Survival Analysis for Colon Cancer Patients

Shane Crinion

20/10/2018

Survival Analysis Study

401 patients with stage B and C cancer were given a treatment of either placebo, levamisole, flourouacil (5-FU) or combination in this study by Laurie JA et. al (DOI: 10.1200/JCO.1989.7.10.1447). The data compiled by this study is used to perform survival analysis on the survival time of patients. The administration of the drug was over 1 year and results were compiled over the 5 year study effect period.

The survival function is used to determine the probability that a patient will survive past a given time. In this example, the diagnosed individual is examined for survival time following drug application. To perform and assess this, the Kaplan-Meier estimatior is used.

The presence of 4 or more sentinel lymph nodes is associated with the presence and prognosis of a tumour. These are the first lymph nodes that cancer is likely to spread meaning the more proximal sentinel lymph nodes, the more likely it is that the cancer will spread. The assesment involves injection of blue dye to find the lymph node and then uses a sample from this to test for cancer cells.

The data is opened and inspected.

The data from the analysis is stored and extracted from the previously annotated "colon_data.csv" file.

```
#data read into the file
colon_data <- read.csv("colon_data.csv")
#look at find stats for this
attach(colon_data)
head(colon_data)
```

```
##        rx sex status node4 time etype
## 1 Lev+5FU   1      1     1  968     1
## 2 Lev+5FU   1      0     0 3087     1
## 3 Lev+5FU   0      1     1  245     1
## 4 Lev+5FU   0      1     1  904     1
## 5 Lev+5FU   0      0     0 3308     1
## 6 Lev+5FU   1      0     0 3309     1
```

```
summary(colon_data)
```

```
##        rx           sex            status           node4
##  Lev+5FU:304   Min.   :0.000   Min.   :0.0000   Min.   :0.0000
##  Obs    :315   1st Qu.:0.000   1st Qu.:0.0000   1st Qu.:0.0000
##                Median :0.000   Median :0.0000   Median :0.0000
##                Mean   :0.496   Mean   :0.4782   Mean   :0.2682
##                3rd Qu.:1.000   3rd Qu.:1.0000   3rd Qu.:1.0000
##                Max.   :1.000   Max.   :1.0000   Max.   :1.0000
##       time          etype
##  Min.   :   8   Min.   :1
##  1st Qu.: 407   1st Qu.:1
##  Median :1800   Median :1
##  Mean   :1450   Mean   :1
```

```
##  3rd Qu.:2296   3rd Qu.:1
##  Max.   :3309   Max.   :1
```

The file is read as a common separated file. The headings are attached for ease when selecting components later. Head is used to inspect formatting. Summary is used to learn more about the values. Summary indicates many 0/1 values which correlates with the binary formation. From the file columns, the sex, drug status, presence/lack of 4 lymph nodes and time in days until event can be determined.

The necessary packages are then installed for the analysis.

Splines and survival are used.

```
library(splines)
library(survival)
```

The packages "splines" and "survival" are installed using install.packages() in the console. "splines" is used for regression modelling and "survival" is used for created the Kaplan-Meier estimator.

Assessments to be performed using survival analysis:

(1) The performance of the adjuvant therapy against the placebo over the duration of the trial:

This is performed by creating a Kaplan-Meier plot.

```
log.rank = survdiff(Surv(colon_data$time,colon_data$status)~colon_data$rx)
log.rank
```

