Introduction

STK4080 - 25. August 2016

S. O. Samuelsen

1. Censoring

2. Survival and hazard function

3. Estimation of survival, Kaplan-Meier

4. Log-rank test for difference in survival

5. Proportional hazards model, Cox-regression

6. Parametric likelihood

7. Counting processes & Martingales (briefly)

8. A few event history schemes

Survival times

or more generally: Time to an event

- Time until death

- Time until a machine stops working

- Time to disease

- Duration of marriage

- Duration of employment

- Age at sexual debut

Typical problem: **Censoring**:

Alive at end of follow-up

Example: Clinical trial

Start of study at time $t = 0$

- New patients are discovered and included in the stud

- Patients are follow-up until death,

- or to when they do not wish to participate in the study anymore

- or to the end of the study period

Example: clinical trial, contd.

Scheme: Death given by ● and censoring by ○.

Fig. on the left: calendar-time, Fig. on the right: time on study.
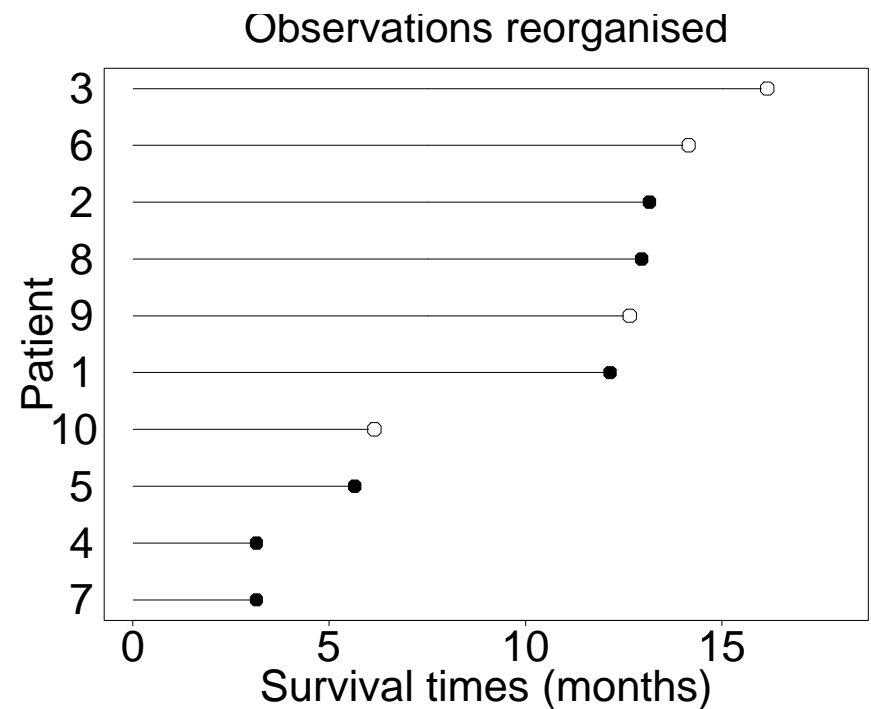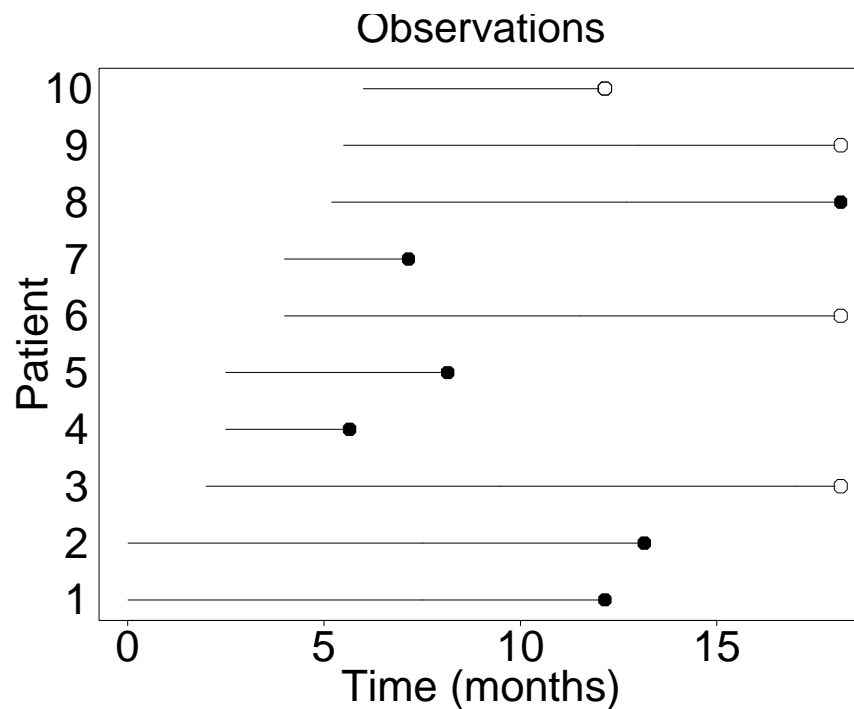
Survival times, formally

