*­Nigeria R3 Dataset*

Data were collected on 2500 people living with HIV who were on stable antiretroviral therapy (ART) for at least 6 months and attending the HIV clinic at the Amino Kano Teaching Hospital. People living with HIV are at increased risk for developing kidney disease. The primary purpose of this study was to investigate factors associated with kidney disease, and in particular, to measure the association between the apolipoprotein 1 (APOL1) high-risk genotypes and markers of poor kidney function. This dataset contains cross-sectional data. Primary study results, as well as details about the study design, were reported in Wudil et al., 2021.

*Reference*

Wudil UJ, Aliyu MH, Prigmore HL, Ingles DJ, Ahonkhai AA, Musa, BM, Muhammad H, Sani MU, Nalado AM, Abdu A, Abdussalam K, Shepherd BE, Dankishiya FS, Burgner AM, Ikizler TA, Wyatt CA, Kopp JB, Kimmel PL, Winkler CA, Wester CW. Apolipoprotein-L1 (APOL1) risk variants and associated kidney phenotypes in an adult HIV cohort in Nigeria. Kidney International 2021; 100: 146–154

*Data Dictionary*

|  |  |
| --- | --- |
| Variable | Description |
| age | Age in years |
| male | Male sex at birth (1=male, 0=female) |
| bmi | Body mass index (kg/m2) |
| smoke | Current smoker (1=yes, 0=no) |
| cd4 | CD4 T Cell count (cells/mm3) |
| tdf | Currently on an antiretroviral therapy regimen containing Tenofovir |
| dtg | Currently on an antiretroviral therapy regimen containing Dolutegravir |
| risk.alleles | Number of high risk APOL1 genotype alleles (0, 1, or 2) |
| htn | Self-reported hypertension (1=yes, 0=no) |
| jnc\_bp | Joint National Committee blood pressure classification (Normal, Pre-hypertension, Stage 1 Hypertension, Stage 2 Hypertension) |
| Years.ART | Years on antiretroviral therapy |
| uACR | Urine albumin creatinine ratio (mg/g) |
| eGFR | Estimated glomerular filtration rate (ml/min per 1.73 m2) |
| potassium |  |
| sodium |  |
| bicarbonate |  |
| chloride |  |
| urea |  |
| SBP | Systolic blood pressure (mm Hg) |
| DBP | Diastolic blood pressure (mm Hg) |
| DM | Diabetes mellitus (1=yes, 0=no) |

*Assignments*

Among people living with HIV in Kano, we are interested in studying kidney function as measured using eGFR. We are also interested in studying factors associated with eGFR.

*Day 1.*

Explore the dataset and create a table that describes the data, including medians (25th and 75th percentiles) for continuous data and counts (and percentages) for discrete data. Investigate missingness of the variables.

*Day 2.*

Discuss possible missing data mechanisms with a clinician (Dr. Aliyu). Based on our discussions, fit models for the missing data. Use inverse probability weighted methods to estimate the mean CD4 count and the proportion of people with 0, 1, and 2 risk alleles. Compute 95% confidence intervals for these quantities using the bootstrap.

*Day 3*.

Based on discussions with a clinician (Dr. Aliyu), fit a regression model for eGFR that includes CD4 count, risk alleles, and other variables thought to be important. Use multiple imputation to account for missing data. Report coefficient estimates and 95% confidence intervals in a table. Interpret findings.

*Day 4.*

Although we know it is a bad idea because it discards information, some people dichotomize eGFR as ≤60 or >60; they refer to eGFR ≤ 60 as kidney disease. Using this definition, fit a logistic regression model for kidney disease that includes CD4 count, risk alleles, and other variables thought to be important. You may have to include fewer variables in this model because the outcome is a somewhat rare dichotomous event. Use multiple imputation to account for missing data. Report coefficient estimates and 95% confidence intervals in a table. Interpret findings.

*Final Homework*

Create a report that puts each day’s homework assignments together using R Markdown. Describe the methods used. Include relevant tables and figures. Summarize findings and interpret results with a few paragraphs.