

# BIOST 537: Homework 3

Department of Biostatistics @ University of Washington

Alejandro Hernandez

Winter Quarter 2025

## PROBLEM 1.

In the CCG803 study, 268 children in remission of acute lymphoblastic leukemia (a type of blood cancer) were recruited from a number of clinical institutions and randomized to one of two maintenance regimes. Patients in the control group were assigned to the standard of care—chemotherapy with 6MP and methotrexate—while patients in the treatment group were also given dactinomycin.

Dactinomycin administration (**rx**), age at baseline (**age**), white blood cell count (**wbc**), clinical institution (**institution**), observed follow-up time (**duration**) and relapse status at end of follow-up (**relapse**) were recorded for each patient. The dataset, called **ccg803.csv**, can be found on the canvas website.

- (a) Describe and fit a proportional hazards model to estimate the association between risk of relapse and treatment. Provide an estimate and 95% confidence interval for the hazard ratio.

We fit a Cox proportional hazards model of time until blood cancer relapse given a patient's treatment group. We estimate the hazards ratio comparing patients who were administered Dactinomycin treatment with standard-of-care to those with only standard-of-care to be 0.738 (95% CI: 0.55-0.992). Because our confidence interval for the hazard ratio ranges from half to equal, we cannot claim with certainty that the sample has sufficient statistical evidence of Dactinomycin being associated with better outcomes. This is illustrated in all three provided statistical tests producing p-values that are approximately equal to  $\alpha$ .

- (b) Describe and fit a proportional hazards model to estimate the association between risk of relapse and treatment adjusting for white blood cell count and age. Provide an estimate and 95% confidence interval for each of the resulting hazard ratios

We fit a Cox proportional hazards model of time until blood cancer relapse given a patient's treatment group, adjusting for white blood cell count and age at baseline. We estimate the hazards ratio comparing two patient populations of the same white blood cell count and age at baseline, who differ in their inclusion of Dactinomycin to a standard-of-care treatment, to be 0.744 (95% CI: 0.553-0.999), with those administered Dactinomycin facing lower risk of relapse. From this model, we again believe the probability of entering cancer relapse is higher for patients who did not receive Dactinomycin, although the statistical evidence is not strong according to our significance level.

- (c) Describe and fit a proportional hazards model to determine whether the association between risk of relapse and treatment adjusting for white blood cell count and age differs in subpopulations of patients defined by white blood cell count being either below 10,000 (**wbc**<100), or above or at 10,000 (**wbc**>=100). Provide an estimate and 95% confidence interval for the hazard ratio corresponding to each of these subpopulations. Perform a test of the hypothesis that this association does not differ in these subpopulations.

We fit a Cox proportional hazards model of time until blood cancer relapse given a patient's treatment group, allowing for effect modification from white blood cell count above 10,000 and adjusting for age at baseline. We estimate the hazards ratio comparing two patient populations of the and age at baseline, who differ in their inclusion of Dactinomycin to a standard-of-care treatment, to be 0.744 (95% CI: 0.553-0.999), with those administered Dactinomycin facing lower risk of relapse.

## **PROBLEM 2.**

**End of report. Code appendix begins on the next page.**

## Code Appendix

```
# Clear environment
rm(list=ls())

# Setup options
knitr::opts_chunk$set(echo=FALSE, warning=FALSE, message=FALSE, results='hide')
options(knitr.kable.NA = '-', digits = 2)
labs = knitr::all_labels()
labs = labs[!labs %in% c("setup", "allcode")]
# Load relevant packages
library(survival) # survival model
library(dplyr)    # data manipulation
library(knitr)    # pretty tables
library(ggplot2)  # data visualization
# library(muhaaz)  # hazard function

# Load data
methadone <- read.csv("../data/methadone.csv")

# Handle missing data

#### Question 1 ####
ccg_dat <- read.csv("../data/ccg803.csv")
ccg_surv <- with(ccg_dat, survival::Surv(duration, relapse))

# Fit a PH model of time to relapse from treatment
survival::coxph(ccg_surv ~ rx, ccg_dat) %>% summary

# Fit a PH model of time to relapse from treatment, WBC, and age
coxph(ccg_surv ~ rx + wbc + age, ccg_dat) %>% summary

# Fit a PH model of time to relapse from treatment, WBC, and age
ccg_dat <- within(ccg_dat, wbc_bin <- ifelse(wbc>=100, 1, 0))

coxph(ccg_surv ~ rx*wbc_bin + age, ccg_dat) %>% summary

#### Question 2 ####
```

End of document.