

2021 Biostatistics II (PPH7092S) - Assignment 3

Release date: 06 October 2021

Due date: 29 October 2021

Note on submission: Please get in touch early to address late submissions – failure to do so (late submission), will result to a penalty (10 % per day).

Aim of the assignment: The aim of this assignment is to demonstrate that you can apply two methods of survival analysis to some real data and accurately interpret the results. You will be asked to plot Kaplan-Meier curves, perform formal tests to explore group differences, fit univariable and multivariable Cox regression models, and interpret the results.

Instructions:

This is an individual assignment without consultation or discussion with other class members. However, ALL coding related questions and/or other problems should be posted on the forums.

- a) Answer the questions in a **typed** report. Put your name, student number, and name of the dataset you analyzed on the first page (cover page). Number your answers as they appear in the assignment. **Submit a pdf file with your answers, as well as the script/code file to Vula.**
- b) **Be concise**, complete, and correct when answering the questions.
- c) All figures and tables should have appropriate labels and captions. The tables should be formatted (**do not copy-paste or screen shot** results from software).
- d) Please round off solutions to **2 decimal places** wherever applicable.
- e) The **page limit** for the report is 3 pages (excluding cover page and R code or do file).
- f) **Submit your code file.** If you are unable to submit a code file, it will be assumed that the work is not yours and **you will receive zero** on the assignment.

Data:

Please note: Each of you will receive a unique dataset, please be sure to indicate the name of the dataset that you analyzed.

The data with 1,151 observations (**this will vary from your given unique dataset**) come from a double-blind, placebo-controlled trial that compared the three-drug regimen of indinavir (IDV), open label zidovudine (ZDV) or stavudine (d4T) and lamivudine (3TC) with the two-drug regimen of zidovudine or stavudine and lamivudine in HIV-infected patients (Hammer et al., 1997). Patients were eligible for the trial if they had no more than 200 CD4 cells per cubic millimeter and at least three months of prior zidovudine therapy. Randomization was stratified by CD4 cell count at the time of screening. The primary outcome measure was time to AIDS defining event or death. A description of variables in the dataset follows:

- **censor** - event indicator for AIDS defining, coded as 1 for AIDS defining diagnosis or death (whichever comes first); and 0 otherwise
- **time** - time to AIDS diagnosis or death (days)
- **tx** - treatment indicator, coded as 1 if treatment includes IDV; and 0 if control group
- **ivdrug** - IV drug use history, coded as 1 if never; 2 if currently and 3 if previously
- **sex** - coded as 1 if male, and 0 if female
- **cd4** - baseline CD4 cell count (cells/milliliter)

- age - age at enrollment (years)

Total [50 Marks]

Question 1 [20 marks]: Non-parametric approach (Descriptive statistics)

- Plot Kaplan-Meier curves for: [8]
 - the different IV drug use history strata (Figure 1), and
 - different treatment strata (Figure 2)
- Compare the survival curves of the groups below using the log-rank test. State the hypothesis being tested and briefly interpret the test results. [8]
 - the three IV use drug history groups,
 - the treatment group indicator
- Based on the plots in Q1a (i) and (ii): provide estimates of the survival probabilities at 90 days with 95% confidence intervals for each group. Interpret the results (ONE SHORT paragraph) [4]

Question 2 [30 marks]: Semi-parametric approach (Modelling)

- Present a single table (Table 1) showing the results of the models below with Hazard Ratios (HR), 95% confidence intervals and p-values. [15]
 - Univariable Cox model, for the covariates: treatment, sex, cd4 and age.
 - Multivariable Cox model investigating the effect of treatment (tx) on the primary outcome, adjusting for sex, cd4, and age as potential confounders.
- Completely interpret your Cox regression analysis. In conclusion, decide which treatment is better. [15]