Heart Transplant Survival Modeling

Vidhi Patel

Background

This paper is an effort to recreate a select number of analyses conducted by John Crawley and Marie Hu in the Covariance Analysis of Heart Transplant Survival Data. Crawley and Hu sought to assess the effects of various covariates on the survival of patients in the Stanford Heart Transplantation Program. Admission to this program follows a rigorous evaluation process, involving a medical consultation where alternative therapeutic options are deemed ineffective for the patient's condition. Donor matching for program participants involves various critieria, including blood type compatibility and specific health concerns may cause patient deferrment from the program. A match can take anywhere from a few days to multiple years to happen. Some patients may die before a suitable match is found. Because of this, a transplant is a time-varying process.

The dataset used encompasses a comprehensive array of patient-specific variables alongside survival outcomes. The dataset's structure encompasses variables related to patients' survival times, including admission date, last seen date, and, for transplant recipients, transplantation date. Survival time is calculated as the difference between the last seen date and admission date, with censoring occurring when the last seen date is either the date of death or the study's closing date. For transplant recipients, waiting time for transplantation and post-transplant survival are also recorded.

Each patient's profile includes a vector of covariates encompassing transplant status (Z_0), waiting time to transplant (Z_1), calendar time at transplant (Z_2), age at program acceptance (Z_3), age at transplant (Z_4), pretransplant risk (Z_5), post-transplant risk (Z_6), and measures of tissue type mismatching (Z_7. Z_8 , Z_9). Age at acceptance and transplant are included to address age-related risks and surgical response variations. Calendar time at transplant is examined to identify programmatic improvements over time, while waiting time to transplant is evaluated for comparison. Variables concerning previous open-heart surgery and tissue type mismatching, notably HLA antigens, are investigated for their potential influence on rejection and survival. Post-transplant risk is a combination of the post_transplant risk and transplant status.

Transplant status is an indicator variable. Due to the time-varying nature of transplants, along with the fact that patients may die before recieving it, treatment should be considered time-varying. Otherwise, we run the risk of assigning unproportionally to the control group. Pretransplant risks, age at acceptance, and previous surgery do not change with time. All other covariates are time-varying.

Model Description

The Cox proportional hazards (PH) model with time-varying covariates is an extension of the classic Cox PH model, which estimates the hazard function $h(t|x)$ as $h(t|x) = h_0(t) \times \exp(\beta_1 x_1 + \beta_2 x_2 + ... + \beta_p x_p)$, where $h_0(t)$ is the baseline hazard function, x_i represents the covariates, and β_i are the corresponding coefficients. In the time-varying model, covariates can change over time, denoted as $x_i(t)$, yielding the hazard function $h(t|x(t)) = h_0(t) \times \exp(\beta_1 x_1(t) + \beta_2 x_2(t) + ... + \beta_p x_p(t))$. This adaptation allows for the incorporation of time-varying effects into the model, providing a more accurate representation of the hazard ratio dynamics throughout the study period. The Breslow tie-handling option is often utilized in Cox PH models to handle tied event times, where multiple subjects experience the event simultaneously. This option assigns weights to tied events, adjusting the likelihood contribution of tied observations to the model fitting process. Incorporating this option ensures more robust parameter estimates and accurate inference.

In this paper, we analysed three different Cox PH models with differing covariates from the dataset, using Breslow tie-handling. The hazard functions for the three models are as follows:

- Line 5: $h(t|x) = h_0(t) \times \exp(\beta_0 Z_0 + \beta_1 Z_4)$
- Line 6: $h(t|x) = h_0(t) \times \exp(\beta_0 Z_0 + \beta_1 Z_3 + \beta_2 Z_4)$
- Line 7: $h(t|x) = h_0(t) \times \exp(\beta_0 Z_0 + \beta_1 Z_4 + \beta_2 Z_5)$

where: - $(h_0(t))$ is the baseline hazard function. - (Z_i) represents the corresponding time-independent covariate - $beta_i$ are the coefficients representing the effects of Z_i on the hazard function

