

Project 1

Joe Nagel

Introduction

HIV (human immunodeficiency virus) is a virus that attacks the immune system; it destroys T-cells and inhibits your body's ability to fight infection. This analysis evaluates the effectiveness of highly active antiretroviral treatment (HAART), the standard treatment for HIV patients. The data come from the Multicenter AIDS Cohort Study, an ongoing study investigating HIV infection in homosexual and bisexual men in the United States. The dataset includes longitudinal demographic information, laboratory measurements, and quality of life measurements on men infected with HIV. Data was collected on subjects at the beginning of HAART treatment (baseline) and every year thereafter for up to 8 years. The primary goal of this analysis is to characterize how treatment response after 2 years of treatment differs based on a subject's hard drug use status at baseline. Specifically, we will investigate the difference in aggregate physical quality of life score, aggregate mental quality of life score, HIV viral load, and CD4+ T cell count between baseline and two years across hard drug use status at baseline while adjusting for age, bmi, smoking status, education, and race. We will compare Bayesian and non-Bayesian analysis approaches.

Preliminary Methods

We will begin by checking the analysis variables for missingness and outliers; we will specifically check if missingness differs between hard drug users and non-hard drug users. We will calculate the change in outcomes as the value at Year 2 minus the value at Year 0. We will also visualize the distributions of the four outcome difference scores using histograms and

boxplots to identify potential outliers and assess the need for any transformations prior to analysis. For the primary analysis, we will generate descriptive statistics to compare baseline demographic and clinical characteristics between those who reported hard drug use and those who did not. To evaluate the effect of baseline hard drug use on treatment response, we will fit four separate multivariable linear regression models—one for each outcome (change in aggregate physical score, aggregate mental score, CD4+ count, and viral load). The primary predictor of interest will be baseline hard drug use status. All models will adjust for the specified demographic variables: age, BMI, smoking status, education, and race. We will approach modeling from both a Bayesian and non-Bayesian approach. For the non-Bayesian, we will use standard Ordinary Least Squares (OLS) regression, reporting point estimates, 95% confidence intervals, and p-values. For the Bayesian, we will fit Bayesian linear regression models. We will specify non-informative or weakly informative prior distributions for our parameters and estimate posterior distributions using Markov Chain Monte Carlo (MCMC) methods. We will report posterior means, 95% Highest Posterior Density (HPD) credible intervals, and posterior probabilities to evaluate the primary research question. We will compare the results from both frameworks and determine if there are any differences in the interpretations. Finally, we will investigate if any relationship between hard drug use and the outcomes is explained by adherence to the treatment protocol.