$$T_i = \text{survival time for ind. no. } i$$

$$C_i = \text{censoring time for ind. no. } i$$

Will not observe all T_i (or C_i), only

$$\widetilde{T}_i = \min(T_i, C_i) = \text{Censored lifetime for ind. no. } i$$

$$D_i = I(T_i = \widetilde{T}_i) = \text{Indicator of death for individ no. } i$$

The responses in survival analysis are the pairs $(\widetilde{T}_i, D_i)$, i.e. the combination of a continuous variable $\widetilde{T}_i$ and a binary variable D_i.

For instance regression on $\widetilde{T}_i$ disregarding D_i does not make sense.

Need specialized methods for survival data!

Distribution functions for survival time T

- Density $f(t)$ given from $\mathrm{P}(T \in [t, t+\Delta >) \approx f(t)\Delta$

- Survival function $S(t) = P(T > t)$

- Hazard function $\alpha(t)$ given from
  $\mathrm{P}(T \in [t, t+\Delta > | T \geq t) \approx \alpha(t)\Delta$

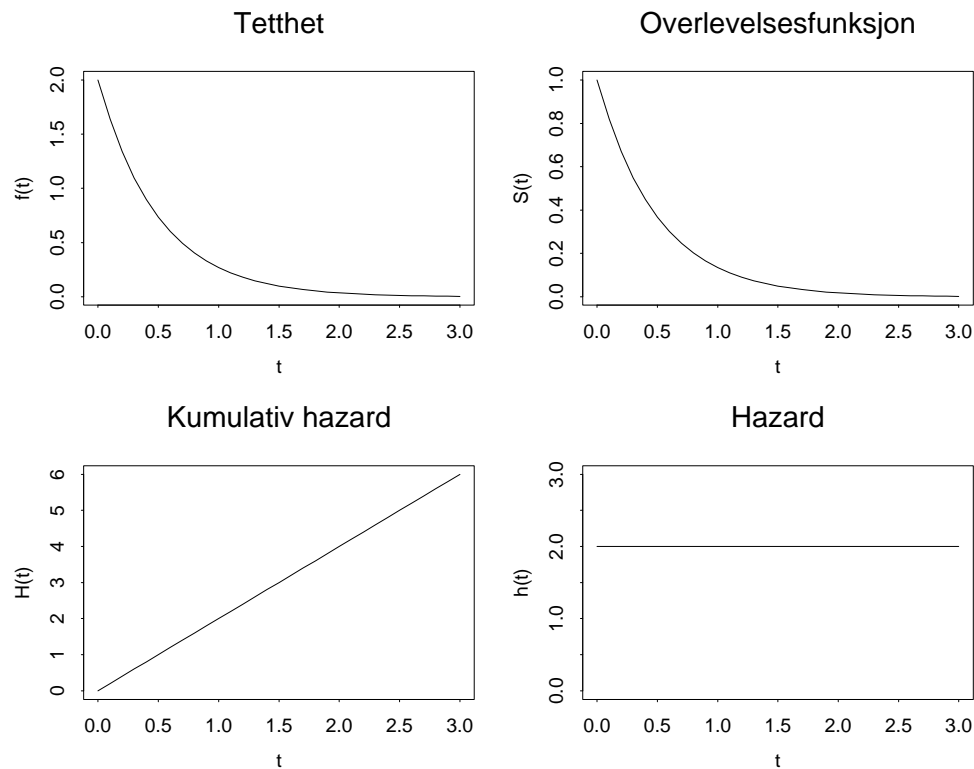- Cumulative hazard $A(t) = \int_0^t \alpha(s)ds$