Estimation Results

From Table 1, it is apparent that in the model from line 5, that age at transplant exhibits statistically significant effects on the hazard of death, with p-values of 0.015 respectively. However, in the model from line 6, the age of acceptance into the program and age at transplant do not, with p-values of 0.156 and 0.144 respectively. This difference between line 5 model and line 6 model indicates that age of acceptance as a pretransplant risk hich is largely explained by age at transplant as a differential post-transplant risk. This is likely because pre-transplant age has little effect of risk, however because surgery and transplant rejection complications

Line	Deviance	Parameter	Z_0	Z_3	Z_4	Z_5
5	589.8	$\hat{\beta}$	-2.40046214		0.05305187	
		$SE(\hat{\beta})$	1.11488031		0.02187030	
		p-value	0.03130988		0.01527679	
6	589.4	$\hat{\beta}$	-1.91707112	0.01183773	0.04140268	
		$SE(\hat{\beta})$	1.35216928	0.01832617	0.02837660	
		p-value	0.15625661	0.51831375	0.14455319	
7	585.0	$\hat{\beta}$	-2.16576097		0.04878214	-0.7300459
		$SE(\hat{\beta})$	1.05631930		0.02067282	0.36058947
		p-value	0.04033613		0.01813513	0.04338174

Table 1: Analysis of the 103 Patients as Table 2 of Crowley and Hu (1977). Z_0 is the time dependent indicator of transplant. Z_3 is age of acceptance into the program in years. Z_4 is age at transplant in years. Z_5 is the history of open-heart surgery indicator. P-value (two-sided) of wald test is provided

| Line	Age at Transplant	Estimation	95% Confidence Interval
7	20.00000	0.3057029	(0.03310819, 2.822693)
	30.00000	0.4979179	(0.04515624, 5.490320)
	40.00000	0.8109906	(0.05853001, 11.237069)
	44.29458	1.0000000	(0.06461043, 15.477377)
	50.00000	1.3209123	(0.07296951, 23.911486)
	60.00000	2.1514543	(0.08835770, 52.386558)

Table 2: Relative Risk at Transplant $exp(\hat{\beta}_0 + \hat{\beta}_4 Z_4)$ for Line 7

can be harder to survive in elder patients, post-transplant age can impact risk. Moreover, in the model from line 7, both age at transplant and history of open-heart surgery indicator also show significance with a p-values less than 0.05. These findings suggest that history of open-heart surgery (as an indicator of pretransplant risk) plays a crucial role in predicting the hazard of death among the patients. By comparsion of the deviances, which acts as a measure of goodness-of-fit, it appears that the line 7 model is best as it has the lowest deviance.

Table 2 presents estimates of relative risk at transplant for different age groups, based on the model from line 7. It appears that older age at transplant is associated with higher hazards of death, with individuals aged 60 years exhibiting approximately twice the risk compared to the reference group. The widening confidence intervals with increasing age underscores the need for cautious interpretation due to the uncertainty inherent in estimating relative risk in older age groups.

Diagnostic Results

Figure A: Estimate of Survival Curve from Residuals
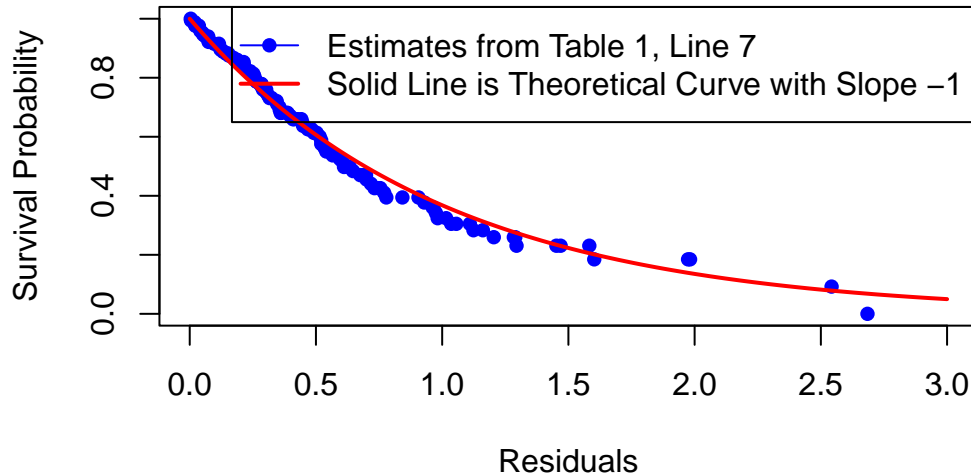
As a diagnostic measure, we estimate a survival curve using the residuals from the Table 1, Line 7 model. By leveraging these residuals to estimate a survival curve, we gain valuable insights into the adequacy of the model's fit to the observed data. The survival probabilities result in a curve close to e^{-t}. This concordance between the observed survival probabilities and the expected exponential decay suggests that the model provides a reasonable approximation of the survival process among heart transplant patients.

