Directives 4 and 7

Department of Biostatistics @ University of Washington

Alejandro Hernandez

Winter Quarter 2025

4. It is generally thought that aGVHD has an anti-leukemic effect. Based on the available data, is occurrence of aGVHD after transplantation associated with improved disease-free survival? Is it associated with a decreased risk of relapse? In view of this, do you consider aGVHD as an important prognostic event?

We fit two Cox proportional hazards models, one for disease-free survival and another for relapse-free survival. In both models, our target is the hazard ratio associated with the development of aGHVD, which in modeled as a time-varying covariate. In the second PH model, death is considered to be a competing event with relapse, because cancer relapse cannot occur after death, and modeling death as a censoring event may misrepresent relapse-free survival.

- For disease-free survival, the unadjusted hazard ratio associated with the occurrence of aGVHD is 1.32 (90% CI: 0.83-2.09). In other words, comparing two patients, one who did develop aGHVD during the study period and another who did not, the instantaneous risk of death or relapse for the first patient is 1.32 greater than times the second, throughout the study period.
- For relapse-free survival, the unadjusted ratio associated with the occurrence of aGVHD is 0.70 (90% CI: 0.33-1.47).

Both confidence interval for the hazard ratio ranges above and below 1, therefore we cannot claim with certainty that the sample has sufficient evidence that aGVHD is associated with better or worse outcomes.

7. Based on the available data, is recovery of normal platelet levels associated with improved disease-free survival? Is it associated with a decreased risk of relapse?

We fit two Cox proportional hazards models, one for disease-free survival and another for relapse-free survival. In both models, our target is the hazard ratio associated with platelet recover, which in modeled as a time-varying covariate. In the second PH model, death is considered to be a competing event with relapse, because cancer relapse cannot occur after death, and modeling death as a censoring event may misrepresent relapse-free survival.

- For disease-free survival, the unadjusted ratio associated with platelet recovery is 0.34 (90% CI: 0.20-0.57).
- For relapse-free survival, the unadjusted ratio associated with platelet recovery is 0.87 (90% CI: 0.37-2.01). In other words, comparing two patients, one whose platelet count returned to a normal level during the study period and another who did not, the instantaneous risk of relapse for the first patient is 0.37-2.01 times the second, throughout the study period.

We can claim with certainty that the sample has sufficient evidence that platelet recovery is positively associated with disease-free survival. The latter confidence interval for the hazard ratio ranges above and below 1, therefore we cannot claim with certainty that the sample has sufficient evidence that platelet is associated with better or worse relapse-free survival.

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 models
library(dplyr)     # data manipulation
library(broom)     # combine and reshape model output
library(ggplot2)   # data visualization

theme_set(theme_bw())

# Load data
bmt <- read.csv("../data/bmt.csv")
dim(bmt) # 137 rows, 22 columns
names(bmt)

# Handle missing data (there is none)
anyNA(bmt)

## Helper function to extract model results
get_robustci <- function (model, var=NULL, alpha=0.1) {
  # Organize model results
  tidy_model <- broom::tidy(model)
  # If desired, filter the results to a specific variable
  if (!is.null(var)) tidy_model = tidy_model %>% dplyr::filter(term==var)
  coef = tidy_model$estimate
  std.error = tidy_model$std.error
  robust.se = tidy_model$robust.se
  crit.val = qnorm(alpha/2, lower.tail=F)

  res <- list(
    term = tidy_model$term,
    # Exponentiate coefficient
    estimate = exp(coef),
    std.error = std.error,
    robust.se = robust.se,
    # Robust confidence interval
    conf.low = exp(coef - crit.val*robust.se),
    conf.high = exp(coef + crit.val*robust.se)
    )

  return(res)
}

## Restructing dataframe to including indicators of competing risks and
## time-varing covariates
```

```r
tbmt <- survival::tmerge(
  data1 = bmt, data2 = bmt, id = id,
  # Death
  death = event(ts, deltas),
  # Relapse
  relapse = event(tdfs, deltar),
  # Death or relapse
  death.relapse = event(tdfs, deltadfs),
  # Occurrence of aGVHD
  post.agvhd = tdc(ta),
  # Occurrence of platelet recovery
  post.recovery = tdc(tp)
  )