Writes $S(t)$ for "Survival"

Interpretation hazard: Prob. for death in small interval around t (divided by Δ) *given* alive by t.

Relationships:

- $\alpha(t) = f(t)/S(t)$

- $S(t) = \exp(-A(t))$

- $A(t) = -\log(S(t))$

For the exponential distribution

- $\alpha(t) = \nu$, i.e. constant

- $A(t) = \nu t$

- $S(t) = \exp(-\nu t)$
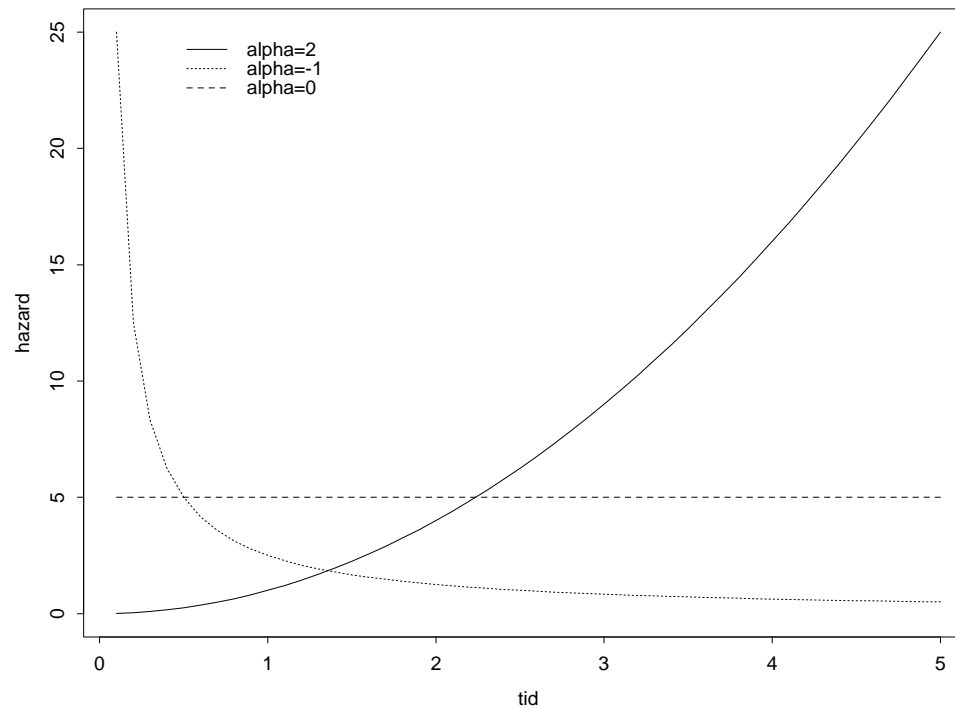
- $f(t) = \nu \exp(-\nu t)$

Weibull distributions: $\alpha(t, b, k) = bt^{k-1}$

With $k = 1$: Exponential distribution, $\alpha(t) = $ constant

Wit $k > 1$: Increasing hazard

With $k < 1$: Decreasing hazard

Kaplan-Meier estimator for the survival function

Define, with $I()$ the indicator function,

$$Y(t) = \text{ number "at risk" by } = \sum_{i=1}^{n} I(\widetilde{T}_i \geq t)$$

