Lecture 01. Introduction to Survival Analysis

Alex Shpenev

1/21/2019

Violet Jessop

Figure 1: Violet Jessop

Titanic

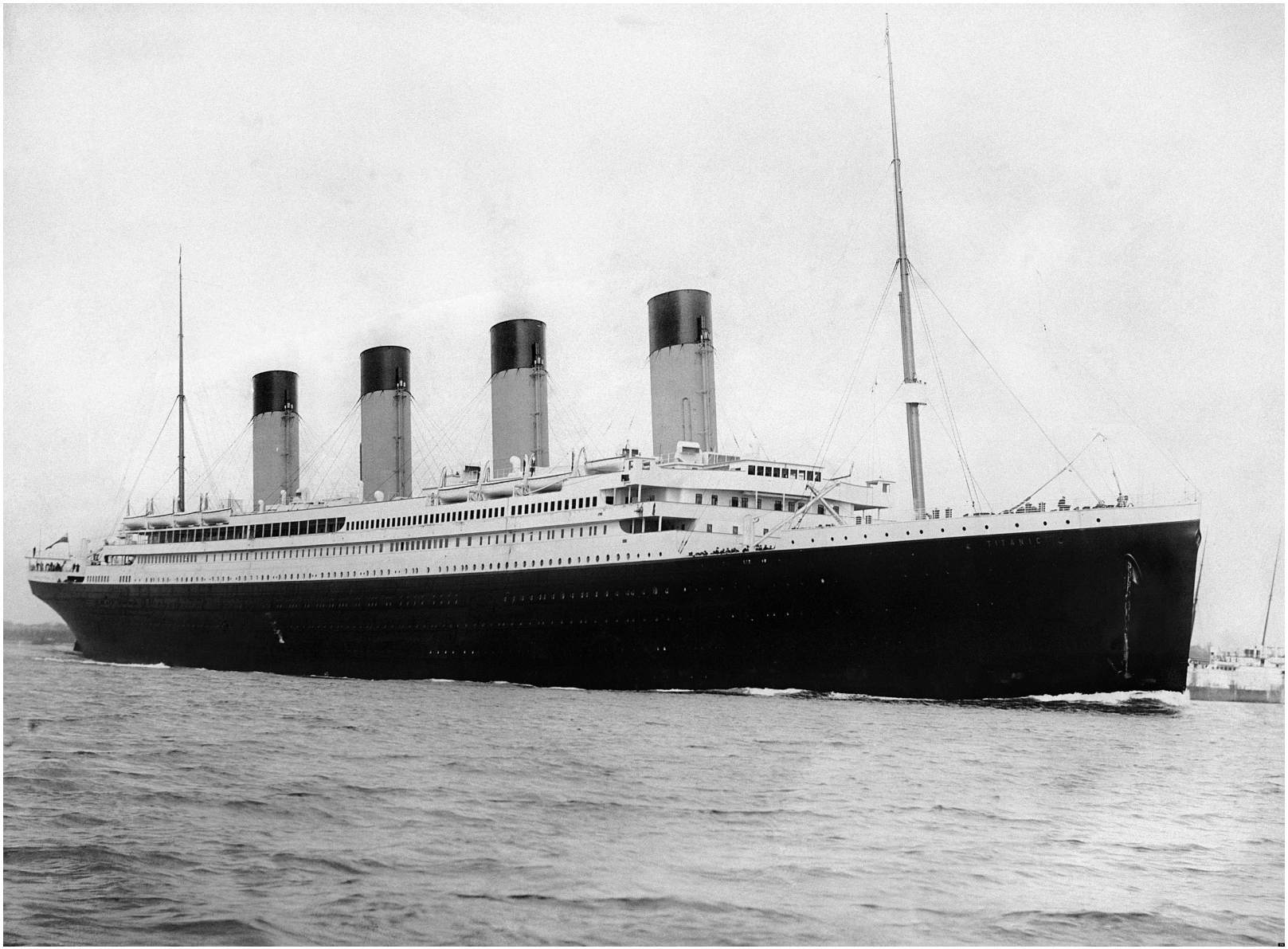


Figure 2: Titanic

What predicts survival on the Titanic

```
library(readr)
library(dplyr)
titanic <- read_csv("titanic.csv")
```

```
glimpse(titanic)
```

```
attach(titanic)
mean(age, na.rm = T)
```

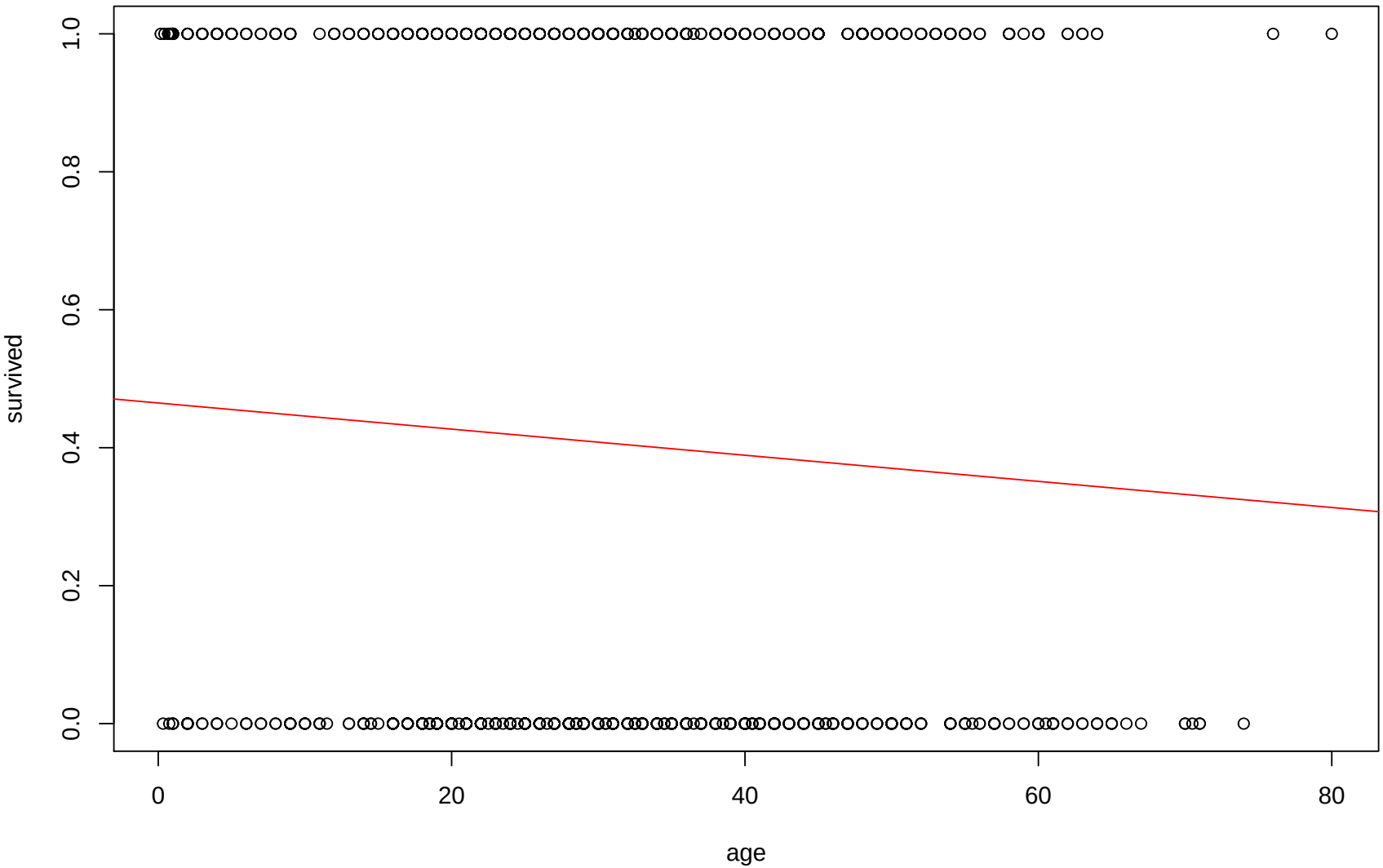
```
## [1] 29.88114
```

```
table(survived, sex)
```

```
##          sex
## survived female male
##        0    127  682
##        1    339  161
```

Linear Regression

```
plot(survived ~ age)
abline(lm(survived~age), col = "red")
```



Linear Regression 2

```r
summary(lm(survived~age))
```

```
##
## Call:
## lm(formula = survived ~ age)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.4642 -0.4156 -0.3796  0.5806  0.6867
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.464813   0.034973  13.291   <2e-16 ***
## age         -0.001894   0.001054  -1.796   0.0727 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
##
## Residual standard error: 0.4912 on 1044 degrees of freed
```

