Cross-Domain Survival Analysis: From Prognosis to Operational Efficiency

Preethi Sree Allam

2023-12-11

Survival Analysis

Brain Cancer Data

```
library(ISLR2)
names(BrainCancer)
```

```
## [1] "sex"       "diagnosis" "loc"       "ki"        "gtv"       "stereo"
## [7] "status"    "time"
```

```
attach(BrainCancer)
table(sex)
```

```
## sex
## Female   Male
##     45     43
```

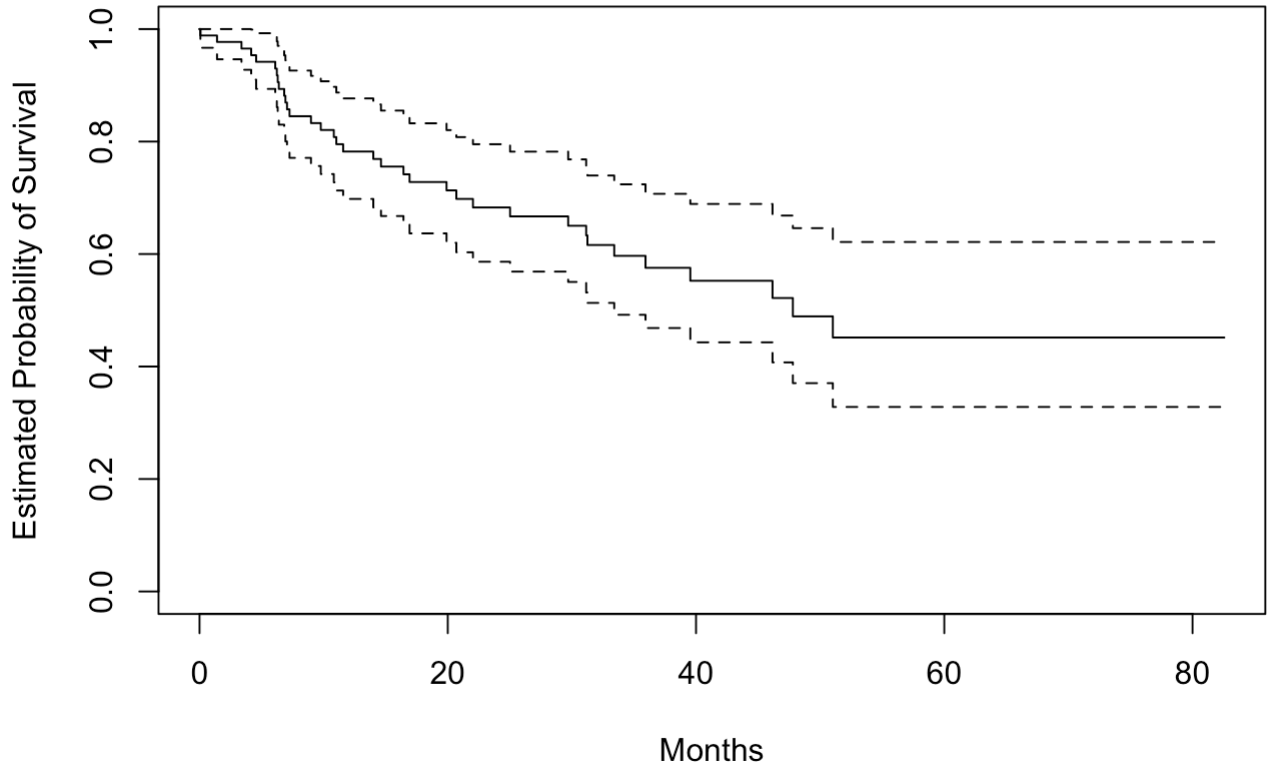
```
table(diagnosis)
```

```
## diagnosis
## Meningioma  LG glioma  HG glioma      Other
##         42          9         22         14
```