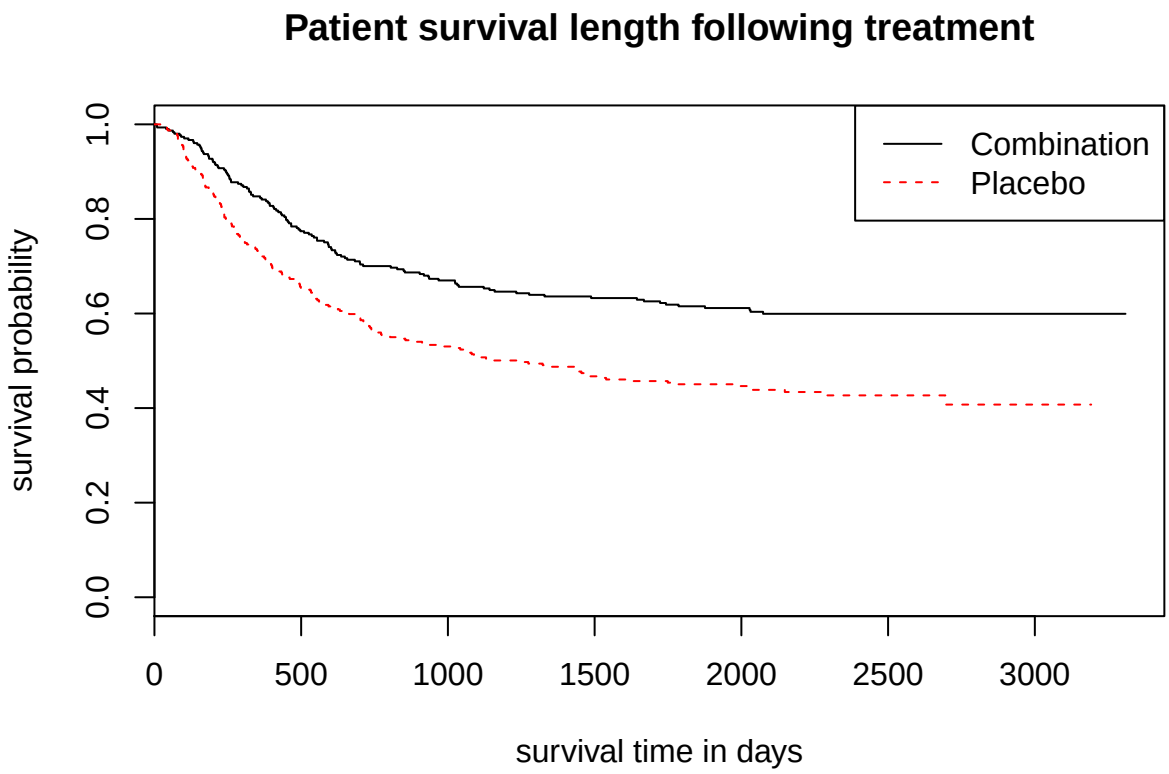
```
## Call:
## survdiff(formula = Surv(colon_data$time, colon_data$status) ~
##     colon_data$rx)
##
##                          N Observed Expected (O-E)^2/E (O-E)^2/V
## colon_data$rx=Lev+5FU 304      119      156      8.96      19.1
## colon_data$rx=Obs     315      177      140     10.05      19.1
##
##  Chisq= 19.1  on 1 degrees of freedom, p= 1e-05
```

```
sfit <- survfit(Surv(colon_data$time,colon_data$status)~colon_data$rx)

plot(sfit,
     lty=c("solid","dashed"),
     col=c("black","red"),
     xlab="survival time in days",
     ylab="survival probability")

title("Patient survival length following treatment")

legend("topright",
c("Combination","Placebo"),
col=c("black","red"),
lty=c("solid","dashed"))
```



survdiff() is used to test the difference between two or more survival curves (time, status) and the predictor (rx). Surv() creates a survival object to use as the response variable in the model. The log ratio values indicate 51.1% survival at study completion. The plot indicates the the drug combination increases the survival of the individual. The horizontal change indicates that it takes longer for the event to occur (greater survival). From this, we can reject the null hypothesis that there is no difference in surival. The drug increases survival length and probability vs. the placebo.