Logistic Regression

```
summary(glm(survived ~ age, family = "binomial"))
```

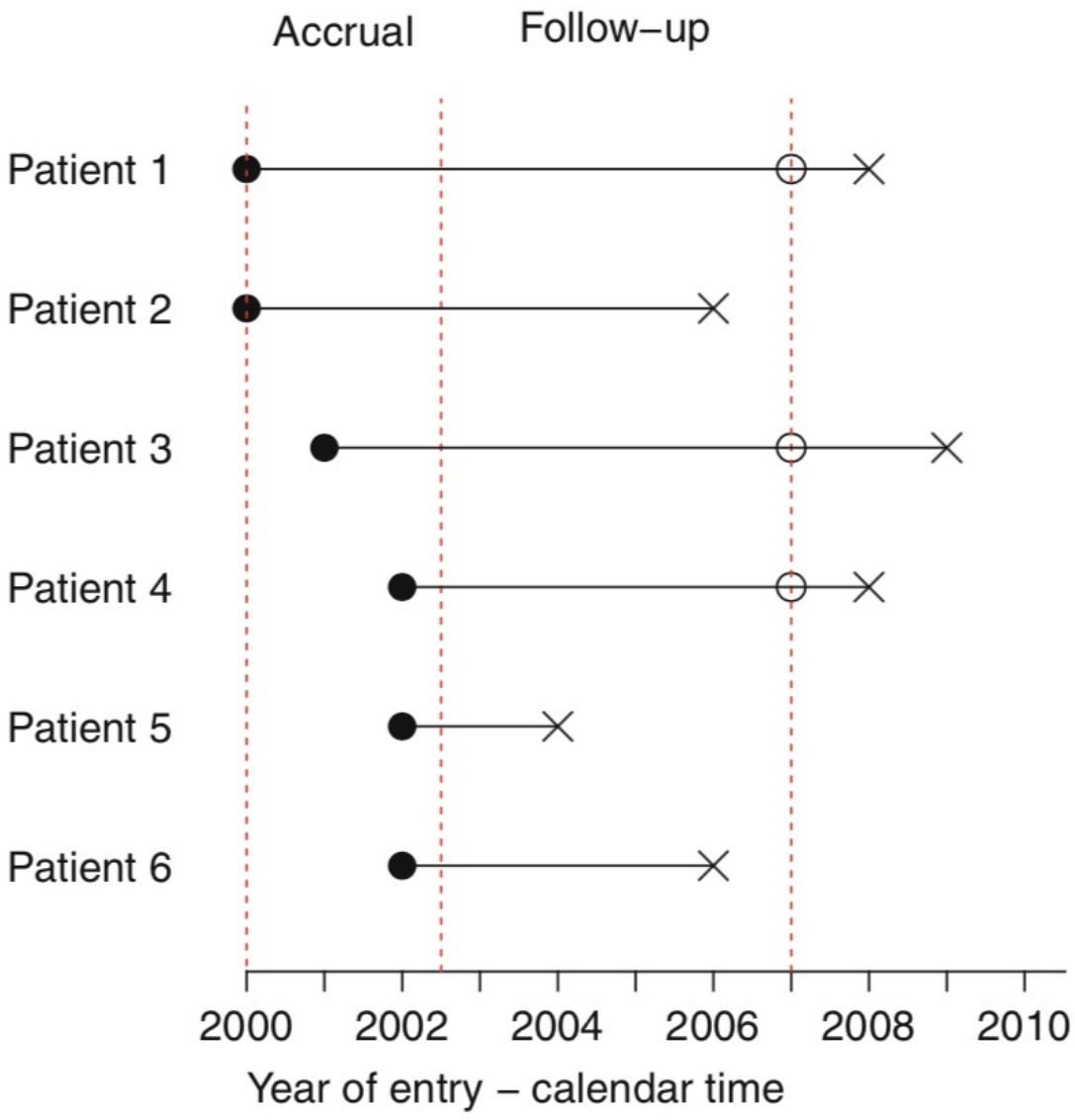
```
##
## Call:
## glm(formula = survived ~ age, family = "binomial")
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.1189  -1.0361  -0.9768   1.3187   1.5162
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.136534   0.144715  -0.943   0.3454
## age         -0.007899   0.004407  -1.792   0.0731 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
##
## (Dispersion parameter for binomial family taken to be 1)
```

Problems with using this approach

- Censoring
  - Right censoring
  - Interval censoring
  - Left censoring
- Truncation
  - Left truncation
  - Right truncation
  - Interval trunncation