# Calculate survival times for distinct end points
death.relapse_surv <- with(tbmt, survival::Surv(tstart, tstop, death.relapse))
relapse_surv <- with(tbmt, Surv(tstart, tstop, relapse))

######################
#### DIRECTIVE 4 ####
######################

## Investigating acute graft-versus-host disease (aGVHD)
# HR of disease-free survival associated with the occurrence of aGVHD
coxfit1 <- survival::coxph(death.relapse_surv ~ post.agvhd + cluster(id),
                           data=tbmt)

tidy(coxfit1, conf.int=TRUE, conf.level=0.90, exponentiate=TRUE)[-c(5,6)]
get_robustci(coxfit1, "post.agvhd")[-1] %>% sapply(round, 4)

# HR of relapse-free survival associated with the occurrence of aGVHD
coxfit2 <- coxph(relapse_surv ~ post.agvhd + cluster(id), data=tbmt)

tidy(coxfit2, conf.int=TRUE, conf.level=0.90, exponentiate=TRUE)[-c(5,6)]
get_robustci(coxfit2, "post.agvhd")[-1] %>% sapply(round, 4)

## Diagnostics

# Martingale residuals (form of patient age)
coxfit <- coxph(death.relapse_surv ~ post.agvhd + age + cluster(id), data=tbmt)
coxfit2 <- coxph(death.relapse_surv ~ post.agvhd + cluster(id), data=tbmt)
mgresid1 <- residuals(coxfit, type="martingale")
mgresid2 <- residuals(coxfit2, type="martingale")
agevals <- tbmt$age

par(mfrow=c(1,2)) # set to print two plots side-by-side
plot(agevals, mgresid2,
     xlab="age", ylab="martingale residuals", # ylim=c(-1,1),
     main="age variable not included")
mgresid1.loess = loess(mgresid1 ~ agevals, degree=1)
lines(sort(agevals), predict(mgresid1.loess, sort(agevals)), col=2, lwd=2)
abline(h=0, lty=3)
```

```r
plot(agevals, mgresid1,
     xlab="age", ylab="", main="age variable included")
mgresid1.loess = loess(mgresid1 ~ agevals, degree=1)
lines(sort(agevals), predict(mgresid1.loess, sort(agevals)), col=2, lwd=2)
abline(h=0, lty=3)
par(mfrow=c(1,1)) # reset to print a single plot

# Deviance residuals (identifying outliers)
coxfit <- coxph(death.relapse_surv ~ post.agvhd + age + cluster(id), data=tbmt)
devresid = residuals(coxfit, type="deviance")
plot(1:coxfit$n, devresid, xlab="observation", ylab="deviance residuals")
abline(h=0, lty=3)

# Schoenfeld residuals (assessing proportional hazards)
coxfit <- coxph(death.relapse_surv ~ cmv + age + cluster(id), data=tbmt)
schoenresid = residuals(coxfit, type="scaledsch")
times = as.numeric(rownames(schoenresid))

plot(times, schoenresid[,1], xlab="time", ylab="scaled Schoenfeld residuals")
schoenresid.loess = loess(schoenresid[,1] ~ times, degree=1)
lines(unique(times), predict(schoenresid.loess, unique(times)), col=2, lwd=2)
abline(h=0, lty=3)

# Assessing proportional hazards of a time-varying covariate
coxfit.tv = coxph(death.relapse_surv ~ tt(post.agvhd) + post.agvhd,
                  tt = function(x,t,...) x*t,
                  data=tbmt)
summary(coxfit.tv)$coef

######################
#### DIRECTIVE 7 ####
######################

## Investigating return of platelet counts to a normal level
# HR of disease-free survival associated with platelet recovery
coxfit <- coxph(death.relapse_surv ~ post.recovery + cluster(id),
                data=tbmt)
tidy(coxfit, conf.int=TRUE, conf.level=0.90, exponentiate=TRUE)[-c(5,6)]
get_robustci(coxfit, "post.recovery")[-1] %>% sapply(round, 4)


# HR of relapse-free survival associated with platelet recovery
coxfit <- coxph(relapse_surv ~ post.recovery + cluster(id),
                data=tbmt)
tidy(coxfit, conf.int=TRUE, conf.level=0.90, exponentiate=TRUE)[-c(5,6)]
get_robustci(coxfit, "post.recovery")[-1] %>% sapply(round, 4)

## Diagnostics
```

End of document.