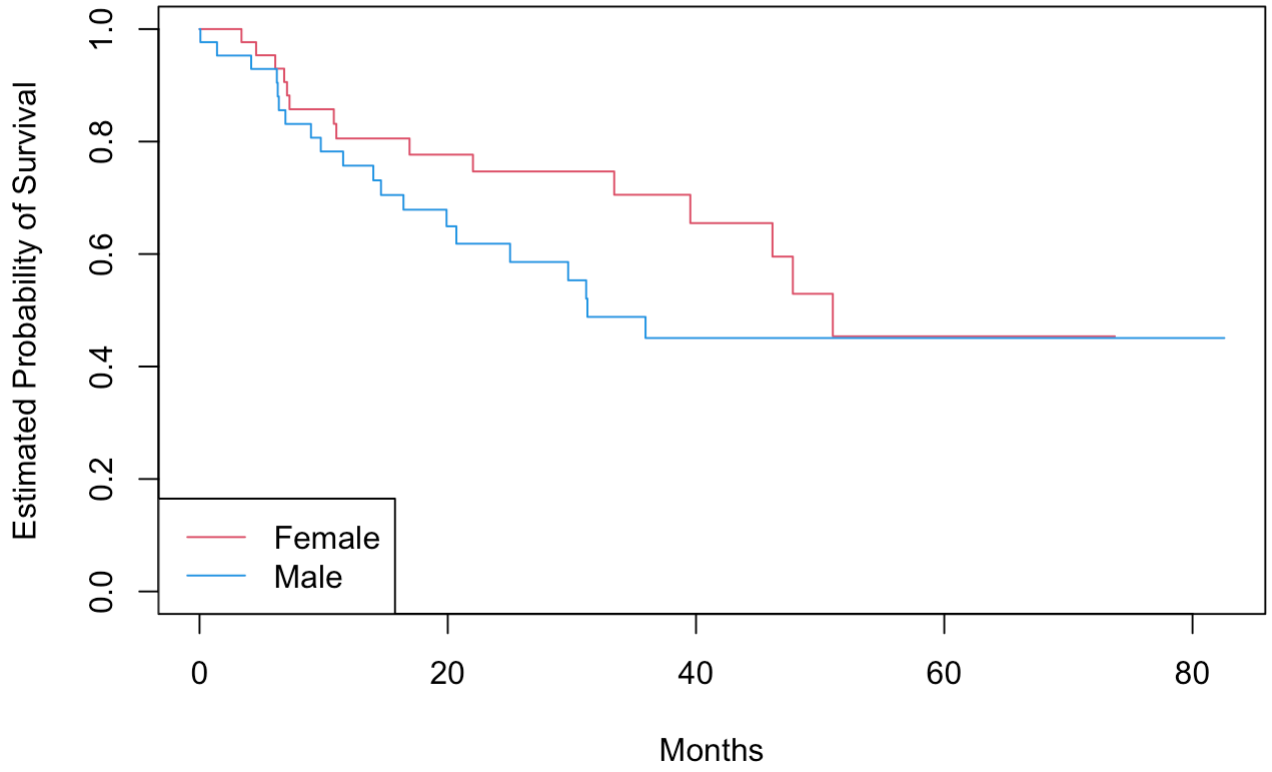
```
table(status)
```

```
## status
##  0  1
## 53 35
```

```
library(survival)
fit_surv <- survfit(Surv(time, status) ~ 1)
plot(fit_surv, xlab = "Months",
     ylab = "Estimated Probability of Survival")
```




```
fit_sex <- survfit (Surv(time,status) ~ BrainCancer$sex)
plot (fit_sex , xlab = "Months",
      ylab = "Estimated Probability of Survival", col = c(2,4))
legend("bottomleft", levels (sex), col = c(2,4), lty = 1)
```




```
logrank_test <- survdiff(Surv(time, status) ~ BrainCancer$sex)
logrank_test
```

```
## Call:
## survdiff(formula = Surv(time, status) ~ BrainCancer$sex)
##
##                          N Observed Expected (O-E)^2/E (O-E)^2/V
## BrainCancer$sex=Female 45       15     18.5     0.676      1.44
## BrainCancer$sex=Male   43       20     16.5     0.761      1.44
##
##  Chisq= 1.4  on 1 degrees of freedom, p= 0.2
```

```
fit.cox <- coxph(Surv(time, status) ~ BrainCancer$sex)
summary(fit.cox)
```

```
## Call:
## coxph(formula = Surv(time, status) ~ BrainCancer$sex)
##
##   n= 88, number of events= 35
##
##                       coef exp(coef) se(coef)     z Pr(>|z|)
## BrainCancer$sexMale 0.4077    1.5033   0.3420 1.192    0.233
##
##                     exp(coef) exp(-coef) lower .95 upper .95
## BrainCancer$sexMale     1.503     0.6652     0.769     2.939
##
## Concordance= 0.565  (se = 0.045 )
## Likelihood ratio test= 1.44  on 1 df,   p=0.2
## Wald test            = 1.42  on 1 df,   p=0.2
## Score (logrank) test = 1.44  on 1 df,   p=0.2
```

```
summary(fit.cox)$logtest[1]
```

```
##     test
## 1.438822
```

```
summary(fit.cox)$waldtest[1]
```

```
## test
## 1.42
```

```
summary(fit.cox)$sctest[1]
```

```
##     test
## 1.440495
```

