BIOST 515/518 Homework 7

Latera Tesfaye Olana

10 March, 2023

Responses:

Question 1:

The event we are interested in is first tumor recurrence.

After applying *Surv()*, the following are the time to event data. Table 1 shows summary of the first 10 survival probabilities.

```
##  [1]  1   1   1   1+  1+  1+  2   2   2   2   2   2   2   2   3   3   3   3   3
## [20]  3   3   3   4   4   4+  5   5   5   6   6   6   6   7   7+  9   9   9+ 10
## [39] 10+ 10+ 12  12  13+ 14+ 16  17  17  17  18  18+ 18+ 22  22+ 23+ 24  25  25+
## [58] 25+ 25+ 26  26+ 28  29  29+ 29+ 29+ 32+ 34+ 35  36+ 37+ 38  38+ 41+ 41+ 41+
## [77] 44+ 45+ 46+ 49+ 49+ 50+ 54+ 59+ 59+
```

Table 1: Summary of the 10 first observation

time	n.risk	n.event	n.censor	estimate	std.error	conf.high	conf.low
1	85	3	3	0.96471	0.02075	1.00000	0.92627
2	79	8	0	0.86701	0.04309	0.94342	0.79680
3	71	8	0	0.76932	0.06038	0.86596	0.68347
4	63	2	1	0.74490	0.06454	0.84535	0.65639
5	60	3	0	0.70765	0.07101	0.81333	0.61571
6	57	4	0	0.65799	0.07979	0.76938	0.56273
7	53	1	1	0.64558	0.08203	0.75819	0.54969
9	51	2	1	0.62026	0.08678	0.73526	0.52325
10	48	1	2	0.60734	0.08929	0.72350	0.50983
12	45	2	0	0.58035	0.09490	0.69899	0.48184

Figure 1 shows the survival curve for thiotepa and placebo arms. On the figure vertical drop represents occurrence of an event (tumor recurrence) and (+) shows censoring. Censoring seems common at later months. The median the survival time for thiotepa arm is approximately around 25 weeks, where as for placebo arm it is approximately 15 weeks. During the first months we can see the survival functions for treatment and control group overlaps. This probably implies, non-proportionality in the hazard function.