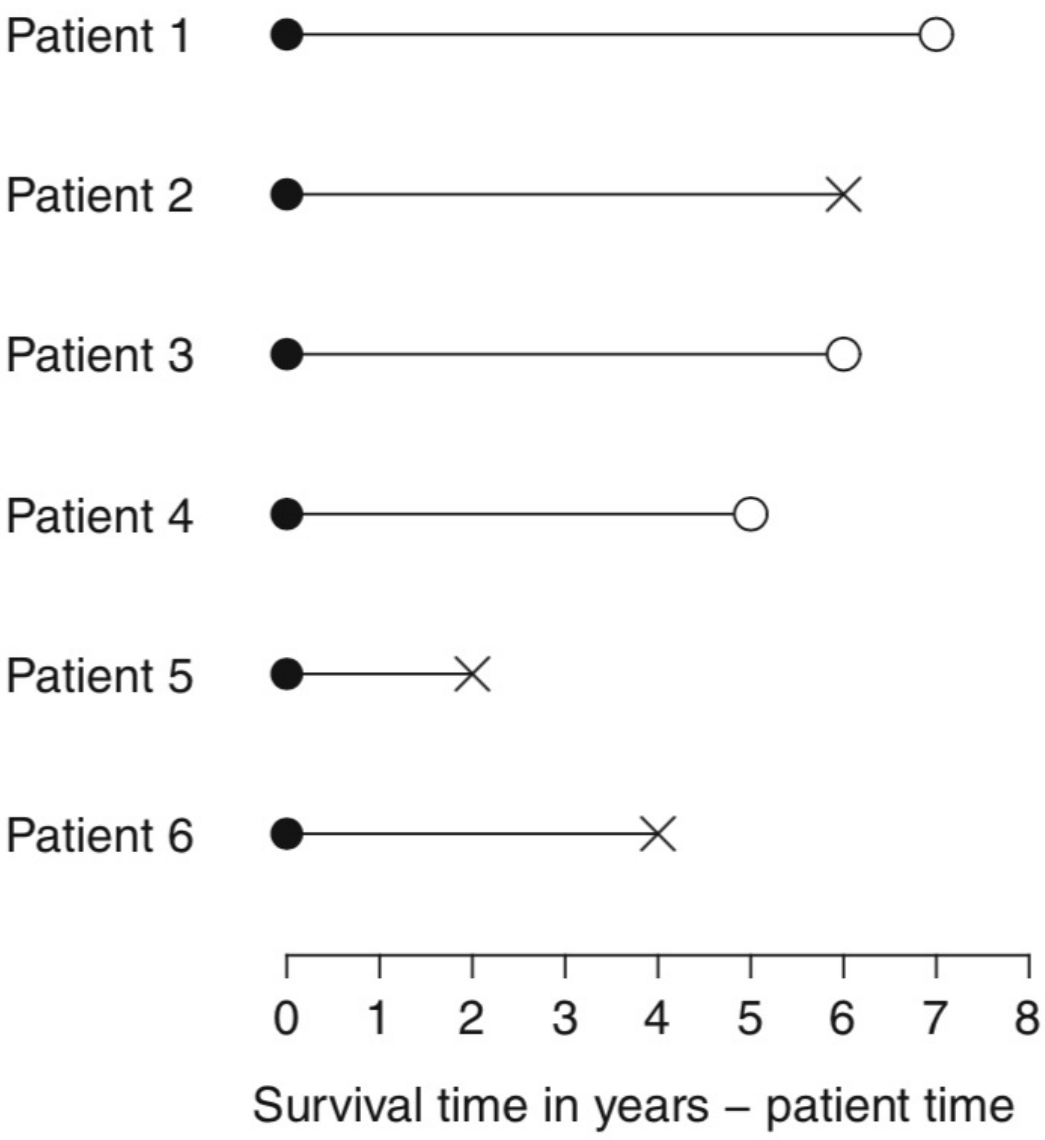
Hence - Survival analysis

Conceptual framework



Source: Moore,

Survival data



Source: Moore,

Some demographic functions

- Survival Function

$$S(t) = P(T > t),\ 0 < t < \infty$$

- Hazard function

$$h(t) = \lim_{\delta \to 0} \frac{P(t < T < t + \delta | T > t)}{\delta}$$

Some examples

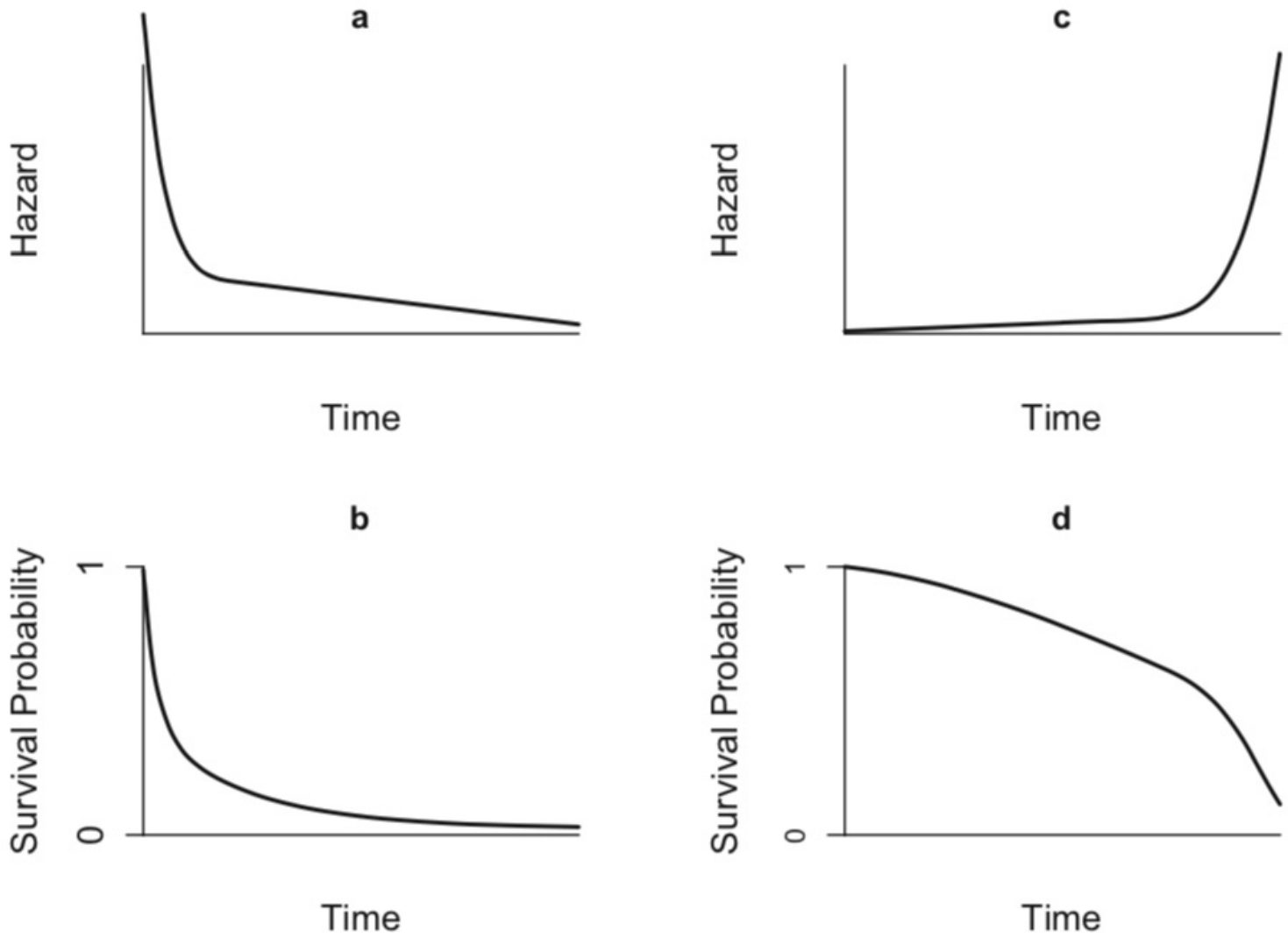


Figure 3: High an low initial hazards

A more realistic example

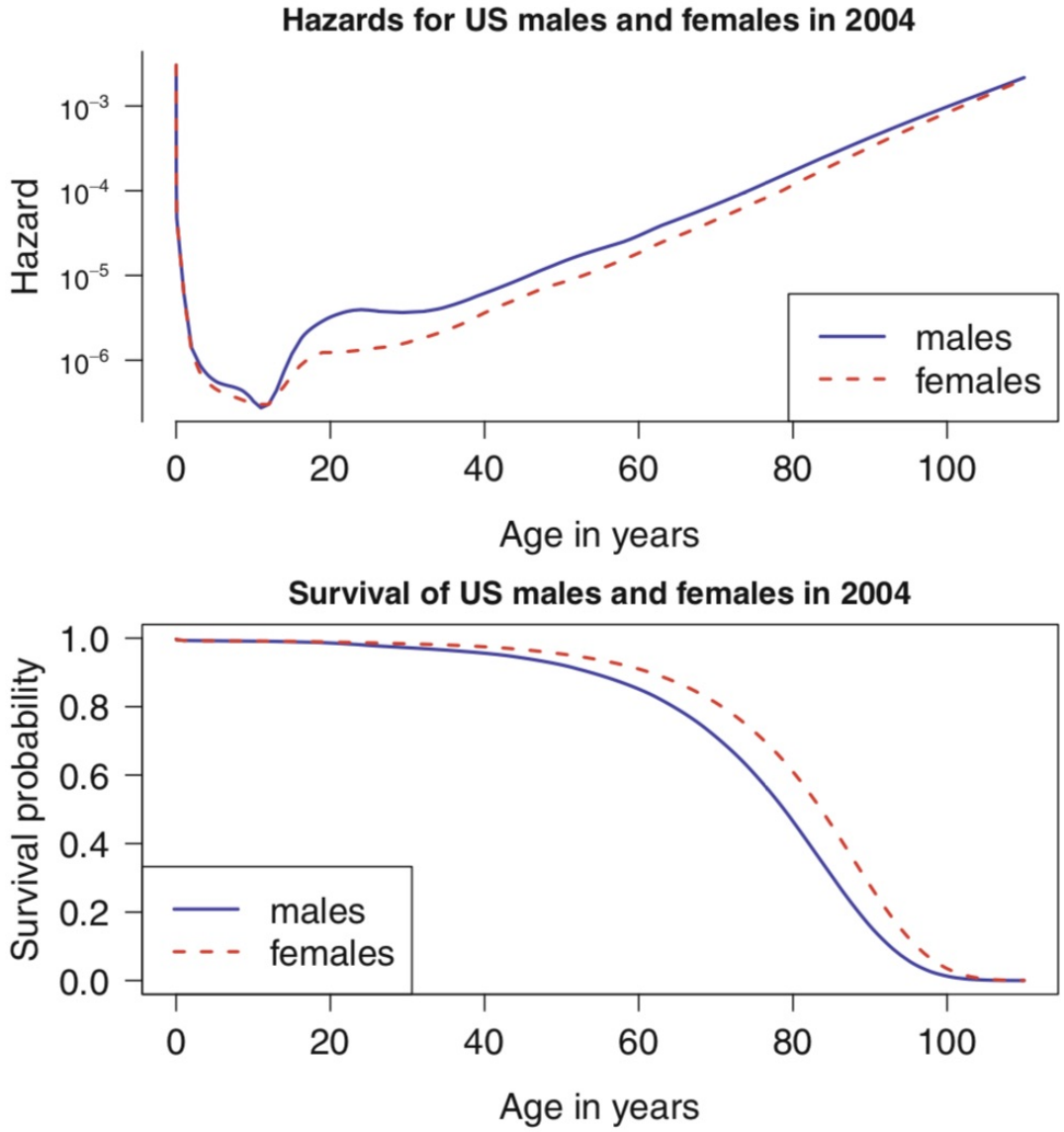


Figure 4: High an low initial hazards

Non-parametric estimation

- Kaplan Meier estimator

$$\hat{S}(t) = \prod_{t_i \leq t} (1 - \hat{q}_i) = \prod_{t_i \leq t} (1 - \frac{d_i}{n_i})$$