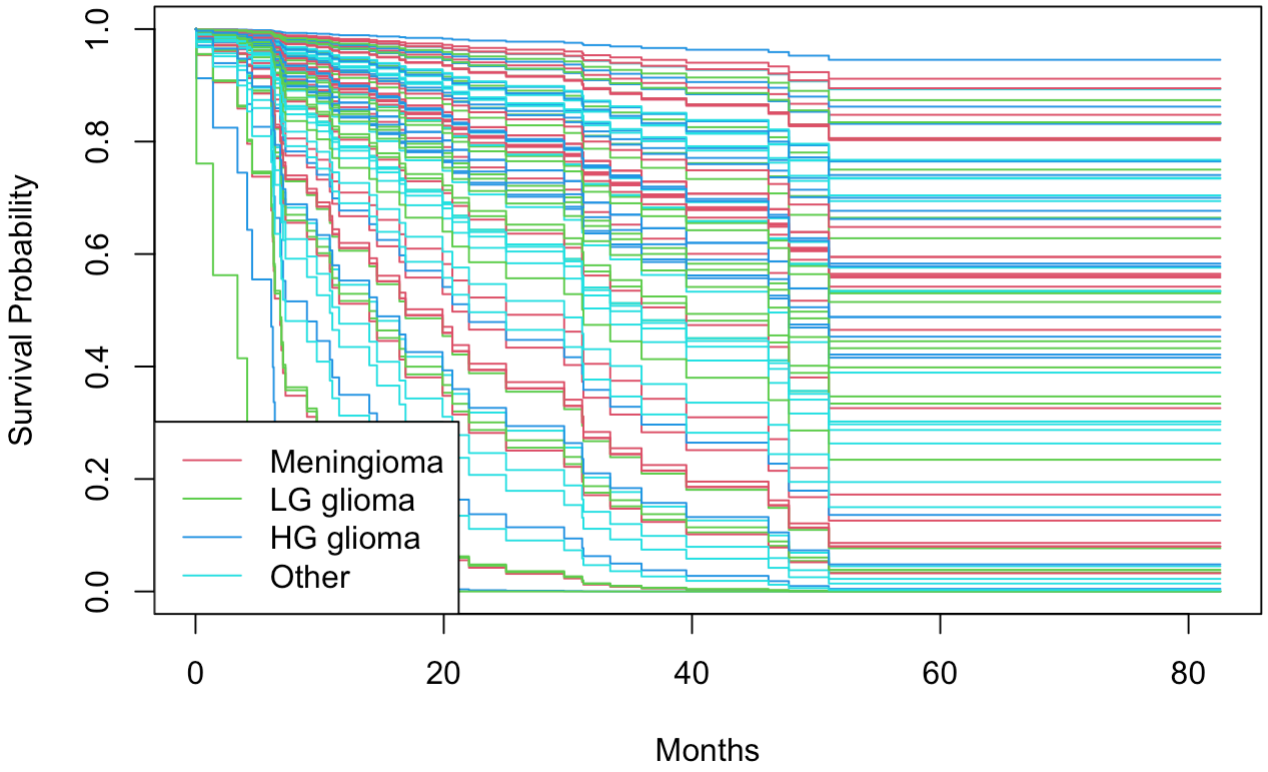
```
logrank_test$chisq
```

```
## [1] 1.440495
```

```
fit.all <- coxph(
  Surv(time, status) ~ BrainCancer$sex + diagnosis + loc + ki + gtv +
    stereo)
fit.all
```

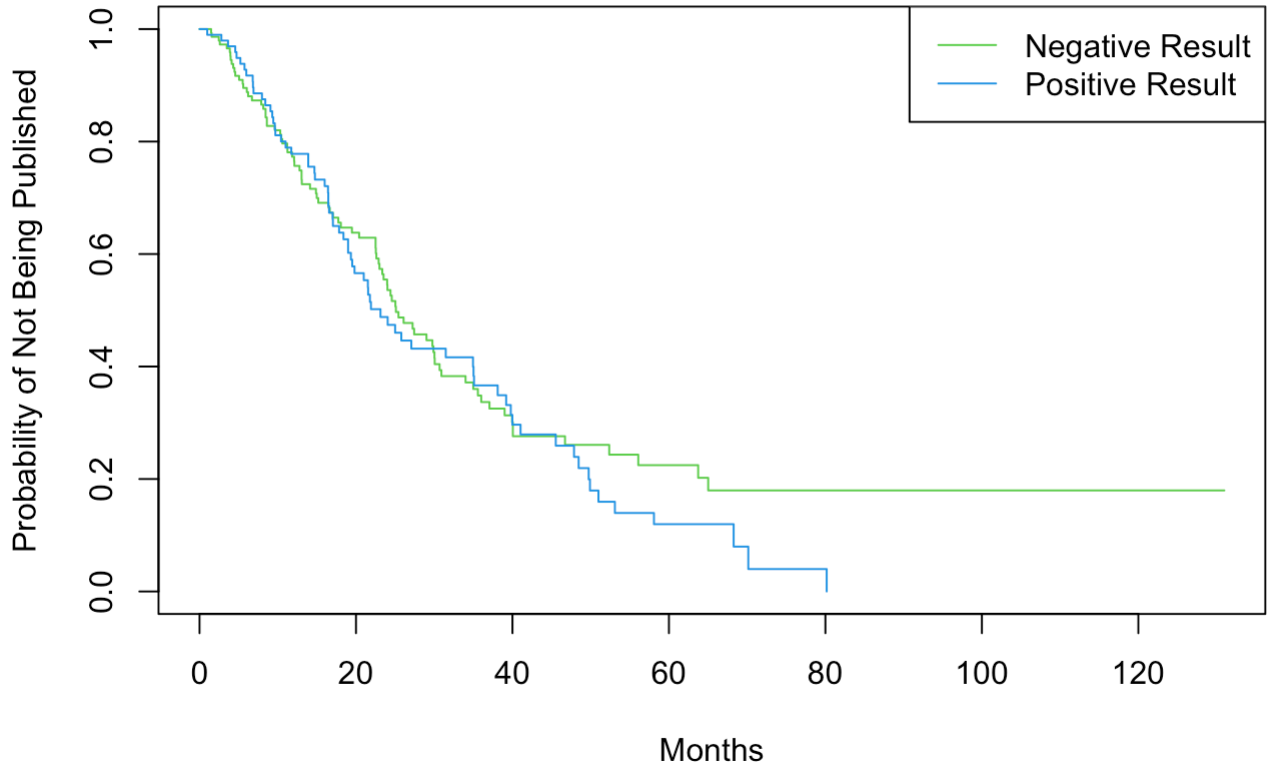
```
## Call:
## coxph(formula = Surv(time, status) ~ BrainCancer$sex + diagnosis +
##     loc + ki + gtv + stereo)
##
##                          coef exp(coef) se(coef)      z        p
## BrainCancer$sexMale   0.18375   1.20171  0.36036  0.510  0.61012
## diagnosisLG glioma    0.91502   2.49683  0.63816  1.434  0.15161
## diagnosisHG glioma    2.15457   8.62414  0.45052  4.782 1.73e-06
## diagnosisOther        0.88570   2.42467  0.65787  1.346  0.17821
## locSupratentorial     0.44119   1.55456  0.70367  0.627  0.53066
## ki                   -0.05496   0.94653  0.01831 -3.001  0.00269
## gtv                   0.03429   1.03489  0.02233  1.536  0.12466
## stereoSRT             0.17778   1.19456  0.60158  0.296  0.76760
##
## Likelihood ratio test=41.37  on 8 df, p=1.776e-06
## n= 87, number of events= 35
##    (1 observation deleted due to missingness)
```

```
modaldata <- data.frame(
  diagnosis = levels(diagnosis),
  sex = rep("Female", 4),
  loc = rep("Supratentorial", 4),
  ki = rep(mean(ki), 4),
  gtv = rep(mean(gtv), 4),
  stereo = rep("SRT", 4)
)
survplots <- survfit(fit.all, newdata = BrainCancer$modaldata)
plot(survplots, xlab = "Months",
     ylab = "Survival Probability", col = 2:5)
legend("bottomleft", levels(diagnosis), col = 2:5, lty = 1)
```