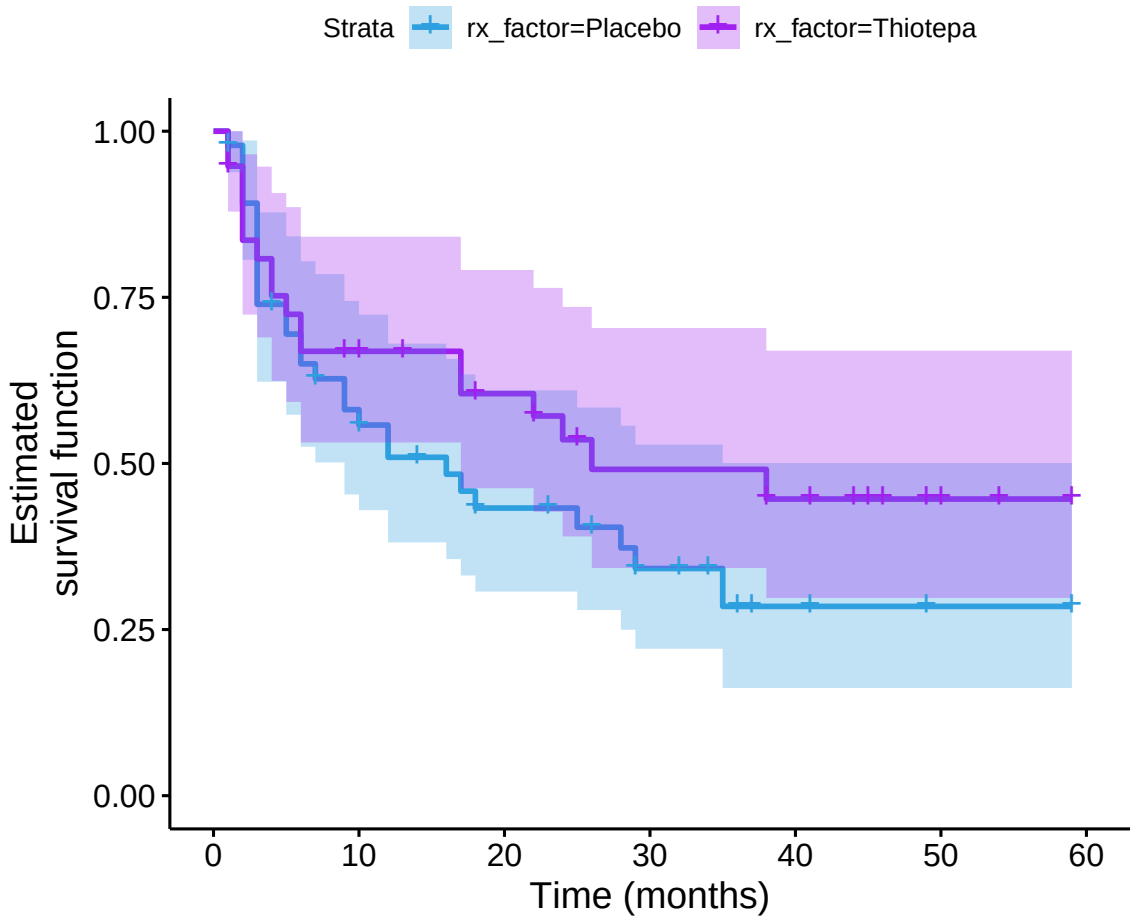

Figure 1: KM curve

Question 2:

```
## Call:
## survdiff(formula = tte ~ rx_factor)
##
##                       N Observed Expected (O-E)^2/E (O-E)^2/V
## rx_factor=Placebo  47       29     24.9     0.671      1.52
## rx_factor=Thiotepa 38       18     22.1     0.757      1.52
##
##  Chisq= 1.5  on 1 degrees of freedom, p= 0.2
```

Based on the log-rank test, there is no significant difference in the distribution of first tumor recurrence between the thiotepa and placebo treatment arms ($\chi^2 = 1.52$, p=0.2). Therefore, we fail to reject the null hypothesis that there is no difference in first tumor recurrence between the two treatment arms (p-value = 0.2).

Question 3:

The model we want to fit can be written as follow: $log(h_i(t)/h_0(t)) = \beta_1 \times X_{i1}$. Where, $h_i(t)$ is the hazard for a population member with $X_1 = X_{i1}$; Where, $h_0(t)$ is the hazard for a population member with $X_1 = 0$, i.e, within a placebo study arm. Accordingly, $h_0(t)$ represents a reference or baseline hazard for a placebo study arm. X_{i1} is 1 for treatment arm group (those receiving thiotepa) and 0 for control arm (those receiving placebo).

The fitted model can be re-written as:
$\hat{h_i}(t) = h_0(t) * e^{\hat{\beta}_1 \times X_{i1}}$. Substituting the fitted number, the formula becomes: $\hat{h_i}(t) = h_0(t) * e^{-0.3706} \times X_{i1}$.
$\hat{h_i}(t)/h_0(t)$ is the fitted hazard ratio.

Characteristic	**log(HR)**	**95% CI**	**p-value**
Treatement Arm			
Placebo	—	—	
Thiotepa	-0.37061	-0.96377, 0.22255	0.2

Based on a Cox proportional hazards model, we estimate that the hazard ratio of first tumor recurrence is 0.69 when comparing populations in both study arms (those who received thiotepa - treatment arm to those who received placebo - control arm), with the control arm having greater estimated hazard (95% CI for hazard ratio: 0.38145 - 1.24926).

Question 4:

Table 3: Summary of hypothesis testing

term	logLik	statistic	df	p.value
~ 1	-185.1376	NA	NA	NA
~ rx_factor	-184.3698	1.535631	1	0.2152694

We fail to reject the null hypothesis that the hazard for first tumor recurrence is equal across populations of both study arms (those who received thiotepa - treatment arm versus those who received placebo - control arm) (Likelihood-Ratio Test - LRT, p = 0.2152).

Question 5:

Yes, they are different.

Even though, both are used to test survival differences in two or more groups, the hypothesis testing in question 4 and 2 are different (cited):

The Cox proportional hazards model tests for the effect of a predictor on the hazard of an event occurring over time, while the log-rank test is a non-parametric test that compares the observed survival times between groups.

The Cox proportional hazards model estimates the hazard ratio, which represents the change in hazard for a one-unit change in the predictor, and provides a confidence interval and p-value for the effect. The log-rank test produces a test statistic and p-value that indicate whether there is a significant difference in survival between groups.