Let's do a quick exercise

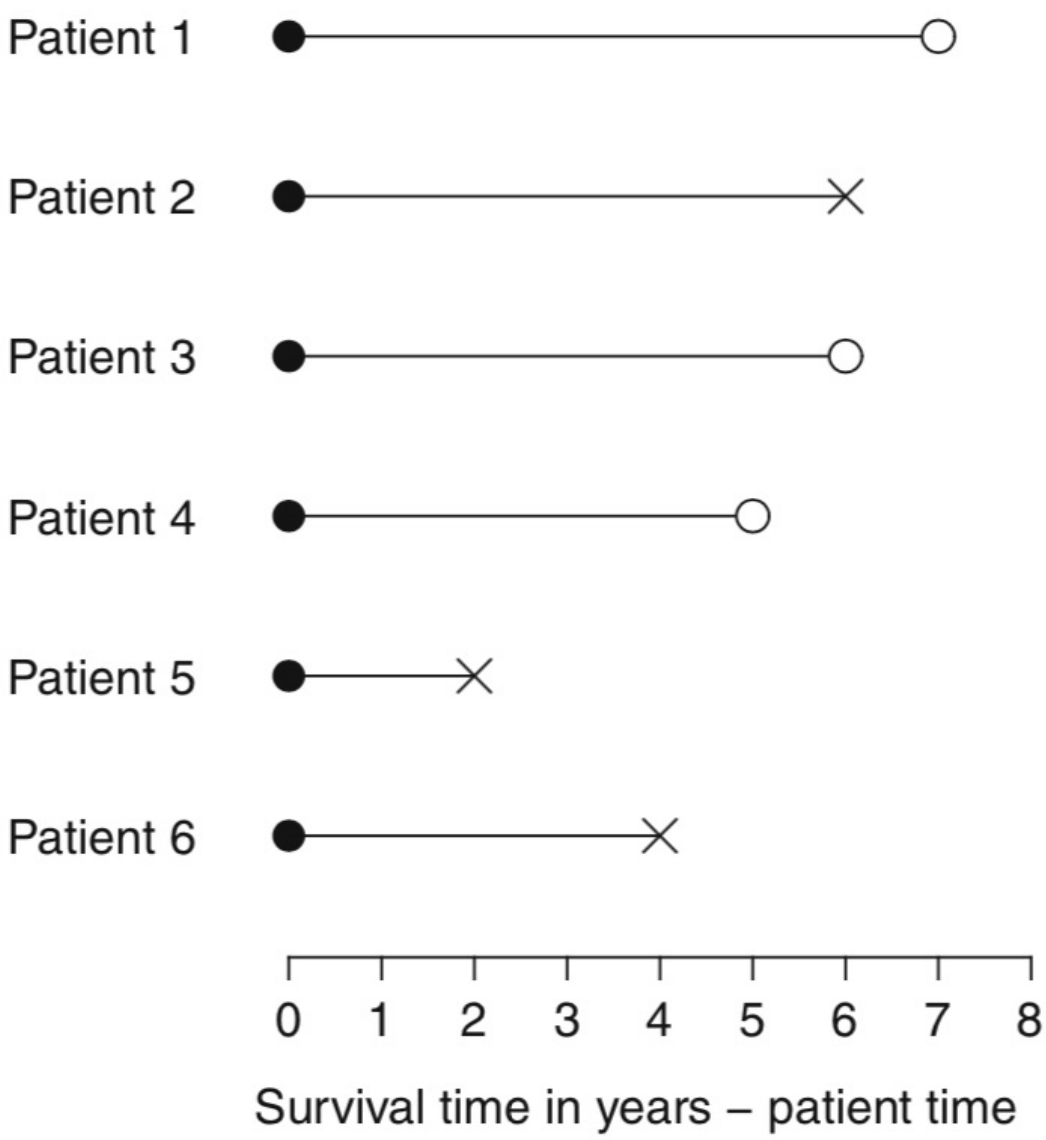


Figure 5: Patient Time

Let's do a quick exercise

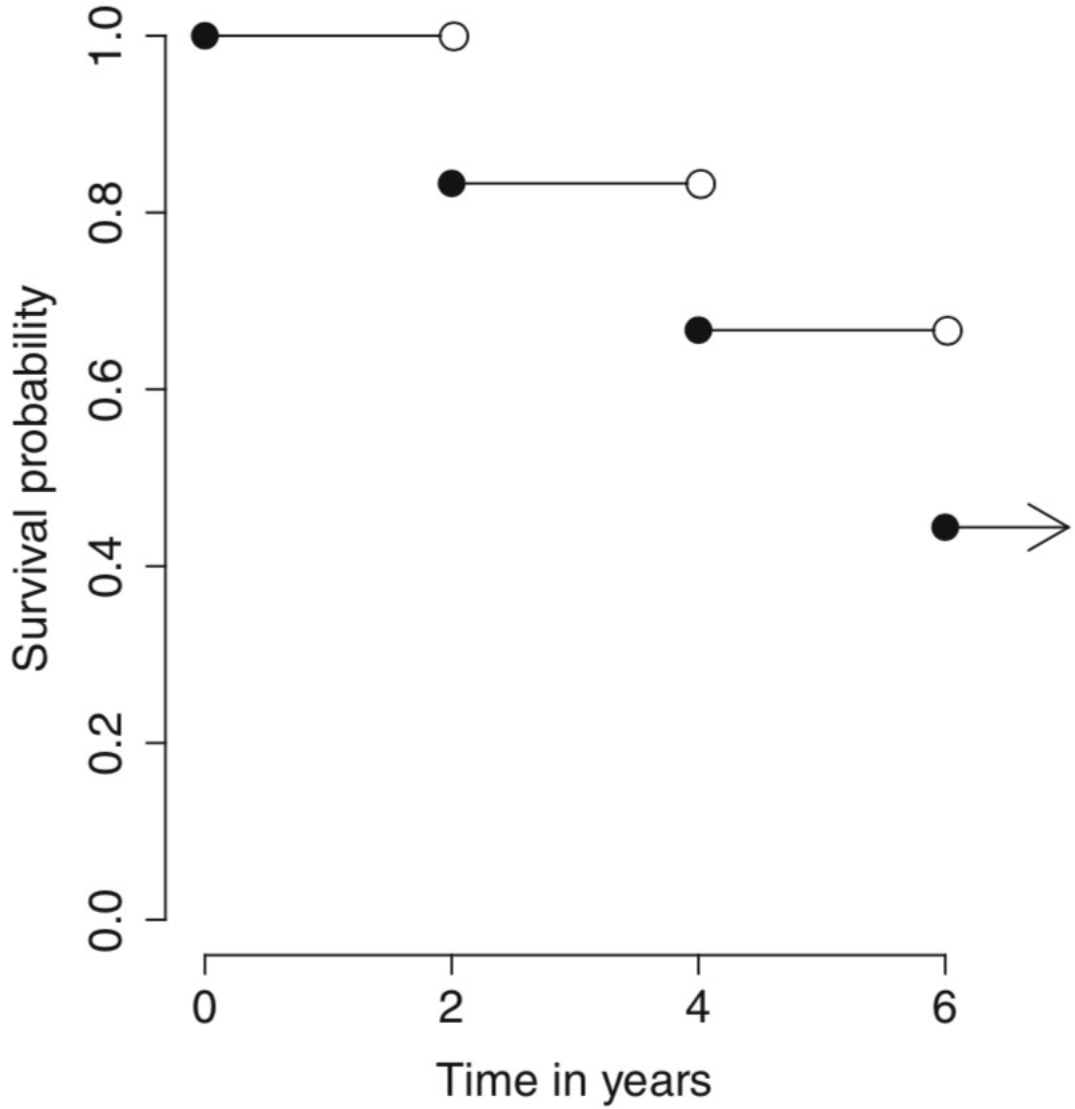


Figure 6: KM estimate

Now let's do it in R

```
library(survival)
tt <- c(7,6,6,5,2,4)
cens <- c(0,1,0,0,1,1)
Surv(tt, cens)
```

```
## [1] 7+ 6  6+ 5+ 2  4
```

Survival model

```
result.km <- survfit(Surv(tt, cens) ~ 1, conf.type="log-log
result.km
```

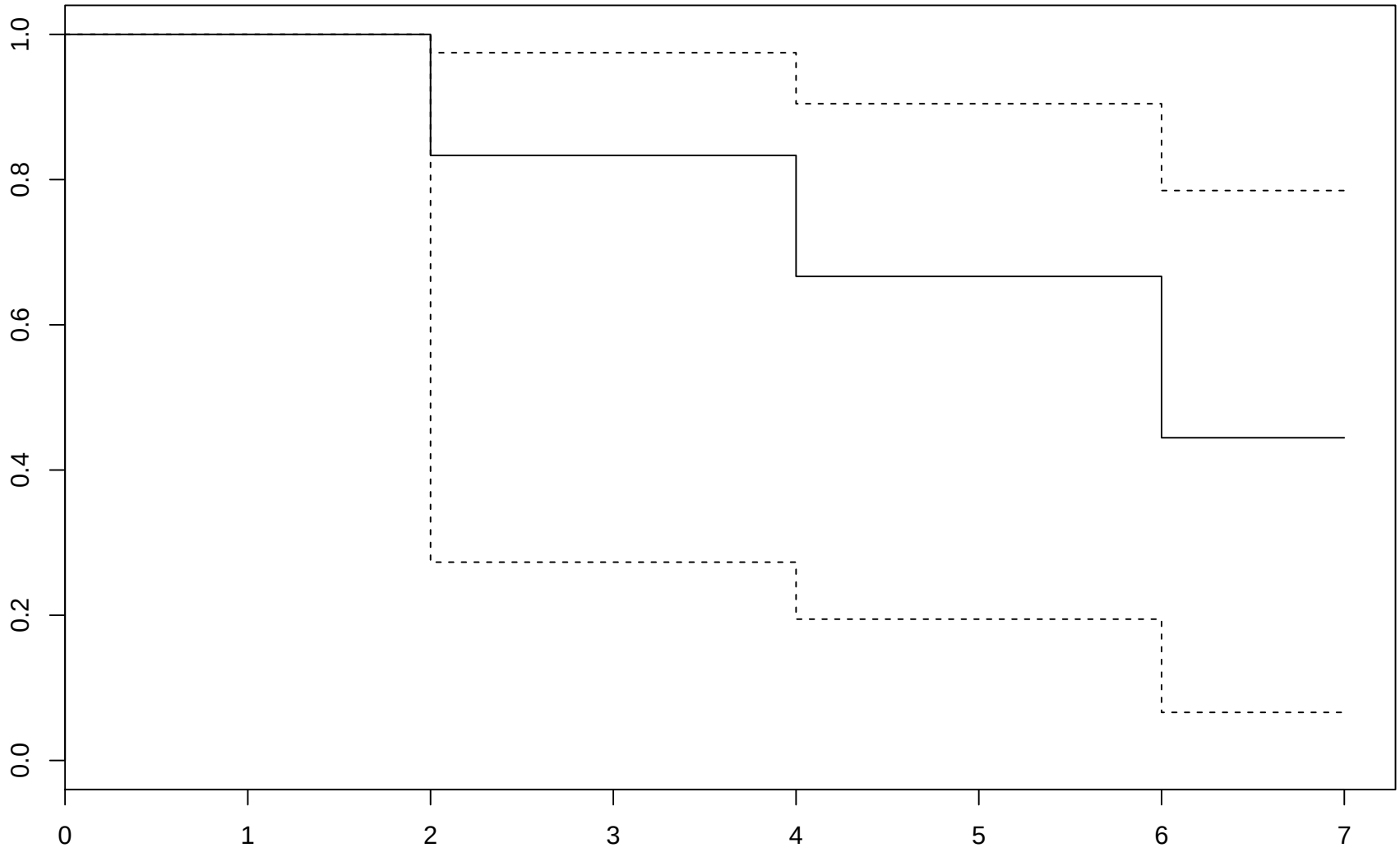
```
## Call: survfit(formula = Surv(tt, cens) ~ 1, conf.type =
##
##       n  events  median 0.95LCL 0.95UCL
##       6       3       6       2      NA
```

```
summary(result.km)
```

```
## Call: survfit(formula = Surv(tt, cens) ~ 1, conf.type =
##
##  time n.risk n.event survival std.err lower 95% CI upper
##     2      6       1    0.833   0.152        0.2731
##     4      5       1    0.667   0.192        0.1946
##     6      3       1    0.444   0.222        0.0662
```

