**Project:** Project 1, BIOS 6623

**Report:** Descriptive Statistics & Bayesian Analysis

**Investigator:** Rachel Johnson

**Date:** 9 October 2017

**Introduction**

The goals of this project are to examine the difference in treatment responses of hard drug users and those who don’t report hard drug use in HIV-1 infected homosexual and bisexual men in the Multicenter AIDS cohort study. Outcomes of interest include two laboratory measures—viral load and CD4+ T cell count—and two quality of life scores—aggregate mental score and aggregate physical score from the SF-36.

These measures were taken on men at baseline and throughout the follow-up period of the study, but this analysis aims to examine how these treatment responses differ between users and non-users of hard drugs from baseline values to their responses at the 2 year follow-up time point in the study.

**Methods**

There were 715 men included in the study at baseline, but 209 of these men did not have data at the 2 year follow-up, so they were excluded. Of the remaining 506, 28 were excluded since they did not have data available for all 4 outcomes, which was acceptable since none of those excluded were hard drug users, so it didn’t exclude any data on our population of interest.

The following demographic variables and potential covariates were selected in addition to hard drug use on the basis of covariate adjustment in previous studies: baseline outcome values, age, BMI, race, baseline marijuana use, baseline alcohol use, baseline smoking status, income, education, and adherence to ART at 2 years.

Some of these variables’ categories were collapsed to match how they were categorized in previous studies. Race was categorized into non-Hispanic white vs. other; alcohol use was categorized into > 13 drinks per weed vs. 13 or fewer drinks per week; smoking was categorized into current smokers vs. never/former smokers; income was categorized into < $10,000, $10,000 - $40,000, and > $40,000; education was categorized into greater than a high school education vs. a high school education or less; and ART adherence at the 2 year mark was categorized into > 95% vs. < 95%.

Continuous variables were all remained continuous for the purposes of the analysis and descriptive statistics. Impossible values for BMI were removed, and viral load was transformed on the log10 scale.

-Bayesian analysis methods

* Introduction: Briefly describe the project (including data received from the investigator) and the scientific hypothesis of interest. Rephrase the scientific hypothesis of interest into testable statistical hypotheses (~.5 page).  Note this is NOT a scientific introduction to a paper and should give information pertinent to the data analysis, not more general biology background.
* Methods: Describe the methods used to clean and analyze the data. Justify and explain your data analysis approach (~ 2 pages).  Should be written in past tense and should not include results.  Do NOT include equations.
* Results: Present results for analyses described in the methods (~1-1.5 pages). Use Tables and Figures as appropriate, including in the text the full interpretation of statistical results for the main findings (i.e. point estimates, confidence bounds, p-values, interpretation of results of test).
* Conclusions: Interpret your results (~.5 to 1 page) in context of scientific question(s). Also discuss any limitations to your analysis that may affect interpretation or that require additional consideration by the investigator.

**Results**

**Table 1.**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Total** | **Hard drugs = Yes** | **Hard drugs = No** |
| **Age at baseline** | 43.26 ± 8.72 | 44.62 ± 9.49 | 43.14 ± 8.65 |
| **BMI at baseline** | 25.21 ± 4.34 | 23.62 ± 3.45 | 25.34 ± 4.39 |
| **Alcohol use at baseline** |  |  |  |
| 13 or fewer drinks per week | 443 (93.07) | 37 (94.87) | 406 (92.91) |
| > 13 drinks per week | 33 (6.93) | 2 (5.13) | 31 (7.09) |
| **Smoking status at baseline** |  |  |  |
| Never/former | 291 (61.13) | 9 (23.08) | 282 (64.53) |
| Current | 185 (38.87) | 30 (76.92) | 155 (35.47) |
| **Marijuana use at baseline** |  |  |  |
| Yes | 194 (40.76) | 12 (30.77) | 182 (41.65) |
| No | 282 (59.24) | 27 (69.23) | 255 (58.35) |
| **Income level at baseline** |  |  |  |
| < $10,000 | 99 (20.8) | 14 (35.9) | 85 (20.24) |
| $10,000 - $40,000 | 196 (41.18) | 16 (41.03) | 180 (42.86) |
| > $40,000 | 164 (34.45) | 9 (23.08) | 155 (36.9) |
| **Education at baseline** |  |  |  |
| HS or less | 101 (21.22) | 16 (41.03) | 85 (19.45) |
| >HS | 375 (78.78) | 23 (58.97) | 352 (80.55) |
| **Adherence at 2 years** |  |  |  |
| <95% | 50 (10.5) | 1 (2.56) | 49 (11.21) |
| >95% | 426 (89.5) | 38 (97.44) | 388 (88.79) |
| **Baseline log10 viral load** | 4.52 ± 0.92 | 4.52 ± 0.86 | 4.52 ± 0.93 |
| **Baseline CD4+ count** | 373.53 ± 200.46 | 352.18 ± 194.67 | 375.44 ± 201.07 |
| **Baseline SF36 MCS score** | 44.89 ± 13.5 | 42.31 ± 11.22 | 45.12 ± 13.68 |
| **Baseline SF36 PCS score** | 51.04 ± 9.05 | 47.7 ± 8.5 | 51.34 ± 9.05 |

**Table 2.**  This table displays the difference in the four outcomes from baseline to the 2 year mark in the study.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Total** | **Hard drugs = Yes** | **Hard drugs = No** |
| **Difference in log10 viral load** | 0.41 ± 0.37 | 0.39 ± 0.25 | 0.41 ± 0.38 |
| **Difference in CD4+ count** | 169.21 ± 183.2 | 13.43 ± 195.73 | 183.11 ± 175.67 |
| **Difference in SF36 MCS score** | 2.34 ± 12.1 | 3.58 ± 15.07 | 2.23 ± 11.82 |
| **Difference in SF36 PCS score** | -1.55 ± 8.37 | -3.85 ± 8.71 | -1.34 ± 8.32 |

**Conclusion**

https://github.com/BIOS6623-UCD/bios6623-johnsra3/tree/master/Project1

**Code**