Publication Data

```
fit.posres <- survfit(
  Surv(time, status) ~ posres, data = Publication
)
plot(fit.posres, xlab = "Months",
     ylab = "Probability of Not Being Published", col = 3:4)
legend("topright", c("Negative Result", "Positive Result"),
       col = 3:4, lty = 1)
```




```
fit.pub <- coxph(Surv(time, status) ~ posres,
                 data = Publication)
fit.pub
```

```
## Call:
## coxph(formula = Surv(time, status) ~ posres, data = Publication)
##
##          coef exp(coef) se(coef)     z    p
## posres 0.1481    1.1596   0.1616 0.916 0.36
##
## Likelihood ratio test=0.83  on 1 df, p=0.3611
## n= 244, number of events= 156
```

```
logrank.test <- survdiff(Surv(time, status) ~ posres,
                         data = Publication)
logrank.test
```

```
## Call:
## survdiff(formula = Surv(time, status) ~ posres, data = Publication)
##
##            N Observed Expected (O-E)^2/E (O-E)^2/V
## posres=0 146       87     92.6     0.341     0.844
## posres=1  98       69     63.4     0.498     0.844
##
##  Chisq= 0.8  on 1 degrees of freedom, p= 0.4
```

```
fit.pub2 <- coxph(Surv(time, status) ~ . - mech,
                  data = Publication)
fit.pub2
```

```
## Call:
## coxph(formula = Surv(time, status) ~ . - mech, data = Publication)
##
##                coef  exp(coef)   se(coef)      z       p
## posres    5.708e-01  1.770e+00  1.760e-01  3.244 0.00118
## multi    -4.086e-02  9.600e-01  2.512e-01 -0.163 0.87079
## clinend   5.462e-01  1.727e+00  2.620e-01  2.085 0.03710
## sampsize  4.678e-06  1.000e+00  1.472e-05  0.318 0.75070
## budget    4.385e-03  1.004e+00  2.465e-03  1.779 0.07518
## impact    5.832e-02  1.060e+00  6.676e-03  8.735 < 2e-16
##
## Likelihood ratio test=149.2  on 6 df, p=< 2.2e-16
## n= 244, number of events= 156
```