Graph

```
plot(result.km)
```



Hazard estimators

- Nelson-Aalen estimator

$$H(t) = \sum_{t_i \leq t} \frac{d_i}{n_i}$$

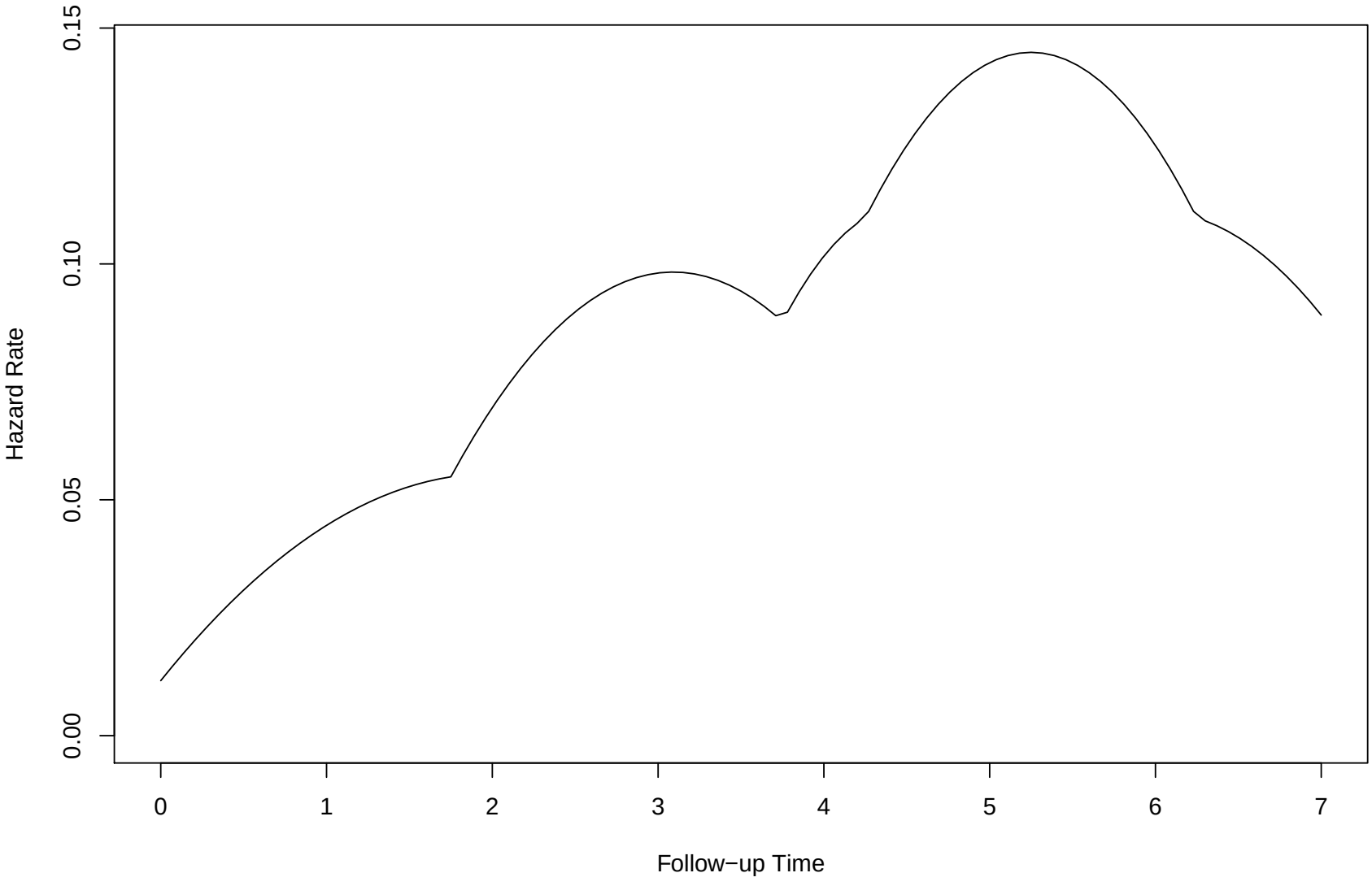
$$S(t) = e^{-H(t)}$$

Nelson-Aalen estimator in R

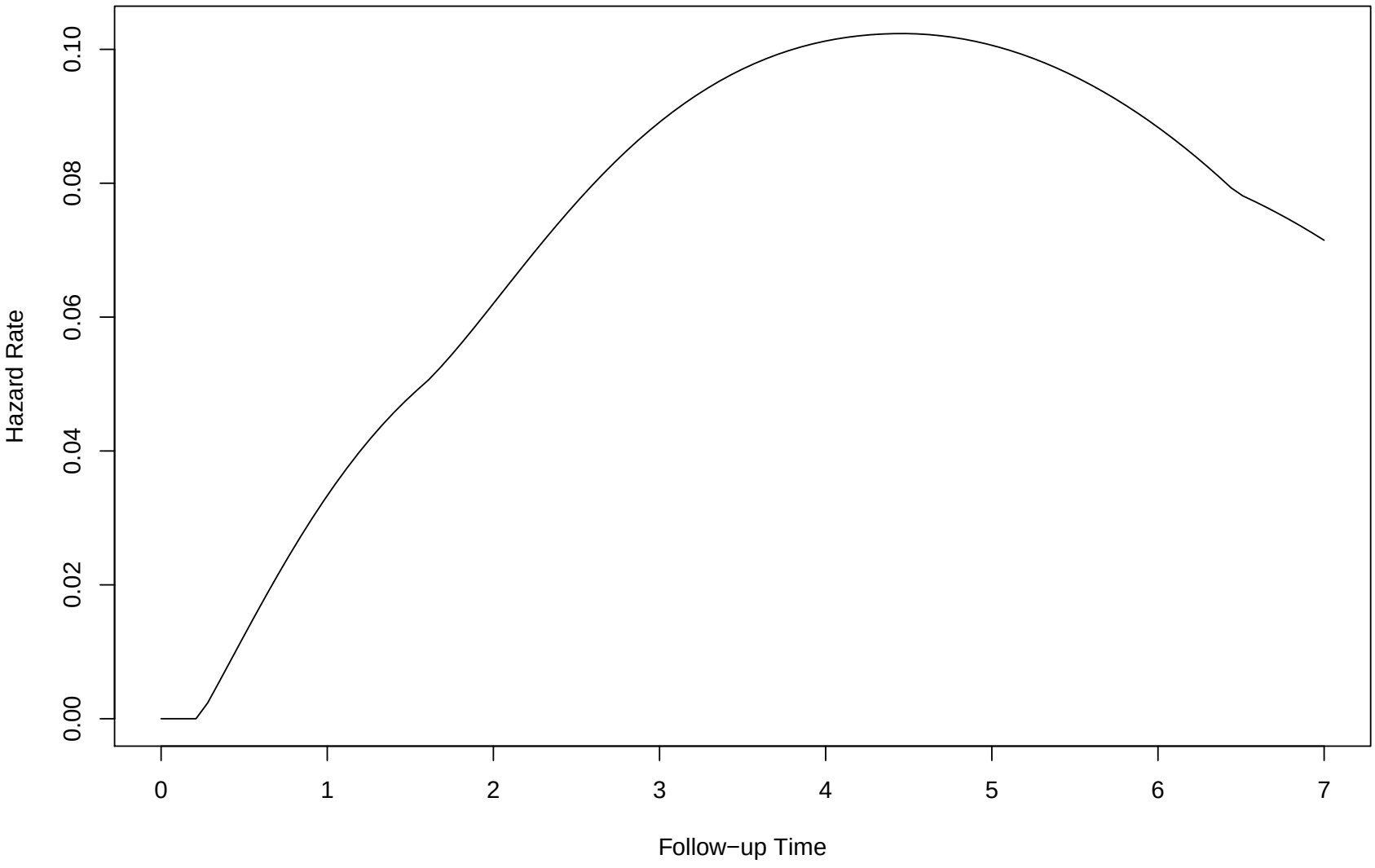
```
result.fh <- survfit(Surv(tt, cens) ~ 1, conf.type = "log-l
summary(result.fh)
```

```
## Call: survfit(formula = Surv(tt, cens) ~ 1, conf.type =
##     type = "fh")
##
##  time n.risk n.event survival std.err lower 95% CI upper
##     2      6       1    0.846   0.155        0.2401
##     4      5       1    0.693   0.200        0.1799
##     6      3       1    0.497   0.248        0.0585
```

Estimating hazard functions



Smoothed hazard



Truncation

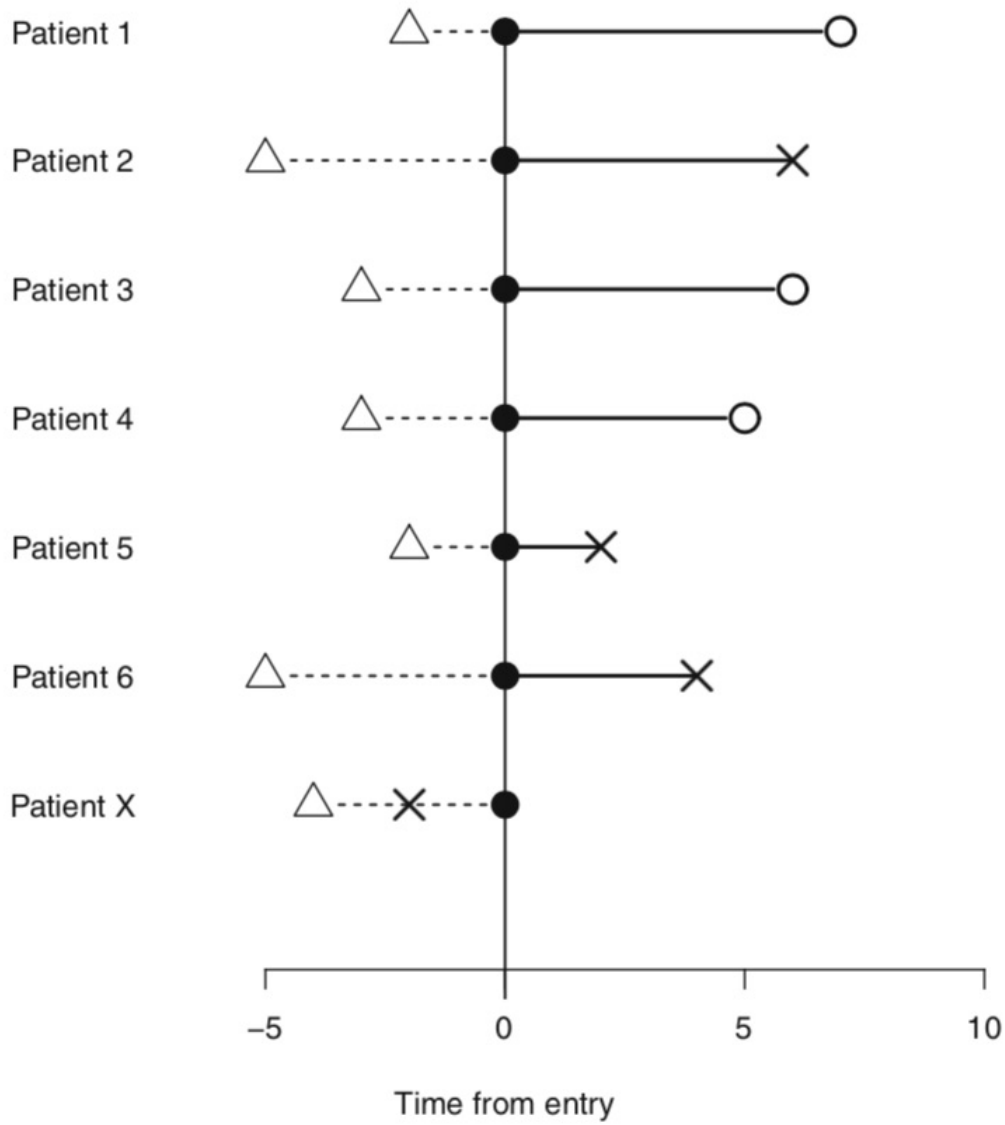


Figure 7: Left Truncation

Truncation (patient time)

