Project 1

Joseph Froelicher

October 11, 2021

Contents

Introduction

The investigative team has been tasked with performing a secondary data analysis of the Multicenter AIDS Cohort Study.These data are from an ongoing prospective cohort study of the natural and treated histories of HIV-1 infection in homosexual and bisexual men. Highly active anti-retroviral treatment (HAART) is the standard treatment for HIV infected patients. These data include up to 8 years of longitudinal laboratory and quality of life measures, in addition to demographic and other health information, on HIV infected men after beginning HAART.

There are four measures of treatment response. The first two are laboratory measures, viral load (VLOAD), which is the number of HIV copies in one mL of blood, and the second, is CD4+ T cell count (LEU3N), a measure of immunologic health. In untreated HIV infection, viral load increases over time and CD4+ T cell counts decline as the immune system is attacked by the virus. Once treatment is initiated, the investigators expect viral load to decrease rapidly and CD4 counts to recover. The final two measures are quality of life measures from the SF-36. The first is the aggregate physical quality of life score (AGG_PHYS) and the second is the aggregate mental quality of life score (AGG_MENT). These scores range from 0 to 100, with higher scores indicating better quality of life. It is unclear what happens to quality of life after initiating treatment.

The research team is interested in how treatment response two years after initiating HAART is different between subjects who report using hard drugs, such as heroine and cocaine, at baseline, and other subjects, who did not report hard drug use at baseline. The investigators have identified several variables of interest, that may have an impact on all four outcome measures, including: age, BMI, smoking status, education, race/ethnicity, and adherence. This translates to the hypothesis: hard drugs will significantly impact each of the measures of quality of life, CD4+ count and viral load. The research team was also tangentially interested in the dropout from baseline to year two in the HAART data.

Methods

Data Cleaning

Value ranges have been provided for numeric variables, and have been replaced as missing. Viral load will be log base 10 transformed due to extremely large values. Typical ranges of Cd4+ are between 500 and 1,500, and thus a value below 200 results in an AIDS Diagnosis. This will be kept in mind during the analysis.

Values of 9 or 999 were recorded for DYSLIP, FP, FRP, BMI, where there were insufficient data, and thus will be treated as missing. The analysis was done based on a complete case analysis, as determined by the researchers. Thus across all analyses there are 463 subjects used for each bayesian linear regression. The summary statistics for each outcome at baseline and known variables of interest for each of those 463 subjects is provided in Table 1 below.

Data Analysis

The analysis performed will be a simple linear regression in a Bayesian framework, with its frequentist counterpart used as a benchmark for analysis. The analysis will examine the four quality of life measures at baseline and 2 years. Each of the variables of interest, Age, BMI, smoking status, education, and race/ethnicity are recorded at baseline. The variable of interest, adherence, is recorded at 2 years.

All parameters for the Byesian analysis were given non-informative priors. After a prior sensitivity analysis, it was determined that too narrow of prior distribution affected the modeling scheme significantly, and thus non-informative priors were used. See prior specification below.

Results

Descriptive Statistics

The descriptive statistics in Table 1 show us that there are a few differences between the hard drug users and non hard drug user groups for the $n = 463$ subjects. Partiuclarly for smoking status, education, and adherence. This would be evident in our multivariate analyses. Of particular note are the differences in smoking and adherence between the two groups.

Bayesian Regression

The investigators noted particular interest in the effects of hard drug use on each of the four outcomes. A change of 10% change between user and non-user groups would represent significant clinical results. Based on our results in Table 2, it is evident that both CD4+ count and Aggregate Physical score are significantly impacted by the the use of hard drugs (posterior probability > 0.95 for a 10% change). Hard drug use had a slightly lower posterior probability of indicating a 10% change in aggregate mental score (~0.9). Log

viral load was not significantly impacted by the use of hard drugs. This is true in both the uni-variate and multivariate iterations of this analysis.

The average CD4+ count after adjusting for adherence, age, BMI, education, race/ethnicity, and smoking status decreased by-167.32 for hard drug users. Viral load increased by a factor of 0.891 for hard drug users after adjusting for covariates. Aggregate mental score decreased by -0.653 on average for hard drug users after adjustment. And aggregate physical score decreased on average by-3.136 after adjustment for hard drug users.

The introduction of each of the covariates education, smoking status, race, age and BMI did not yield significantly different results in the two outcomes CD4+ count and aggregate physical score. However, for each of the two outcomes where the effects of hard drugs was not significant (log viral load and aggregate mental score), the estimate of the effects of hard drugs changed significantly, in magnitude and direction. For log based 10 viral load this translates to an decrease in the percent change in the effect of hard drugs on viral load after adjusting for covariates.

Frequentist Regression

For the purposes of benchmarking the estimates of our Bayesian linear regression results from a frequentist iteration of the same analysis have been provided. The estimates yielded similar results, and thus there should not be any concern that the posterior distribution was not properly sampled, or that the prior distributions had any significant impact over the reported results.

Dropout

There was noticeable dropout from baseline to two years for each of the four outcomes, as seen in table 4. Additionally, many subjects were missing data for some of: education, smoking, adherence, race/ethnicity, BMI, and age. This will be noted as a study limitation.

Conclusion

The purpose of this analysis was to identify the effects of hard drugs use on four outcomes, in a study of HIV positive subjects. Those outcomes were CD4+ count, log based 10 transformed viral load, SF-36 aggregate mental score, and SF-36 aggregate physical score. For CD4+ count and aggregate physical score, there was

a significant effect in the outcome as a results of hard drugs use. Hard drug users did not experience as much success as a part of the HAART treatment.