Call center data

```
set.seed(4)
N <- 2000
Operators <- sample(5:15, N, replace = T)
Center <- sample(c("A", "B", "C"), N, replace = T)
Time <- sample(c("Morn.", "After.", "Even."), N, replace = T)
X <- model.matrix( ~ Operators + Center + Time)[, -1]
X[1:5, ]
```

```
##   Operators CenterB CenterC TimeEven. TimeMorn.
## 1        12       1       0         0         1
## 2        15       0       0         0         0
## 3         7       0       1         1         0
## 4         7       0       0         0         0
## 5        11       0       1         0         1
```

```
true.beta <- c(0.04, -0.3, 0, 0.2, -0.2)
h.fn <- function(x) return(0.00001 * x)
library(coxed)
```

```
## Loading required package: rms
```

```
## Loading required package: Hmisc
```

```
##
## Attaching package: 'Hmisc'
```

```
## The following objects are masked from 'package:base':
##
##     format.pval, units
```

```
## Loading required package: mgcv
```

```
## Loading required package: nlme
```

```
## This is mgcv 1.8-42. For overview type 'help("mgcv-package")'.
```

```
queuing <- sim.survdata(N = N, T = 1000, X = X,
                        beta = true.beta, hazard.fun = h.fn)
```

```
## Warning in FUN(X[[i]], ...): 9 additional observations right-censored because the user-supplied hazard functio
n
##                                  is nonzero at the latest timepoint. To avoid these extra censored observatio
ns, increase T
```

```
names(queuing)
```

```
## [1] "data"            "xdata"           "baseline"        "xb"
## [5] "exp.xb"          "betas"           "ind.survive"     "marg.effect"
## [9] "marg.effect.data"
```

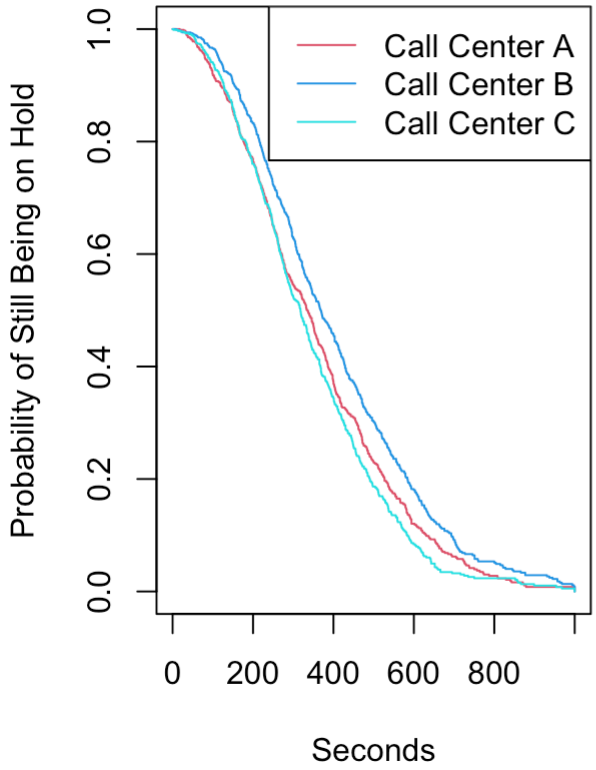
```
head(queuing$data)
```

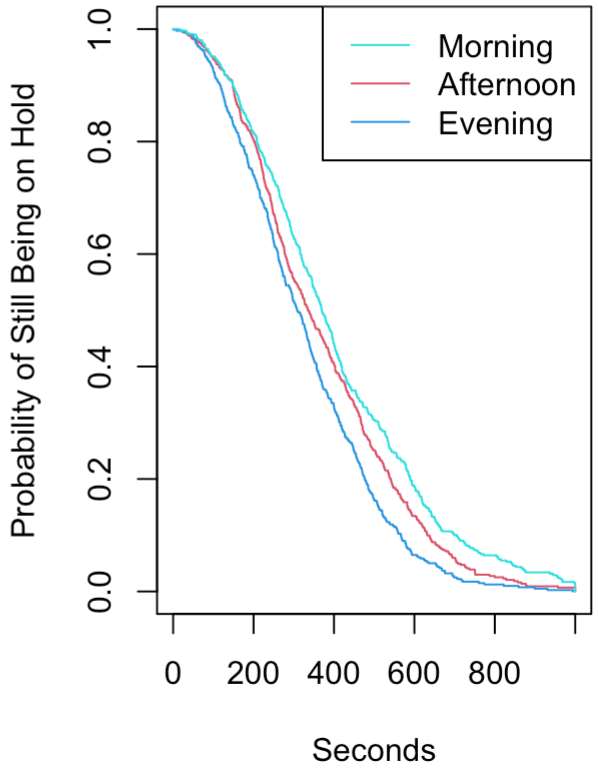
```
##   Operators CenterB CenterC TimeEven. TimeMorn.   y failed
## 1        12       1       0         0         1 344   TRUE
## 2        15       0       0         0         0 241   TRUE
## 3         7       0       1         1         0 187   TRUE
## 4         7       0       0         0         0 279   TRUE
## 5        11       0       1         0         1 954   TRUE
## 6         7       1       0         0         1 455   TRUE
```

```
mean(queuing$data$failed)
```

```
## [1] 0.89
```

```
par(mfrow = c(1, 2))
fit.Center <- survfit(Surv(y, failed) ~ Center,
                      data = queuing$data)
plot(fit.Center, xlab = "Seconds",
     ylab = "Probability of Still Being on Hold",
     col = c(2, 4, 5))
legend("topright",
       c("Call Center A", "Call Center B", "Call Center C"),
       col = c(2, 4, 5), lty = 1)
fit.Time <- survfit(Surv(y, failed) ~ Time,
                    data = queuing$data)
plot(fit.Time, xlab = "Seconds",
     ylab = "Probability of Still Being on Hold",
     col = c(2, 4, 5))
legend("topright", c("Morning", "Afternoon", "Evening"),
       col = c(5, 2, 4), lty = 1)
```






```
survdiff(Surv(y, failed) ~ Center, data = queuing$data)
```

```
## Call:
## survdiff(formula = Surv(y, failed) ~ Center, data = queuing$data)
##
##            N Observed Expected (O-E)^2/E (O-E)^2/V
## Center=A 683      603      579     0.971      1.45
## Center=B 667      600      701    14.641     24.64
## Center=C 650      577      499    12.062     17.05
##
##  Chisq= 28.3  on 2 degrees of freedom, p= 7e-07
```

```
survdiff(Surv(y, failed) ~ Time, data = queuing$data)
```

```
## Call:
## survdiff(formula = Surv(y, failed) ~ Time, data = queuing$data)
##
##               N Observed Expected (O-E)^2/E (O-E)^2/V
## Time=After. 688      616      619    0.0135     0.021
## Time=Even.  653      582      468   27.6353    38.353
## Time=Morn.  659      582      693   17.7381    29.893
##
##  Chisq= 46.8  on 2 degrees of freedom, p= 7e-11
```

