|  |  |  |
| --- | --- | --- |
| Assignment 2 | August 6  2019 | |
| Paul-Willem Janse van Rensburg - 15338673 | | Survival Analysis |

Table of Contents

[Question 1 2](#_Toc15482114)

[Question 2 4](#_Toc15482115)

[Question 3 5](#_Toc15482116)

[Appendix A 6](#_Toc15482117)

# Question 1

Below is a table summarizing the survival function estimates, , along with their standard errors. As we are estimating the survival function, we employed the Kaplan-Meier method as it estimates the survival function directly (as opposed to the cumulative hazard rate, which would require a conversion)

For the surgical placement, we have:

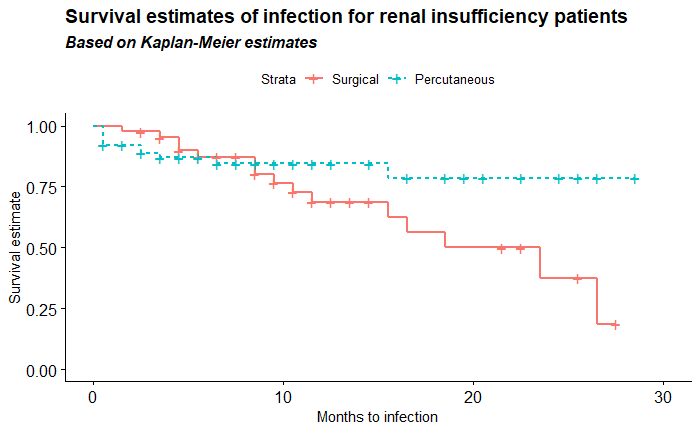
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Event Time | # at risk | # of events | # of cenosred | est. S(t) | Std Error |
| 1 | 1.5 | 43 | 1 | 0 | 0.97674 | 0.02353 |
| 2 | 2.5 | 42 | 0 | 2 | 0.97674 | 0.02353 |
| 3 | 3.5 | 40 | 1 | 3 | 0.95233 | 0.03456 |
| 4 | 4.5 | 36 | 2 | 1 | 0.89942 | 0.05319 |
| 5 | 5.5 | 33 | 1 | 1 | 0.87216 | 0.06145 |
| 6 | 6.5 | 31 | 0 | 2 | 0.87216 | 0.06145 |
| … | … | … | … | … | … | … |
| 8 | 8.5 | 25 | 2 | 1 | 0.80239 | 0.08517 |
| … | … | … | … | … | … | … |
| 18 | 21.5 | 8 | 0 | 2 | 0.49978 | 0.22228 |
| 19 | 22.5 | 6 | 0 | 2 | 0.49978 | 0.22228 |
| 20 | 23.5 | 4 | 1 | 0 | 0.37484 | 0.36434 |
| 21 | 25.5 | 3 | 0 | 1 | 0.37484 | 0.36434 |
| 22 | 26.5 | 2 | 1 | 0 | 0.18742 | 0.79545 |
| 23 | 27.5 | 1 | 0 | 1 | 0.18742 | 0.79545 |

For the percutaneous catheter placement:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Event Time | # at risk | # of events | # of cenosred | est. S(t) | Std Error |
| 24 | 0.5 | 76 | 6 | 10 | 0.92105 | 0.03358 |
| 25 | 1.5 | 60 | 0 | 4 | 0.92105 | 0.03358 |
| 26 | 2.5 | 56 | 2 | 5 | 0.88816 | 0.0423 |
| 27 | 3.5 | 49 | 1 | 5 | 0.87003 | 0.04706 |
| 28 | 4.5 | 43 | 0 | 3 | 0.87003 | 0.04706 |
| 29 | 5.5 | 40 | 0 | 5 | 0.87003 | 0.04706 |
| … | … | … | … | … | … | … |
| 32 | 8.5 | 30 | 0 | 3 | 0.84517 | 0.05527 |
| … | … | … | … | … | … | … |
| 42 | 20.5 | 7 | 0 | 1 | 0.7848 | 0.09246 |
| 43 | 22.5 | 6 | 0 | 1 | 0.7848 | 0.09246 |
| 44 | 24.5 | 5 | 0 | 1 | 0.7848 | 0.09246 |
| 45 | 25.5 | 4 | 0 | 1 | 0.7848 | 0.09246 |
| 46 | 26.5 | 3 | 0 | 2 | 0.7848 | 0.09246 |
| 47 | 28.5 | 1 | 0 | 1 | 0.7848 | 0.09246 |

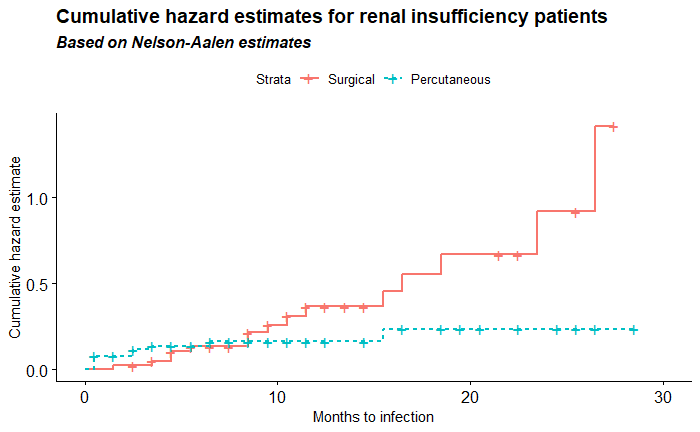
We can see that the survival estimate is strictly decreasing, which is as expected, with the error increasing as we move along the y-axis. Looking at 8.5 months, we can see the estimated survival for the percutaneous placement at 0.84517 with a standard deviation of 0.05527 is higher than that of the surgical placement estimate at 0.80239 with a standard error of 0.08517. From there on the estimates diverge even further, with the percutaneous techniques estimated survival rate remaining above that of surgical techniques.