Over all, the p-value obtained from a Cox proportional hazards model represents the significance of the association between the predictor and the hazard ratio, taking into account the time-to-event data and the censoring information. where as, the p-value obtained from a log-rank test represents the significance of the difference in survival curves between the two groups.

Question 6 :

Individuals who present with largest tumors in the bladder or urinary tract are at a heightened risk for cancer recurrence when contrasted with those small tumors. This suggests that the size of initial tumors may serve as a predictor for the likelihood of subsequent recurrences. Based on these fact, there is a scientific rationale for having 7 different categories of bladder tumor sizes.

The fitted model can be written as:

$\hat{h_i}(t) = h_0(t) * e^{-0.43\times X_{i1} + -1.63\times X_{i2} + -0.058\times X_{i3} + -0.54\times X_{i4} + 0.18\times X_{i5} + 0.45\times X_{i6} + 0.16\times X_{i7}}$. Where, $h_0(t)$ is the baseline hazard for a study participant who had one previous bladder tumor and assigned to control arm to receive placebo. X_{i1} is 1 for treatment arm group (those receiving thiotepa) and 0 for control arm (those receiving placebo); X_{i2} is 1 if the i^{th} study participant has two centimeters of bladder tumors before enrollment, other wise zero; X_{i3} is 1 if study the i^{th} study participant has three centimeters of bladder tumors before enrollment, other wise zero; X_{i4} is 1 if study the i^{th} study participant has four centimeters of bladder tumors before enrollment, other wise zero; X_{i5} is 1 if study the i^{th} study participant has five centimeters of bladder tumors before enrollment, other wise zero; X_{i6} is 1 if study the i^{th} study participant has six centimeters of bladder tumors before enrollment, other wise zero; X_{i7} is 1 if study the i^{th} study participant has eight centimeters of bladder tumors before enrollment, other wise zero.

Characteristic	**log(HR)**	**95% CI**	**p-value**
Treatement Arm			
Placebo	—	—	
Thiotepa	-0.42896	-1.04442, 0.18651	0.2
Initial tumors			
One	—	—	
Two	-1.63128	-3.07277, -0.18979	0.027
Three	-0.05841	-0.80729, 0.69047	0.9
Four	-0.54990	-1.99364, 0.89384	0.5
Five	0.18483	-1.83346, 2.20312	0.9
Six	0.45218	-0.75353, 1.65789	0.5
seven	0.16302	-1.85083, 2.17688	0.9

From a proportional hazards model, we estimate that for two groups of study participants with the same size of initial bladder tumor but who differ in their enrollment to a control (placebo) or treatment arm (thiotepa),

the hazard ratio for first tumor recurrence is 0.65, with the control arm having greater estimated hazard (95% CI for hazard ratio: 0.35190, 1.20503).

From a proportional hazards, we estimate that for two groups of study participants that are in the same study arm with one group that had two centimeters of initial bladder tumors while the other had one centimeter of initial bladder tumors, the hazard ratio for first tumor recurrence is 0.20, with study participants with two centimeters of initial bladder tumors having lesser estimated hazard (95% CI for hazard ratio: 0.04629, 0.82714).

From a proportional hazards, we estimate that for two groups of study participants that are in the same study arm with one group that had three centimeters of initial bladder tumors while the other had one centimeter of initial bladder tumors, the hazard ratio for first tumor recurrence is 0.94, with study participants with three centimeters of initial bladder tumors having lesser estimated hazard (95% CI for hazard ratio: 0.13620, 2.44451).

From a proportional hazards, we estimate that for two groups of study participants that are in the same study arm with one group that had four centimeters of initial bladder tumors while the other had one centimeter of initial bladder tumors, the hazard ratio for first tumor recurrence is 0.58, with study participants with four centimeters of initial bladder tumors having lesser estimated hazard (95% CI for hazard ratio: 0.52184, 6.03504).

From a proportional hazards, we estimate that for two groups of study participants that are in the same study arm with one group that had five centimeters of initial bladder tumors while the other had one centimeter of initial bladder tumors, the hazard ratio for first tumor recurrence is 0.20, with study participants with five centimeters of initial bladder tumors having lesser estimated hazard (95% CI for hazard ratio: 0.15986, 9.05322).

From a proportional hazards, we estimate that for two groups of study participants that are in the same study arm with one group that had six centimeters of initial bladder tumors while the other had one centimeter of initial bladder tumors, the hazard ratio for first tumor recurrence is 1.57, with study participants with six centimeters of initial bladder tumors having higher estimated hazard (95% CI for hazard ratio: 0.47070, 5.24821).