---

(2) The effect of gender on survival within the placebo, and in the treated cohorts

This is tested by changing the predictor variable in surv() function to sex.

```
par(mfrow=c(1,2))
#subset by placebo first

colon.subs <- colon_data
lev.sub <- subset(colon.subs,colon.subs$rx=="Lev+5FU")
obs.sub <- subset(colon.subs,colon.subs$rx=="Obs")

log.rank1 = survdiff(Surv(obs.sub$time,obs.sub$status)~obs.sub$sex)
log.rank1
```

```
## Call:
## survdiff(formula = Surv(obs.sub$time, obs.sub$status) ~ obs.sub$sex)
##
##                 N Observed Expected (O-E)^2/E (O-E)^2/V
```

```
## obs.sub$sex=0 149       81     82.8    0.0398    0.0749
## obs.sub$sex=1 166       96     94.2    0.0350    0.0749
##
##  Chisq= 0.1  on 1 degrees of freedom, p= 0.8
```

```
ofit <- survfit(Surv(obs.sub$time,obs.sub$status)~obs.sub$sex)

plot(ofit,
     lty=c("solid","dashed"),
     col=c("black","blue"),
     xlab="survival time in days",
     ylab="survival probability")

title("Survival M vs. F (placebo)")

legend("topright",
c("Female","Male"),
lty=c("solid","dashed",
col="black","blue"))

###now for those with the drug

log.rank2 = survdiff(Surv(lev.sub$time,lev.sub$status)~lev.sub$sex)
log.rank2
```

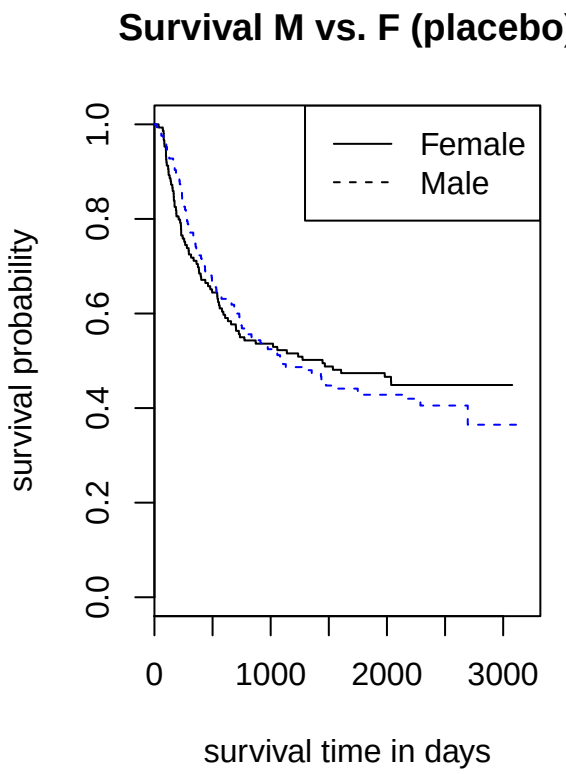
```
## Call:
## survdiff(formula = Surv(lev.sub$time, lev.sub$status) ~ lev.sub$sex)
##
##                  N Observed Expected (O-E)^2/E (O-E)^2/V
## lev.sub$sex=0 163       74     59.6      3.48      6.98
## lev.sub$sex=1 141       45     59.4      3.49      6.98
##
##  Chisq= 7  on 1 degrees of freedom, p= 0.008
```