# Question 2



It seems initially that the estimated survival rate for the surgical placement is the better technique for delaying infection, but the percutaneous technique’s survival estimate overtakes the surgical technique at around about 8.5 months and remains relatively stable thereafter. This tells us that the surgical technique for catheter placement is less likely to cause infection initially (within the first 8.5 months), however the overall likelihood of infection is lower with the percutaneous technique.

# Question 3



For the surgical treatment, we crudely estimate the hazard rate at 5 months as following:

For the percutaneous treatment, we crudely estimate the hazard rate at 5 months as following:

# Appendix A

library(dplyr)

library(data.table)

library(readr)

library(survival)

library(survminer)

data <- read\_delim('H:/Werk/Survival Analysis/Data sets/Section1\_4.dat',

delim = ' ', col\_names = c('Time\_To\_Infection','Censored','Treatment'),

col\_types = cols(Time\_To\_Infection = 'n', Censored = 'n', Treatment = 'n'))

kmp<-survfit(Surv(data$Time\_To\_Infection,data$Censored)~data$Treatment,type="kaplan-meier")

df\_summary <- fortify(kmp)

df\_summary\_surg <- df\_summary[df\_summary$strata == 1,]

df\_summary\_perc <- df\_summary[df\_summary$strata == 2,]

df\_surg\_head\_tail <- data.frame(head(df\_summary\_surg))

df\_perc\_head\_tail <- data.frame(head(df\_summary\_perc))

temp <- data.frame('...','...','...','...','...','...','...','...','...')

names(temp) <- names(df\_summary)

highlight\_surg <- df\_summary\_surg[df\_summary\_surg$time == 8.5,]

highlight\_perc <- df\_summary\_perc[df\_summary\_perc$time == 8.5,]

df\_surg\_head\_tail <- rbind(df\_surg\_head\_tail, temp, highlight\_surg, temp,tail(df\_summary\_surg))

df\_perc\_head\_tail <- rbind(df\_perc\_head\_tail, temp, highlight\_perc, temp,tail(df\_summary\_perc))

drops <- c('upper','lower','strata')

df\_surg\_head\_tail <- df\_surg\_head\_tail[ , !(names(df\_surg\_head\_tail) %in% drops)]

df\_perc\_head\_tail <- df\_perc\_head\_tail[ , !(names(df\_perc\_head\_tail) %in% drops)]

names(df\_surg\_head\_tail) <- c('Event Time','# at risk','# of events','# of cenosred','est. S(t)','Std Error')

names(df\_perc\_head\_tail) <- c('Event Time','# at risk','# of events','# of cenosred','est. S(t)','Std Error')

print(kable(df\_surg\_head\_tail))

print(kable(df\_perc\_head\_tail))

ggsurv <- ggsurvplot(kmp , data = data, xlab="Months to infection",

ylab="Survival estimate",

legend.labs = c("Surgical", "Percutaneous"), linetype = 'strata') +

labs(title = "Survival estimates of infection for renal insufficiency patients",

subtitle = "Based on Kaplan-Meier estimates")

ggsurv <- ggpar(ggsurv, font.title = c(14, "bold"),

font.subtitle = c(12, "bold.italic"),

font.x = c(11, "plain"),

font.y = c(11, "plain"))

print(ggsurv)

kmp\_na <-survfit(Surv(data$Time\_To\_Infection,data$Censored)~data$Treatment,type="fleming-harrington")

ggsurv\_ch <- ggsurvplot(kmp\_na, data = data, xlab="Months to infection",

ylab="Cumulative hazard estimate",

legend.labs = c("Surgical", "Percutaneous"),

linetype = 'strata', fun="cumhaz") +

labs(title = "Cumulative hazard estimates for renal insufficiency patients",

subtitle = "Based on Nelson-Aalen estimates")

ggsurv\_ch <- ggpar(ggsurv\_ch,

font.title = c(14, "bold"),

font.subtitle = c(12, "bold.italic"),

font.x = c(11, "plain"),

font.y = c(11, "plain"))

print(ggsurv\_ch)

df\_na\_summary <- fortify(kmp\_na)

cumhaz\_45 <- kmp\_na$cumhaz[(kmp\_na$time == 4.5) & (df\_na\_summary$strata == 1)]

cumhaz\_55 <- kmp\_na$cumhaz[(kmp\_na$time == 5.5) & (df\_na\_summary$strata == 1)]

print(cumhaz\_45)

print(cumhaz\_55)

hazard\_estimate\_surg <- cumhaz\_45 \* 0.5 + cumhaz\_55 \* 0.5

print(hazard\_estimate\_surg)

cumhaz\_35 <- kmp\_na$cumhaz[(kmp\_na$time == 3.5) & (df\_na\_summary$strata == 2)]

cumhaz\_65 <- kmp\_na$cumhaz[(kmp\_na$time == 6.5) & (df\_na\_summary$strata == 2)]

print(cumhaz\_35)

print(cumhaz\_65)

hazard\_estimate\_perc <- cumhaz\_35 \* 1.5/3 + cumhaz\_65 \* 1.5/3

print(hazard\_estimate\_perc)