P1 Data Analysis Plan

Jacob Kebe

Introduction

In a cohort of HIV infected patients undergoing Highly Active Antiretrovial Treatment (HAART) over an 8 year period, there is interest whether treatment response differs between patients who reported hard drug use and those who did not. 4 measures of treatment response will be considered and tested: number of HIV copies per mL of blood ("viral load"), CD4+ T cell count (indicative of level of immunological response), and aggregate physical quality of life score and aggregate mental quality of life score from the SF-36 assessment. Under treatment, viral load is expected to decline and T cell count is expected to increase over time. We will test, for each outcome, whether there is a statisically and clinically significant relationship between the difference in the outcome at 2 years vs. baseline, and hard drug use reported at baseline.

Preliminary Methods

First, I will implement the data missingness definitions from the data dictionary, and then keep a table of baseline values for all variables. I will add to this table variables for the difference

between 2 years and baseline for each outcome of interest. I also looked at missingness patterns over time. For the first 2 years, there were not concerning patterns, but some variables have high missingness rates at baseline.

I will be testing both Bayesian and frequentist linear models of the difference between each outcome at 2 years vs. baseline against hard drug use adjusting for other relevant cofactors, for a total of 8 models. Of interest will be the significance and magnitude of the coefficient related to hard drug use in each model. I will adjust for Age, BMI, Smoking Status, Education, and Race, since these variables were deemed relevant and traditionally included cofactors for modelling by the investigator. While other variables are available for adjustment, many are highly correlated with each other which causes concern for multicollinearity, and many have high missingness rates, so they will not be included in the analysis.
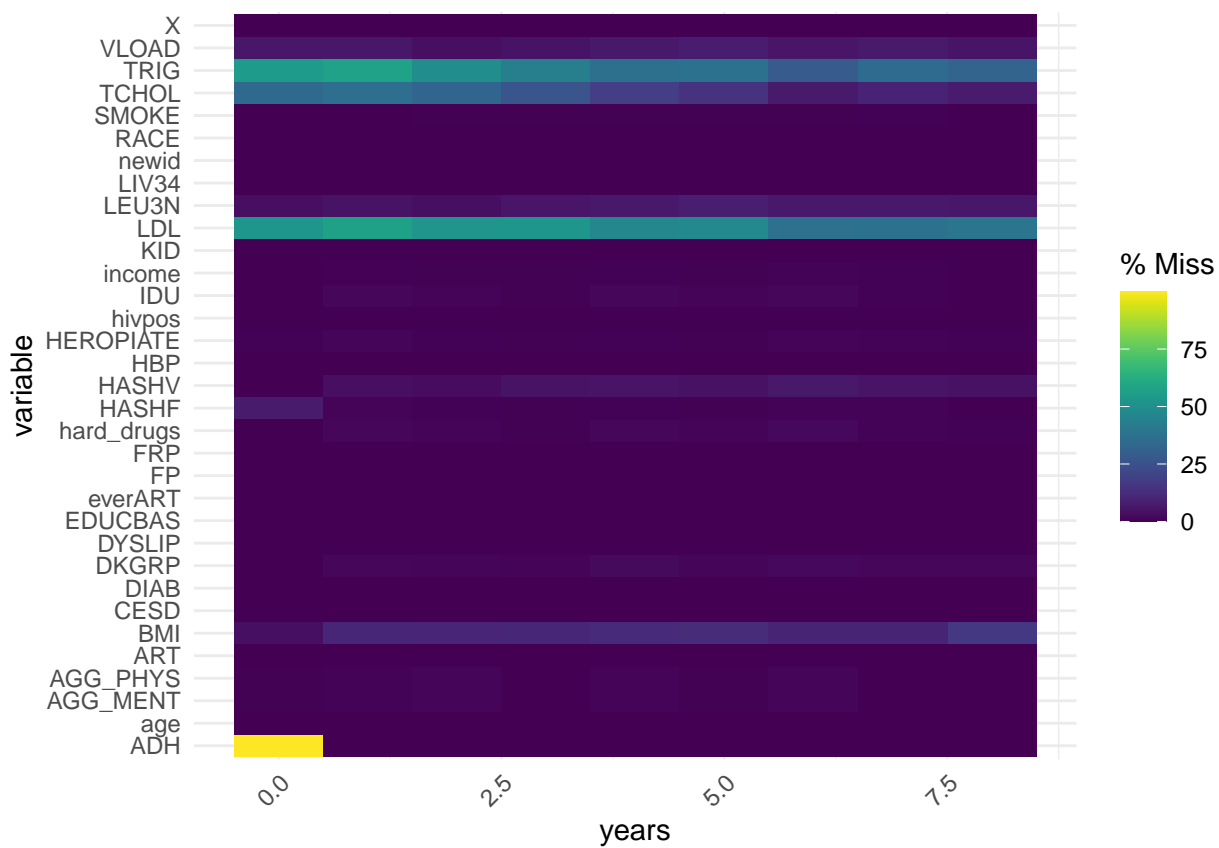
Figure 1: Missingness Rates over Time for HIV Dataset Variables