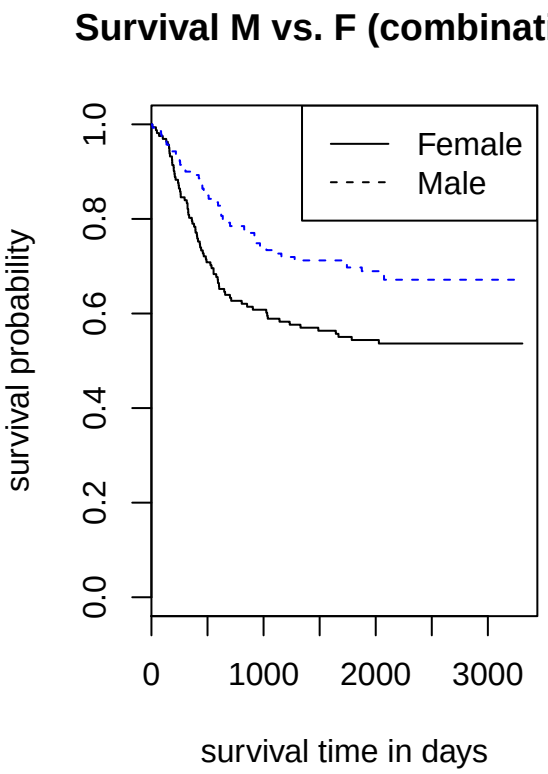
```
lfit <- survfit(Surv(lev.sub$time,lev.sub$status)~lev.sub$sex)

plot(lfit,
     lty=c("solid","dashed"),
     col=c("black","blue"),
     xlab="survival time in days",
     ylab="survival probability")

title("Survival M vs. F (combination)")

legend("topright",
c("Female","Male"),
lty=c("solid","dashed",
col="black","blue"))
```





Subsets are created of the placebo patients and treated patients using subset(). This means the sex can be used as the variable instead of the treatment status. Results indicate the survival in placebo is of similar length and probability with the female eventually being higher. Log ratio indicates that about 50% of females survive to 1274 in nontreated. 53% of treated females survive to 2027 days. For males, 36.5% live to 2695 and 59.7% live to 2074 days which indicates longer survival but lower probability following treatment.Survival is then starkly different following treatment for males vs. female. The males on treatment performed better than the females on treatment. Females on treatment showed increased survival probability and similar survival time to non-treated cases. The combination drug appears more affective in males than females. It appears to increase survival probability more than survival time. The differences in response may be due to physical differences such as hormonal/weight differences. It could also be due to optimal treatment level being found for males but suboptimal for females.

---

(3) The impact having more than 4 positive sentinel lymph nodes to the resection location

The predictor variable is replaced with node4.

```
par(mfrow=c(1,2))
#subset by placebo first
log.rank3 = survdiff(Surv(obs.sub$time,obs.sub$status)~obs.sub$node4)
log.rank3
```

```
## Call:
## survdiff(formula = Surv(obs.sub$time, obs.sub$status) ~ obs.sub$node4)
##
##                   N Observed Expected (O-E)^2/E (O-E)^2/V
```

```
## obs.sub$node4=0 228      114    140.6      5.03      24.7
## obs.sub$node4=1  87       63     36.4     19.42      24.7
##
##  Chisq= 24.7  on 1 degrees of freedom, p= 7e-07
```

```
nfit <- survfit(Surv(obs.sub$time,obs.sub$status)~obs.sub$node4)

plot(nfit,
     lty=c("solid","dashed"),
     col=c("black","blue"),
     xlab="survival time in days",
     ylab="survival probability")

title("Effect of 4+ LN (Placebo)")

legend("topright",
c("Not present","Present"),
lty=c("solid","dashed",
col="black","blue"))

###now for those with the drug

log.rank4 = survdiff(Surv(lev.sub$time,lev.sub$status)~lev.sub$node4)
log.rank4
```