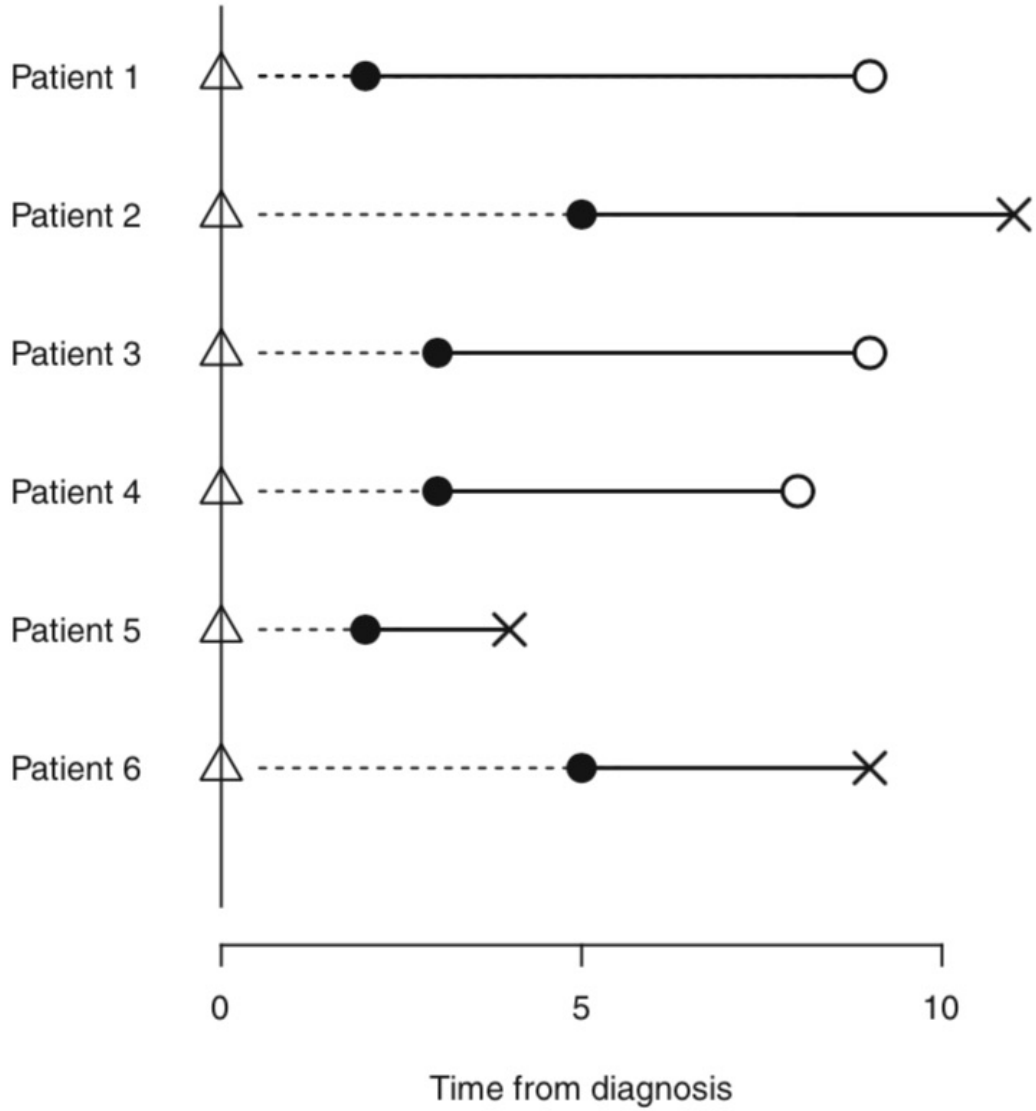


Figure 8: Left truncation in patient time

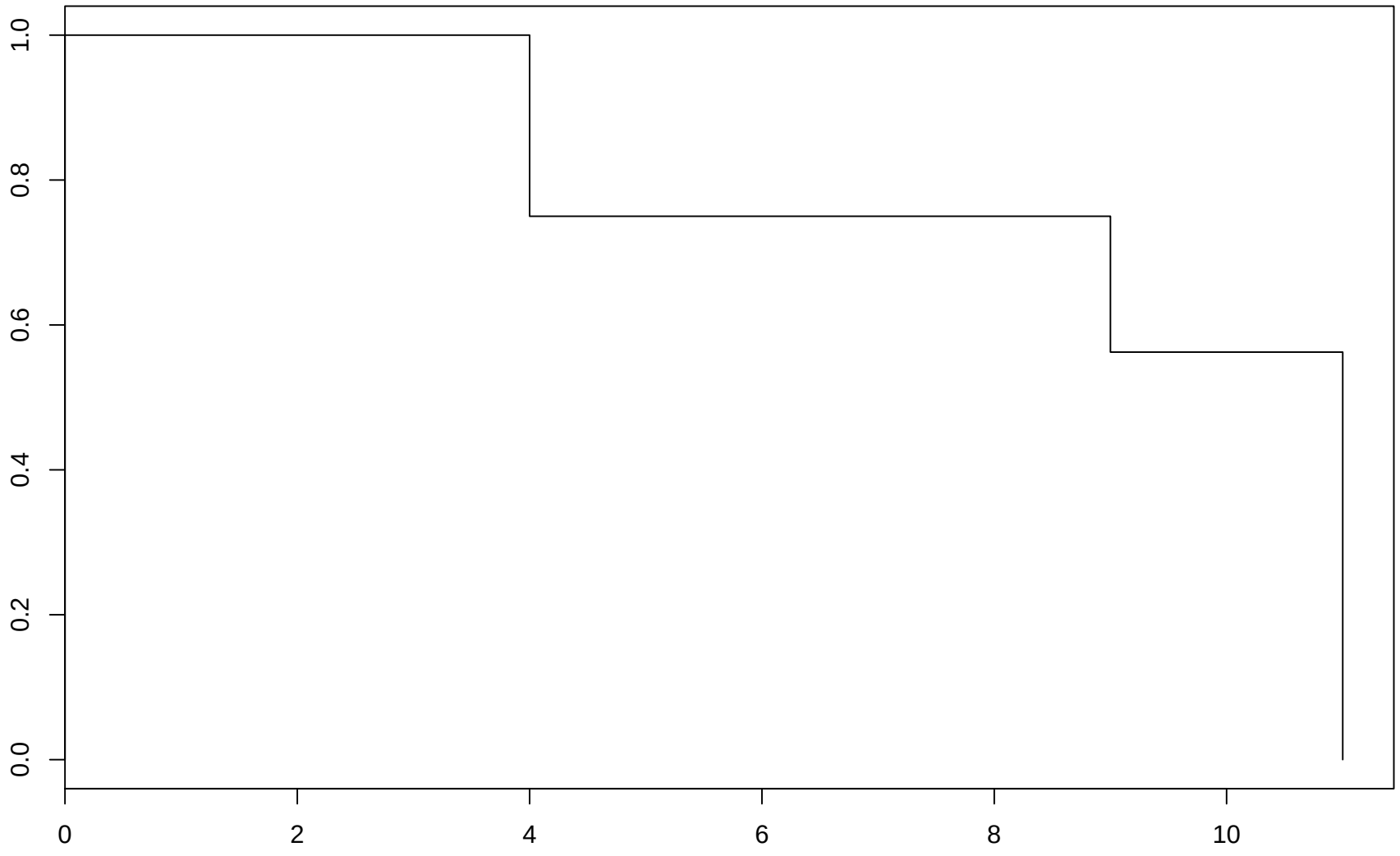
Left truncation in R

```
tt <- c(7, 6, 6, 5, 2, 4)
status <- c(0, 1, 0, 0, 1, 1)
backTime <- c(-2, -5, -3, -3, -2, -5)
tm.enter <- -backTime
tm.exit <- tt - backTime
result.left.trunc.km <- survfit(Surv(tm.enter, tm.exit, sta
summary(result.left.trunc.km)
```

```
## Call: survfit(formula = Surv(tm.enter, tm.exit, status,
##     1, conf.type = "none")
##
##  time n.risk n.event censored survival std.err
##     4      4       1        0    0.750   0.217
##     9      4       1        3    0.562   0.230
##    11      1       1        0    0.000     NaN
```

Left trunctation in R 2

```
plot(result.left.trunc.km)
```



Comparing survival curves

```
tt <- c(6, 7, 10, 15, 19, 25)
delta <- c(1, 0, 1, 1, 0, 1)
trt <- c(0, 0, 1, 0, 1, 1)
survdiff(Surv(tt, delta) ~ trt)
```

```
## Call:
## survdiff(formula = Surv(tt, delta) ~ trt)
##
##         N Observed Expected (O-E)^2/E (O-E)^2/V
## trt=0 3         2     1.08     0.776      1.27
## trt=1 3         2     2.92     0.288      1.27
##
##  Chisq= 1.3  on 1 degrees of freedom, p= 0.3
```

Real world example from historical demography

The demographic database of Umea University
http://www.ddb.umu.se

```
library(eha)
data(oldmort)
head(oldmort)
```

```
##          id  enter   exit event birthdate m.id f.id    s
## 1 765000603 94.510 95.813  TRUE  1765.490   NA   NA fema
## 2 765000669 94.266 95.756  TRUE  1765.734   NA   NA fema
## 3 768000648 91.093 91.947  TRUE  1768.907   NA   NA fema
## 4 770000562 89.009 89.593  TRUE  1770.991   NA   NA fema
## 5 770000707 89.998 90.211  TRUE  1770.002   NA   NA fema
## 6 771000617 88.429 89.762  TRUE  1771.571   NA   NA fema
##    ses.50 birthplace imr.birth   region
## 1 unknown     remote  22.20000    rural
## 2 unknown     parish  17.71845 industry
## 3 unknown     parish  12.70903    rural
```