Appendix

Data Processing

```
library(data.table)
library(survival)
library(tidyverse)
dt <- data.table(jasa)
# add 0.5 day to those died on day 0
dt[, futime:= pmax(0.5,futime)]
```

4

```
# remove 0.5 day in wait.time for those who had the transplant
# and death at the same date
dt[wait.time==futime & !is.na(wait.time), wait.time:= (wait.time -0.5)]
# add ID as patient unique unique identifiers
dt[,id:= (1:nrow(dt))]


dt$accept.age = ifelse(dt$transplant == 1,
                       as.numeric(dt$accept.dt - dt$birth.dt) / 365.25, dt$age)


dt$tx.age =  as.numeric(dt$tx.date - dt$birth.dt) / 365.25
dt$tx.age[is.na(dt$tx.age)] = 0


dt$tx.cal = dt$tx.date - as.Date("1967-10-01")
#dt$tx.cal[is.na(dt$tx.cal)] = 0


s.dt = tmerge(dt, dt, id=id,
              death = event(futime, fustat),
              tx = tdc(wait.time))


# transplant accept.age, surgery, mismatch, hla.a2, mscore are fixed
# tx.cal, wait.time, tx.age are time independent


s.dt1 <- s.dt %>%
  select(id, tx, tstart, tstop, wait.time, tx.cal, accept.age, tx.age,
         surgery, mismatch, hla.a2, mscore, reject, death) %>%
  rename(z_0 = tx, z_1 = wait.time, z_2 = tx.cal, z_3 = accept.age,
         z_4 = tx.age, z_5 = surgery, z_7 = mismatch, z_8 = hla.a2, z_9 = mscore)
```