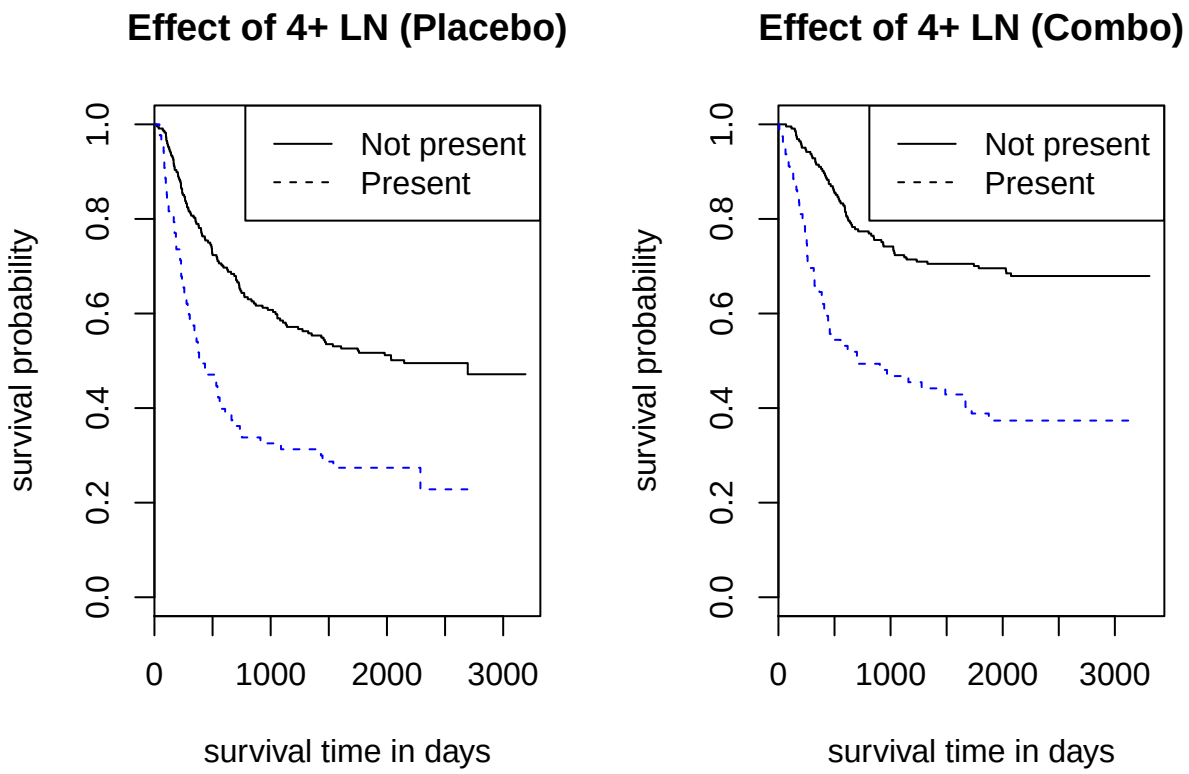
```
## Call:
## survdiff(formula = Surv(lev.sub$time, lev.sub$status) ~ lev.sub$node4)
##
##                    N Observed Expected (O-E)^2/E (O-E)^2/V
## lev.sub$node4=0 225       70     94.8      6.48        32
## lev.sub$node4=1  79       49     24.2     25.38        32
##
##  Chisq= 32  on 1 degrees of freedom, p= 2e-08
```

```
n1fit <- survfit(Surv(lev.sub$time,lev.sub$status)~lev.sub$node4)

plot(n1fit,
     lty=c("solid","dashed"),
     col=c("black","blue"),
     xlab="survival time in days",
     ylab="survival probability")

title("Effect of 4+ LN (Combo)")

legend("topright",
c("Not present","Present"),
lty=c("solid","dashed",
col="black","blue"))
```



The results indicate that probabiltiy of survival is much higher in the absence of 4+ sentinel lymph node. The log rank indicates that the chances of survival are greatly improved also. The probability of survival is only slightly increased in the presence of LM. However if the patient does not have 4 or more proximal LM and the are recieving combination treatment then their probability of survival is greatly improved. The greatest benefit for those who have 4 or more LM is that their survival time is improved but their prognosis is still bad. From this data, we can reject the null hypothesis that there is no difference in survival as it is obvious. This information is useful in decision making on how to proceed with cancers that are effected by lymph node proximity. Research into breast cancer survival following sentinel lymph node biopsy has indicated that there was no difference in survival between control groups and those who had the surgery. Using survival analysis for breast cancer or colorectal carcinoma following drug treatment or surgery in useful in finding the most effective treatment for increase survival probability and length.