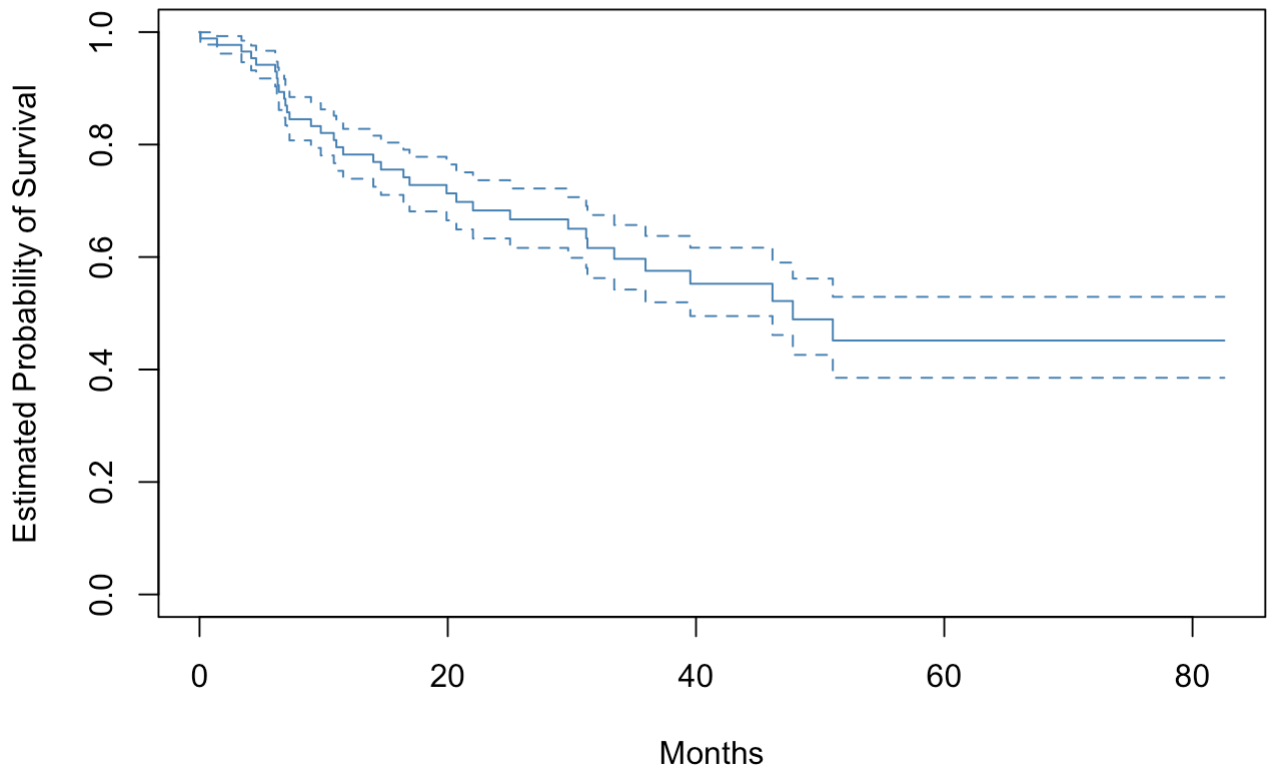
```
fit.queuing <- coxph(Surv(y, failed) ~ .,
                     data = queuing$data)
fit.queuing
```

```
## Call:
## coxph(formula = Surv(y, failed) ~ ., data = queuing$data)
##
##               coef exp(coef) se(coef)      z        p
## Operators  0.04174   1.04263  0.00759  5.500  3.8e-08
## CenterB   -0.21879   0.80349  0.05793 -3.777 0.000159
## CenterC    0.07930   1.08253  0.05850  1.356 0.175256
## TimeEven.  0.20904   1.23249  0.05820  3.592 0.000328
## TimeMorn. -0.17352   0.84070  0.05811 -2.986 0.002828
##
## Likelihood ratio test=102.8  on 5 df, p=< 2.2e-16
## n= 2000, number of events= 1780
```