From a proportional hazards, we estimate that for two groups of study participants that are in the same study arm with one group that had seven centimeters of initial bladder tumors while the other had one centimeter of initial bladder tumors, the hazard ratio for first tumor recurrence is 1.18, with study participants with seven centimeters of initial bladder tumors having higher estimated hazard (95% CI for hazard ratio: 0.15711, 8.81871).

Question 7 :

It is possible that the size of the largest tumor (pre-treatment) could be a potential confounder for time to recurrence in this study. As we have discussed in class confounding variable is a variable that is related to both the treatment group and the outcome variable, and could potentially influence the observed relationship between the two. In this case, the size of the largest tumor could be related to both the treatment group and the time to first recurrence. Patients with larger tumors may have a higher risk of recurrence (1), and they may also be more likely to be assigned to the thiotepa treatment arm if the randomization is not properly stratified by tumor size. It would be important to examine whether the distribution of tumor sizes is similar between the treatment groups. One way to do this would be to include tumor size as a covariate in a our model, along with treatment group and any other relevant variables.

Reference

1. Chang C, Chen J, Chang WY, Chiang AJ. Tumor Size Has a Time-Varying Effect on Recurrence in Cervical Cancer. J Low Genit Tract Dis. 2016 Oct;20(4):317-20. doi: 10.1097/LGT.0000000000000238. PMID: 27438585.

Code Appendix

```
### Setting up the packages
library(knitr)
knitr::opts_chunk$set(echo = FALSE)
# check if packages are installed; if not, install them
packages <- c("tidyverse", "readr", "ggExtra", "plotly",
              "ggplot2","ggstatsplot","ggside","rigr","nlme","lmtest",
              "sandwich","ggpubr")
not_installed <- setdiff(packages, rownames(installed.packages()))
if (length(not_installed)) install.packages(not_installed)

# load packages
library("MASS")
library(sandwich)
library(readr)
library(lmtest)
library(nlme)
library(ggstatsplot)
library(ggside)
library(gtsummary)
library(survival)
library(caret)
library(survminer)
library(rigr)
library(ggExtra)
library(broom)
library(plotly)
library(ggplot2)
library(table1)
library(tidyverse) # don't load tidyverse package due to conflict with dplyr
### ------------------------------------------------------------
#Loading working directory of the raw data

#Please load your data/directory by changing it with your work directory
#Throughout this code module you will see a tone of places, where
#data is read and written, so please make sure to change them to your
#working directory folder format

working_directory_data <- setwd("C:/Users/latera/Desktop/Bio_ass")

survi <- read_csv("Data/thiotepa.csv")

tte <- survi %>% with(Surv(stop, event))
sort(tte)
knitr::kable(tidy(survfit(tte ~ 1) )%>% slice(1:10),

      digits = 5,caption = "Summary of the 10 first observation")
rx_factor <- factor(survi$rx, levels = c(1, 2), labels = c("Placebo", "Thiotepa"))
survfit(tte ~ rx_factor, survi) %>%
 ggsurvplot(data = survi, conf.int=T, xlab = "Time (months)",
 ylab = "Estimated\nsurvival function",
 palette = c("#2E9FDF", "purple"))
```

```r
# Perform log-rank test
logrank_test <- survdiff(tte ~ rx_factor)

logrank_test
model <- coxph(tte ~ rx_factor, data=survi)

model %>%
  tbl_regression(
    label = list(

      rx_factor ~ "Treatement Arm"
    ),
    exponentiate = F,
    intercept = TRUE,
    estimate_fun = purrr::partial(style_number, digits = 5)
    )
ph0 <- coxph(tte ~ 1, data=survi)
ph1 <- coxph(tte ~ rx_factor, data=survi)
k <- anova(ph0, ph1, test="LRT")
knitr::kable(tidy(k), caption = "Summary of hypothesis testing")

survi$size <- factor(survi$size, levels = c(1, 2,3,4,5,6,7),
                     labels = c("One", "Two","Three","Four","Five",
                                "Six","seven"))

ph1 <- coxph(tte ~ rx_factor + size, data=survi)

ph1 %>%
  tbl_regression(
    label = list(

      rx_factor ~ "Treatement Arm",
      size ~ "Initial tumors"
    ),
    exponentiate = F,
    intercept = TRUE,
    estimate_fun = purrr::partial(style_number, digits = 5)
    )
```