Schematic representation of old age life in Swedent

Schematic representation of old age life in Swedent

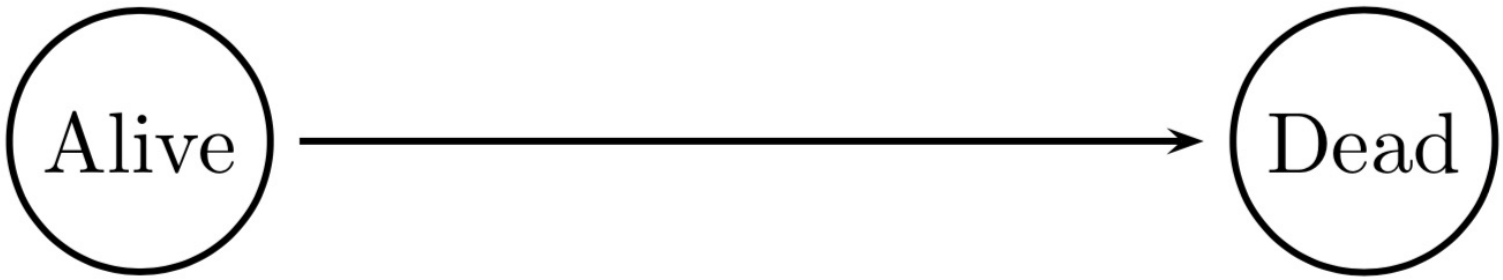
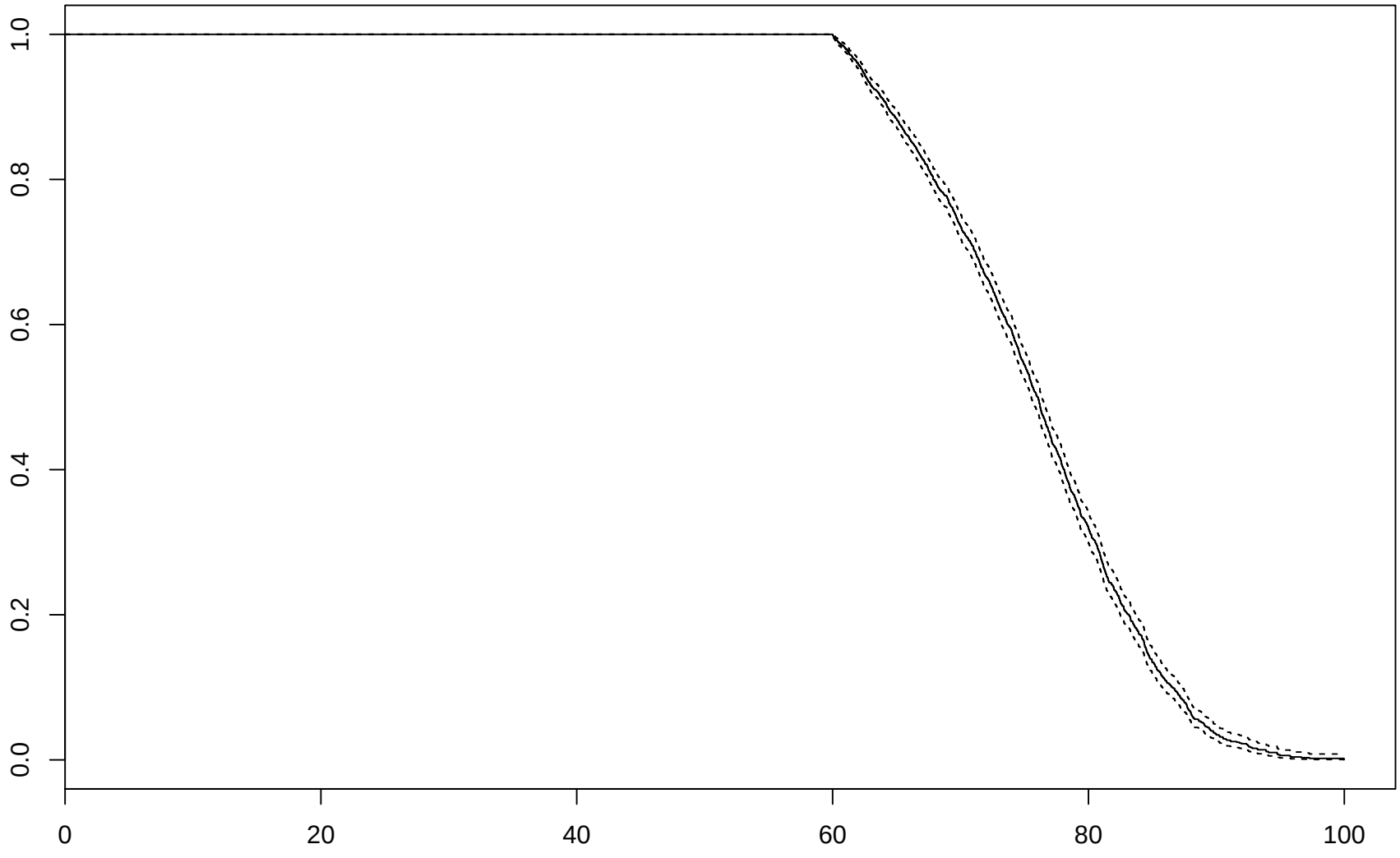


Figure 9: Life in Sweden in the 19th century

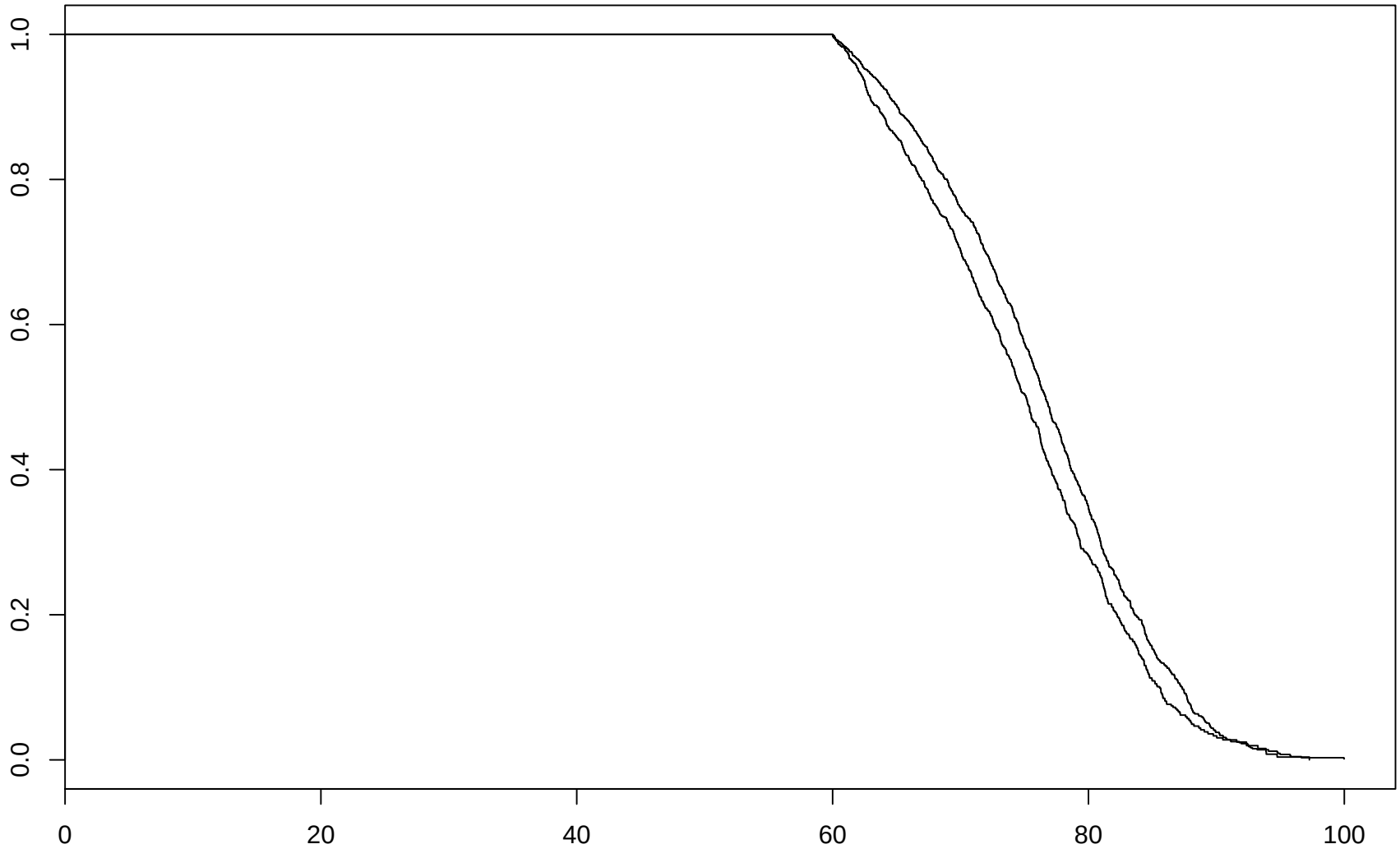
Analysis in R

```
with(oldmort, plot(survfit(Surv(enter, exit, event)~1)))
```



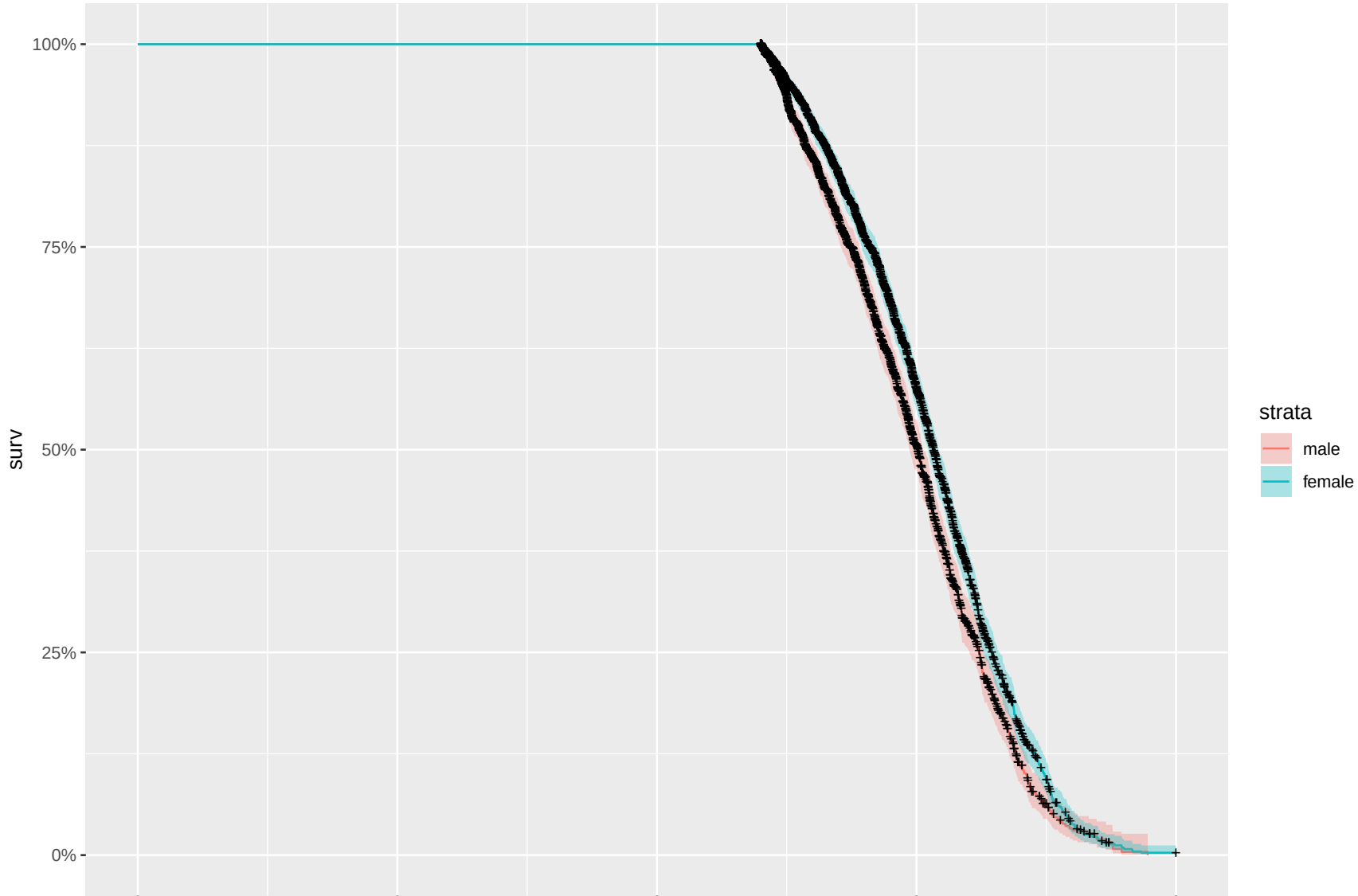
Survival by sex

```
with(oldmort, plot(survfit(Surv(enter, exit, event)~sex)))
```