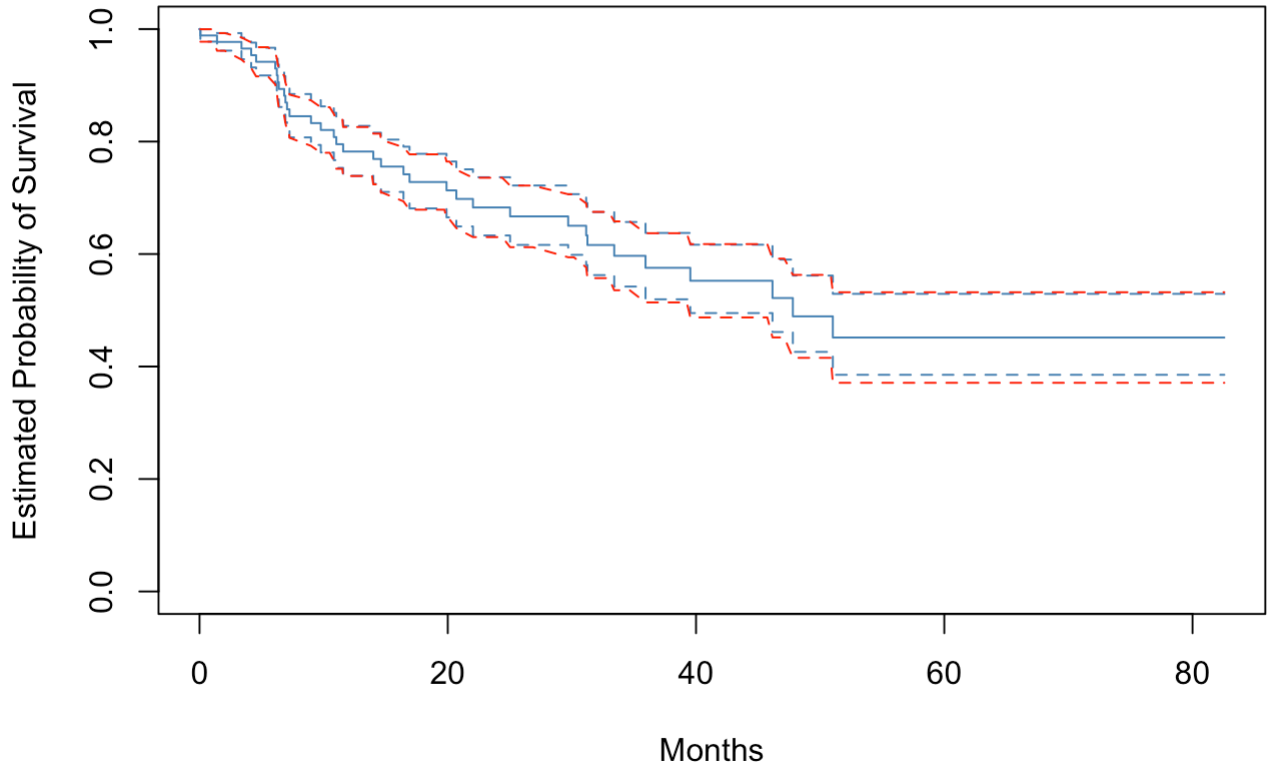
```
library(ISLR2)
library(tidyverse)
```

```
## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──
## ✔ dplyr     1.1.2     ✔ readr     2.1.4
## ✔ forcats   1.0.0     ✔ stringr   1.5.0
## ✔ ggplot2   3.4.2     ✔ tibble    3.2.1
## ✔ lubridate 1.9.2     ✔ tidyr     1.3.0
## ✔ purrr     1.0.1
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## ✖ dplyr::collapse()  masks nlme::collapse()
## ✖ dplyr::filter()    masks stats::filter()
## ✖ dplyr::lag()       masks stats::lag()
## ✖ dplyr::src()       masks Hmisc::src()
## ✖ dplyr::summarize() masks Hmisc::summarize()
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
x <- Surv(BrainCancer$time, BrainCancer$status)
plot(survfit(x ~ 1),
     xlab = "Months",
     ylab = "Estimated Probability of Survival",
     col = "steelblue",
     conf.int = 0.67
)
```



```
plot(survfit(x ~ 1),
     xlab = "Months",
     ylab = "Estimated Probability of Survival",
     col = "steelblue",
     conf.int = 0.67
)
fit <- survfit(x ~ 1)
dat <- tibble(time = c(0, fit$time))
for (i in 1:200) {
  y <- survfit(sample(x, 88, replace = TRUE) ~ 1)
  y <- tibble(time = c(0, y$time), "s{i}" := c(1, y$surv))
  dat <- left_join(dat, y, by = "time")
}
res <- fill(dat, starts_with("s")) |>
  rowwise() |>
  transmute(sd = sd(c_across(starts_with("s"))))
se <- res$sd[2:nrow(res)]
lines(fit$time, fit$surv - se, lty = 2, col = "red")
lines(fit$time, fit$surv + se, lty = 2, col = "red")
```

