Project 1 data analysis plan

Shuai Zhu

2024-09-23

1 Introduction

The secondary data analysis of the Multicenter AIDS Cohort Study, a prospective cohort study designed to understand the natural and treated histories of HIV-1 infection in homosexual and bisexual men in 4 major in the United States. The dataset used in this analysis includes eight years of longitudinal data from 715 HIV-infected men, capturing laboratory measurements, quality of life scores, demographic information, and other health-related data collected after initiating highly active antiretroviral therapy (HAART). Multicenter AIDS Cohort Study is a prospective cohort study designed to understand the natural and treated histories of HIV-1 infection in homosexual and bisexual men in 4 major in the United States. The main research question is to know how treatment response two years after initiating HAART differs between individuals who reported using hard drugs at baseline and those who did not.

2 Method

The dataset was filtered to include only baseline and year 2 observations. The data was reshaped from long to wide format for ease of analysis. A new binary variable, adh, was created, where adh = 1 for participants with an adherence (ADH) score of 3 or less, and adh = 0 otherwise. The outcome variables include viral load, CD4+ T cell count, physical quality of life score, and mental quality of life score. Potential predictor variables, such as age, hard drug use, BMI, education status, Adherence, and baseline measurements, were considered for univariable analysis. Predictor variables that showed a significant association with the outcomes were included in multivariable regression models for each of

the four outcome measures. The assumptions of linearity, homoscedasticity, and normality will be assessed using diagnostic plots. P values <0.05 will be considered significant. R version 4.3.3 was used for data cleaning and analysis.

3 Preliminary Results

In the univariable regression analysis, several predictor variables demonstrated significant associations with the outcomes:

- Viral Load: Education status, adherence, and baseline viral load were significantly associated with viral load at two years.
- CD4+ T Cell Count: Significant predictors for CD4+ T cell count included education status and baseline CD4+ T cell count.
- Physical Quality of Life Score: Age, education status, baseline physical quality of life score, and hard drug use were significantly associated with the physical quality of life score.
- Mental Quality of Life Score: Age, BMI, education status, and baseline mental quality of life score showed significant associations with the mental quality of life score.