Table 1

```
line5 = coxph(Surv(tstart, tstop, death) ~ z_0 + z_4:z_0,
              data=s.dt1, ties="breslow")
line6 = coxph(Surv(tstart, tstop, death) ~ z_0 + z_3 + z_4:z_0,
              data=s.dt1, ties="breslow")
line7 = coxph(Surv(tstart, tstop, death) ~ z_0 + z_4:z_0 + z_5,
              data=s.dt1, ties="breslow")


coef5 <- summary(line5)$coefficients[, 1]
```

```r
se5 <- summary(line5)$coefficients[, 3]
p_value5 <- summary(line5)$coefficients[, 5]


coef6 <- summary(line6)$coefficients[, 1]
se6 <- summary(line6)$coefficients[, 3]
p_value6 <- summary(line6)$coefficients[, 5]

coef7 <- summary(line7)$coefficients[, 1]
se7 <- summary(line7)$coefficients[, 3]
p_value7 <- summary(line7)$coefficients[, 5]

coef_table1 <- matrix(NA, nrow = 3, ncol = 4)
rownames(coef_table1) <- c(" ^", "SE( ^)", "p-value")
colnames(coef_table1) <- c("Z0", "Z3", "Z4", "Z5")

# Create a data frame to store the results
coef_table1[" ^", ] = c(coef5['z_0'], NA , coef5['z_0:z_4'], NA)
coef_table1["SE( ^)", ] = c(se5['z_0'], NA , se5['z_0:z_4'], NA)
coef_table1["p-value", ] = c(p_value5['z_0'], NA , p_value5['z_0:z_4'], NA)


coef_table2 <- matrix(NA, nrow = 3, ncol = 4)
rownames(coef_table2) <- c(" ^", "SE( ^)", "p-value")
colnames(coef_table2) <- c("Z0", "Z3", "Z4", "Z5")

coef_table2[" ^", ] = c(coef6['z_0'], coef6['z_3'] , coef6['z_0:z_4'], NA)
coef_table2["SE( ^)", ] = c(se6['z_0'], se6['z_3'] , se6['z_0:z_4'], NA)
coef_table2["p-value", ] = c(p_value6['z_0'], p_value6['z_3'] , p_value6['z_0:z_4'], NA)

coef_table3 <- matrix(NA, nrow = 3, ncol = 4)
rownames(coef_table3) <- c(" ^", "SE( ^)", "p-value")
colnames(coef_table3) <- c("Z0", "Z3", "Z4", "Z5")

coef_table3[" ^", ] = c(coef7['z_0'], NA , coef7['z_0:z_4'], coef7['z_5'])
coef_table3["SE( ^)", ] = c(se7['z_0'], NA , se7['z_0:z_4'], se7['z_5'])
coef_table3["p-value", ] = c(p_value7['z_0'], NA , p_value7['z_0:z_4'], p_value7['z_5'])

dev5 = round(line5$loglik[2]*(-2), 1)
dev6 = round(line6$loglik[2]*(-2), 1)
dev7 = round(line7$loglik[2]*(-2), 1)
```

```r
# Stack tables on top of each other
stacked_table <- as.data.frame(rbind(coef_table1, coef_table2, coef_table3))
rownames(stacked_table) <- NULL
stacked_table$line <- c(5, 5, 5, 6, 6, 6, 7, 7, 7)
stacked_table$deviance <- c(dev5, dev5, dev5, dev6, dev6, dev6, dev7, dev7, dev7)
print(stacked_table)
```

Table 2

```r
coefficients <- coef(line7)
std_errors <- summary(line7)$coefficients[, 3]

age_relative_risk_1 <- (0 - coefficients[1]) / coefficients[3]
line_age <- c(20, 30, 40, age_relative_risk_1, 50, 60)

# Calculating the standard error of the relative risk using the delta method
rel_risk_std_errors <- sqrt((std_errors[1])^2 + (line_age * std_errors[3])^2)

# calculating confidence intervals for the relative risk
lower_ci <- exp(coefficients[1] + coefficients[3]
                * line_age -  qnorm(0.975) * rel_risk_std_errors)
upper_ci <- exp(coefficients[1] + coefficients[3] * line_age
                +  qnorm(0.975) * rel_risk_std_errors)

# Creating a table
result_table <- data.frame(
   Line = 7,
   `Age at Transplant` = line_age,
   Estimation = exp(coefficients[1] + coefficients[3] * line_age),
   `Lower CI` = lower_ci,
   `Upper CI` = upper_ci
)

print(result_table)
```

Figure A

```r
line7 = coxph(Surv(tstart, tstop, death) ~ z_0 + z_4:z_0 + z_5,
              data=s.dt1, ties="breslow")
resid = residuals(line7)
fitted = s.dt1$death - resid

plotdt = survfit(Surv(fitted, s.dt1$death)~1)
surv_prob <- plotdt$surv
trans_time <-  plotdt$time

# Plot dots for each coordinate
plot(trans_time, surv_prob, type = "p", pch = 16,
     xlim =c(0, 3), col = "blue", xlab = "Residuals",
     ylab = "Survival Probability",
     main = "Figure A: Estimate of Survival Curve from Residuals")

# Add e^-t line
t <- seq(0, 3, length.out = 100)  # Range of residuals
y <- exp(-t)  # Calculate e^-t values
lines(t, y, col = "red", lwd = 2)  # Add line to the plot

legend("topright", legend = c("Estimates from Table 1, Line 7",
                              "Solid Line is Theoretical Curve with Slope -1"),
       col = c("blue", "red"), lty = c(1, 1),
       lwd = c(1, 2), pch = c(16, NA))
```