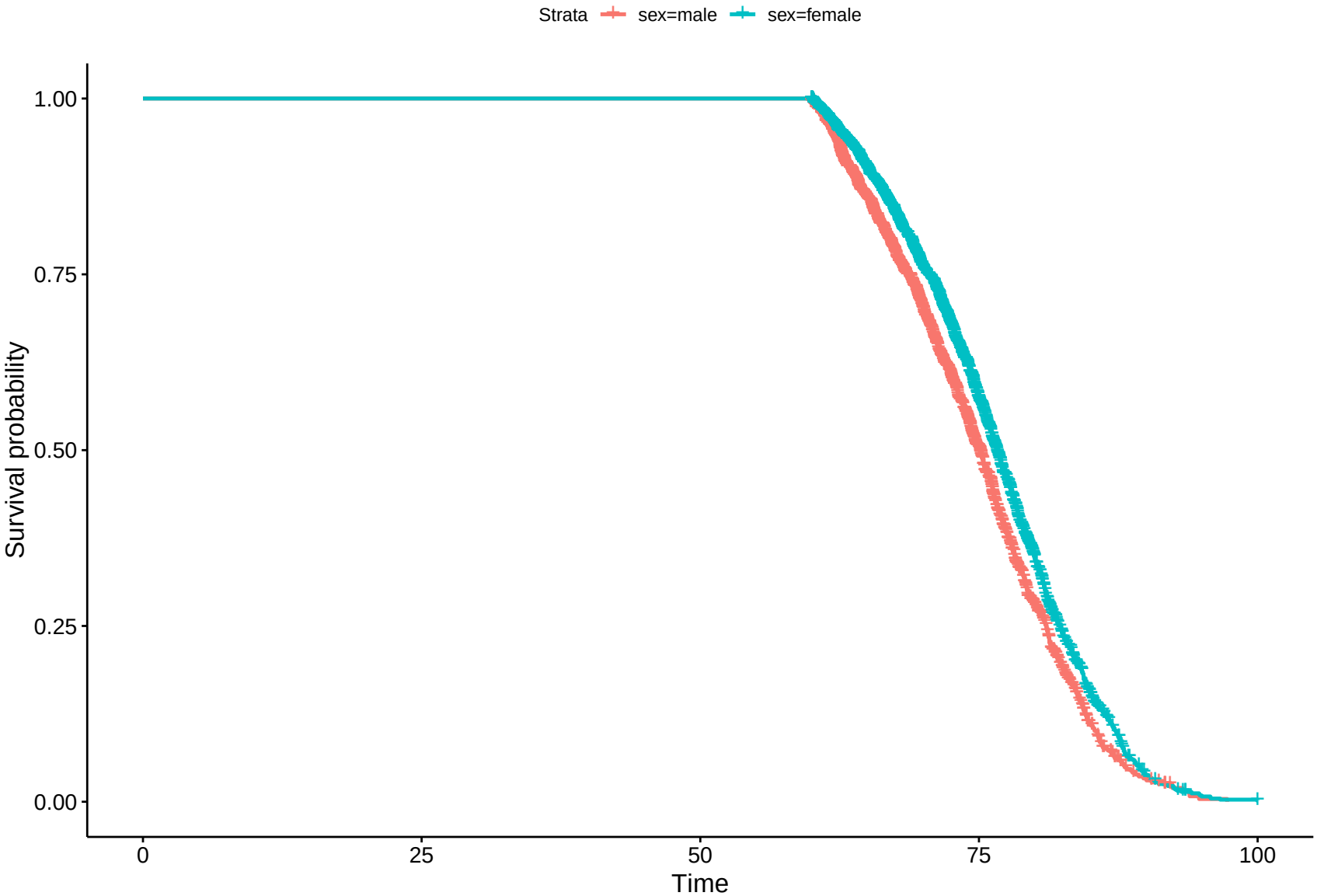
Using ggplot to plot slightly nicer graphs

```
oldm <- survfit(Surv(enter, exit, event)~sex)
autoplot(oldm)
```



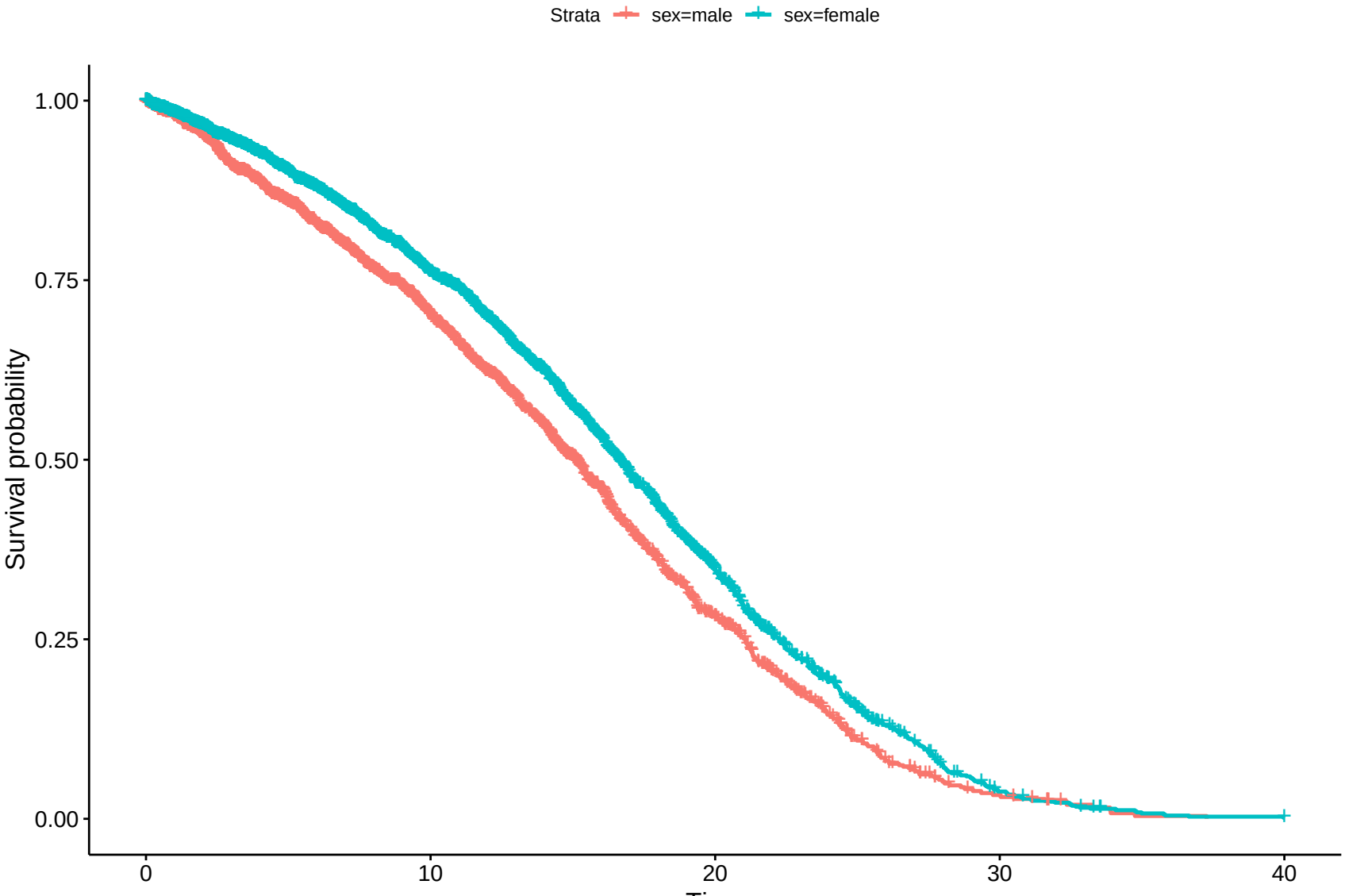
An alternative approach

```
ggsurvplot(oldm, data = oldmort)
```



Final adjustment to the graph

```
 oldm_mod <- survfit(Surv(enter - 60, exit - 60, event)~sex
ggsurvplot(oldm_mod, data = oldmort)
```



A teaser for tomorrow

This will not work:

```
survdiff(Surv(enter, exit, event)~sex, data = oldmort)
```