Adjusting for the variables age, BMI, education, race/ethnicity, adherece, and smoking status provided better context for interpreting viral load and aggregate mental score. For CD4+ white blood cell count and aggregate physical score, these variables had negligible effects on the affect of hard drugs for these two outcomes. The analysis has confirmed both the importance of hard drug use to the HAART treatment, as well as the importance of including contextual variables like smoking status, adherence, age, BMI, education and race/ethnicity.

Limitations

There were two limitations of not for this analysis. There was significant loss to follow up from baseline to year two, which in a complete case analysis heavily affects our sample size (see table 4). Our Sample size was large enough (n = 463) though that we are still adequately powered to perform this analysis. Secondly, there were large amounts of missing data, and again in a complete case analysis this means that our sample size was decreased due to missing data. This was particularly evident the the variable BMI. Between these two effects several hundred observations were removed. Lastly, the variable of interest, hard drug use, was observed for all subjects at baseline, which does not account for new use of hard drugs after baseline survey results. The results of this analysis would suggest that researchers may be interested in a new line of questioning involving hard drug use after baseline survey results.

Tables

Table 1

	Hard Drugs	Absence	Presence
n		427	36
Mental Score (Mean ± sd)		45.15 ± 13.77	42.01 ± 11.64
Physical Score (Mean ± sd)		51.35 ± 9.00	49.16 ± 7.05
Log Viral Load (Mean ± sd)		4.53 ± 0.93	4.58 ± 0.87
CD4 positive cell count (Mean ± sd)		374.45 ± 202.29	360.04 ± 200.80
BMI (Mean ± sd)		25.34 ± 4.39	23.62 ± 3.45
Age (Mean ± sd)		43.10 ± 8.68	43.89 ± 9.52
Race/Ethnicity (%)	White, non-Hispanic	382 (89.5)	35 (97.2)
	Black, non-Hispanic	45 (10.5)	1 (2.8)
	Hispanic	277 (64.9)	19 (52.8)
	Other	19 (4.4)	0 (0.0)
Education (%)	< High School	108 (25.3)	14 (38.9)
	High School	23 (5.4)	3 (8.3)
	> High School	19 (4.4)	10 (27.8)
Adherence (%)	Not Adherent	61 (14.3)	3 (8.3)
	Adherent	347 (81.3)	23 (63.9)
Smoker (%)	Non-Smoker	278 (65.1)	9 (25.0)
	Smoker	149 (34.9)	27 (75.0)

Table 1. Summary Statistics for complete case analysis of outcomes and potential confounding variables.

Table 2

Treatment Response	Model Type	Hard Drugs	95% HPDI	DIC	P(10% Change)
CD4+ Count	Crude	-170.473	(-231.764, -110.731)	6113.295	1.00000
CD4+ Count	Adjusted	-167.320	(-228.88, -103.216)	6117.275	1.00000
Log 10 Viral load	Crude	0.013	(-0.414, 0.429)	1510.859	0.20380
Log 10 Viral load	Adjusted	-0.050	(-0.449, 0.358)	1434.023	0.19342
SF36 MCS	Crude	1.458	(-2.703, 5.53)	3626.065	0.92856
SF36 MCS	Adjusted	-0.653	(-4.182, 2.852)	3447.172	0.89880
SF36 PCS	Crude	-3.463	(-6.242, -0.664)	3264.651	0.99580
SF36 PCS	Adjusted	-3.136	(-5.882, -0.442)	3215.578	0.99428

Table 2. Table of Results of the Bayesian Regression Analysis, with posterior mean estimates for the Variable of interest, Hard Drug use. Reported are the Highest posterior density intervals, and the deviance criterion for each model.

Table 3

Treatment Response	Model Type	Hard Drugs	95% CI
CD4+ Count	Crude	-170.4574180	(-230.967, -109.947)
CD4+ Count	Adjusted	-167.5137043	(-230.531, -104.497)
Log 10 Viral load	Crude	0.0142904	(-0.406, 0.434)
Log 10 Viral load	Adjusted	-0.0503837	(-0.451, 0.35)
SF36 MCS	Crude	1.4671839	(-2.657, 5.591)
SF36 MCS	Adjusted	-0.6471135	(-4.176, 2.882)
SF36 PCS	Crude	-3.4648546	(-6.256, -0.673)
SF36 PCS	Adjusted	-3.1380177	(-5.883, -0.393)

Table 3. Table of Results of the Frequentist Regression Analysis, with posterior mean estimates for the Variable of interest, Hard Drug use. Reported are the 95% confidence intervals.

Table 4

Variable	Time Point	N	% Dropout
CD4+	baseline	691	
	2 years	487	29.52
Viral Load	baseline	673	
	2 years	487	27.64
SF-36 MCS	baseline	713	
	2 years	500	29.87
SF-PCS	baseline	713	
	2 years	500	29.87

Table 3. Table of analysis of dropout at year two.

Prior Specification

$$\beta_{intercept} \sim \mathcal{N}(0.5, 0.0000001)$$

$$\beta_i \sim \mathcal{MVN}(0.5, 0.0000001)$$

$$\epsilon_i \sim \mathcal{G}amma(0.0001, 0.0001)$$

Reproducibility

Code for this analysis is availables from `https://github.com/BIOS6624-UCD/bios6624-JoeFroelicher/project1`, under the branch `project1`. Special note to `@willippit` and `@ehccooper` for calculation of probabilities and inspiration for tables respectively.