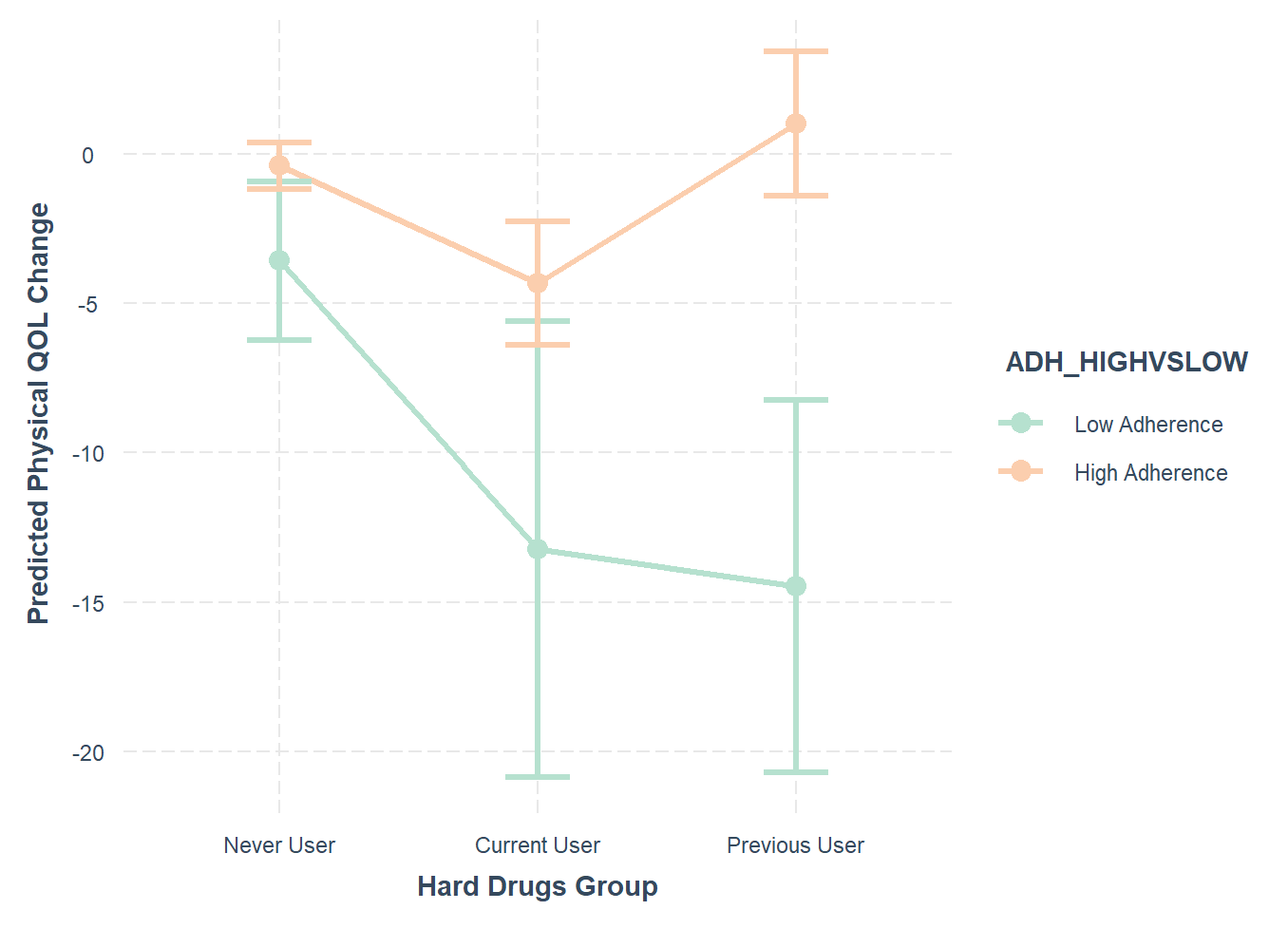
***Figure 3a.***

******

***Figure 4.***

******

***Figure 5.***

******