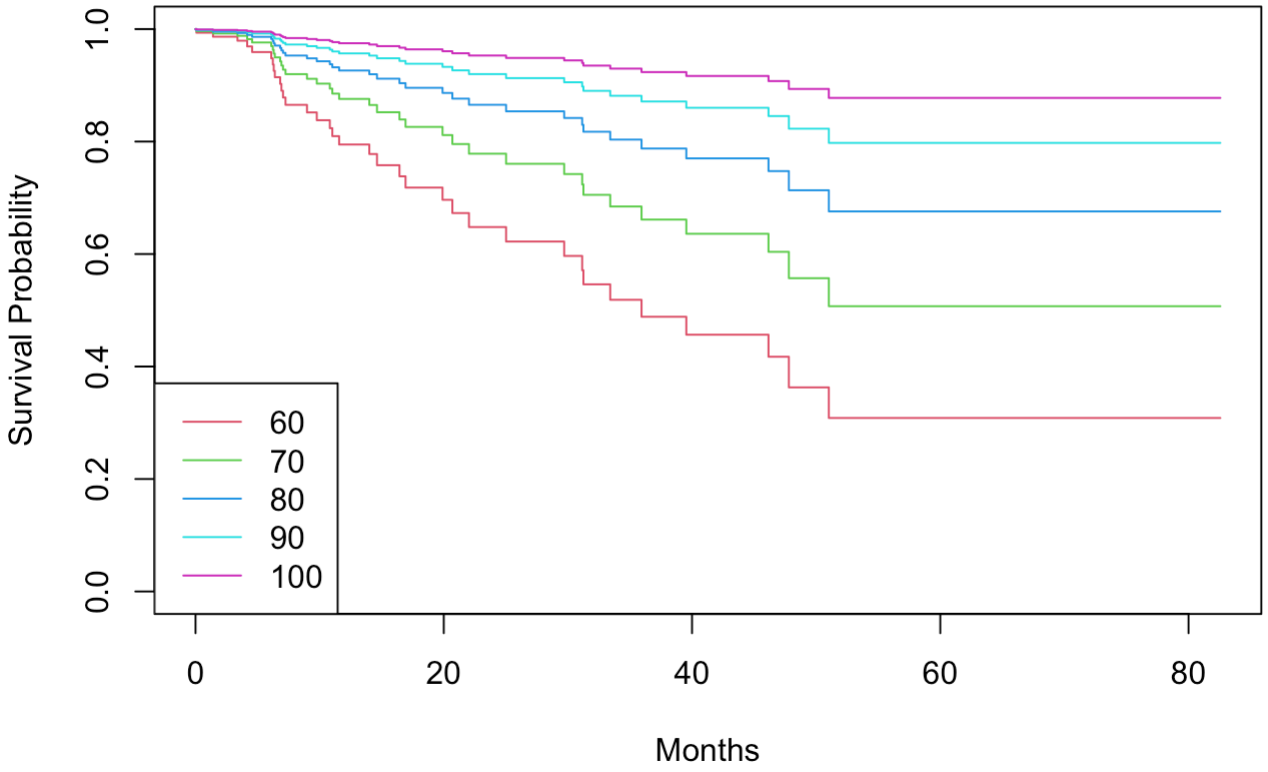



```
fit <- coxph(Surv(time, status) ~ sex + diagnosis + loc + ki + gtv + stereo, data = BrainCancer)
fit
```

```
## Call:
## coxph(formula = Surv(time, status) ~ sex + diagnosis + loc +
##     ki + gtv + stereo, data = BrainCancer)
##
##                        coef exp(coef) se(coef)      z        p
## sexMale             0.18375   1.20171  0.36036  0.510  0.61012
## diagnosisLG glioma  0.91502   2.49683  0.63816  1.434  0.15161
## diagnosisHG glioma  2.15457   8.62414  0.45052  4.782 1.73e-06
## diagnosisOther      0.88570   2.42467  0.65787  1.346  0.17821
## locSupratentorial   0.44119   1.55456  0.70367  0.627  0.53066
## ki                 -0.05496   0.94653  0.01831 -3.001  0.00269
## gtv                 0.03429   1.03489  0.02233  1.536  0.12466
## stereoSRT           0.17778   1.19456  0.60158  0.296  0.76760
##
## Likelihood ratio test=41.37  on 8 df, p=1.776e-06
## n= 87, number of events= 35
##    (1 observation deleted due to missingness)
```

`diagnosisHG` and `ki` are highly significant.

```
library(ggfortify)
modaldata <- data.frame(
  sex = rep("Female", 5),
  diagnosis = rep("Meningioma", 5),
  loc = rep("Supratentorial", 5),
  ki = c(60, 70, 80, 90, 100),
  gtv = rep(mean(BrainCancer$gtv), 5),
  stereo = rep("SRT", 5)
)
survplots <- survfit(fit, newdata = modaldata)
plot(survplots, xlab = "Months", ylab = "Survival Probability", col = 2:6)
legend("bottomleft", c("60", "70", "80", "90", "100"), col = 2:6, lty = 1)
```


Survival Probability
0.0
0.2
0.4
0.6
0.8
1.0
60
70
80
90
100
0
20
40
60
80
Months