May then estimate $S(t)$ by the Kaplan-Meier estimator

$$
\begin{aligned}
\hat{S}(t) \;&= \prod_{\widetilde{T}_i \leq t}[1 - \frac{D_i}{Y(\widetilde{T}_i)}] \\
&= (1 - \frac{D_1}{Y(\widetilde{T}_1)})(1 - \frac{D_2}{Y(\widetilde{T}_2)}) \cdots (1 - \frac{D_k}{Y(\widetilde{T}_k)})
\end{aligned}
$$

when the data are indexed so that $\widetilde{T}_1 < \widetilde{T}_2 < \cdots < \widetilde{T}_n$
$\widetilde{T}_{k-1} < t \leq \widetilde{T}_k$.

Explanation for Kaplan-Meier:

$$\hat{P}(\text{Die at time } \widetilde{T}_i | \text{ Alive just before } \widetilde{T}_i) = D_i / Y(\widetilde{T}_i)$$

$$\hat{P}(\text{Survive to time } \widetilde{T}_i | \text{ Alive just before } \widetilde{T}_i) = 1 - D_i / Y(\widetilde{T}_i)$$

Thus, since nobody dies in the intervals $(\widetilde{T}_{j-1}, \widetilde{T}_j)$

$$
\begin{aligned}
\hat{P}(\text{Survive to } \widetilde{T}_i) = \quad & \hat{P}(\text{Survive to } \widetilde{T}_i | \text{ Survive to } \widetilde{T}_{i-1}) \\
& *\hat{P}(\text{Survive to } \widetilde{T}_{i-1} | \text{ Survive to } \widetilde{T}_{i-2}) \\
& * \cdots \cdot \\
& *\hat{P}(\text{Survive to } \widetilde{T}_1) \\
= \quad & \hat{S}(\widetilde{T}_i)
\end{aligned}
$$

We may say that the Kaplan-Meier estimator is non-parametric since we have made no parametric assumptions.

A constructed data set:

Censored lifetimes $\widetilde{T}_i = 2, 3, 5^*, 6, 7^*, 8, 8, 10^*, 12$
where * indicates a censored value of T_i with $D_i = 0$.

Time t_j	At risk Y_i	Death D_i	D_j/Y_i	$1 - D_i/Y_i$	$\hat{S}(t)$
0	9	0	0	1	1
2	9	1	$\frac{1}{9}$	$\frac{8}{9}$	$\frac{8}{9}$
3	8	1	$\frac{1}{8}$	$\frac{7}{8}$	$\frac{8}{9}\frac{7}{8} = \frac{7}{9}$
5	7	0	0	1	$\frac{7}{9}$
6	6	1	$\frac{1}{6}$	$\frac{5}{6}$	$\frac{7}{9}\frac{5}{6} \approx 0.648$
7	5	0	0	1	0.648
8	4	2	$\frac{2}{4}$	$\frac{1}{2}$	0.324
10	2	0	0	1	0.324
12	1	1	1	0	0

R-calculation of Kaplan-Meier

```
> tid<-c(2, 3, 5, 6, 7, 8, 8, 10, 12)
> d<-c(1, 1, 0, 1, 0, 1, 1, 0, 1)
> library(survival)
> survtest<-survfit(Surv(tid,d)~1)
> survtest
Call: survfit(formula = Surv(tid, d)~1)
      n   events   median 0.95LCL 0.95UCL
      9        6        8       6     Inf
> names(survtest)
 [1] "n"          "time"       "n.risk"     "n.event"    "surv"      "type"
 [7] "std.err"    "upper"      "lower"      "conf.type" "conf.int"  "call"
> cbind(survtest$time,survtest$n.risk,survtest$n.event,survtest$surv)
     [,1] [,2] [,3]       [,4]
[1,]    2    9    1 0.8888889
[2,]    3    8    1 0.7777778
[3,]    5    7    0 0.7777778
[4,]    6    6    1 0.6481481
[5,]    7    5    0 0.6481481
[6,]    8    4    2 0.3240741
[7,]   10    2    0 0.3240741
[8,]   12    1    1 0.0000000
```
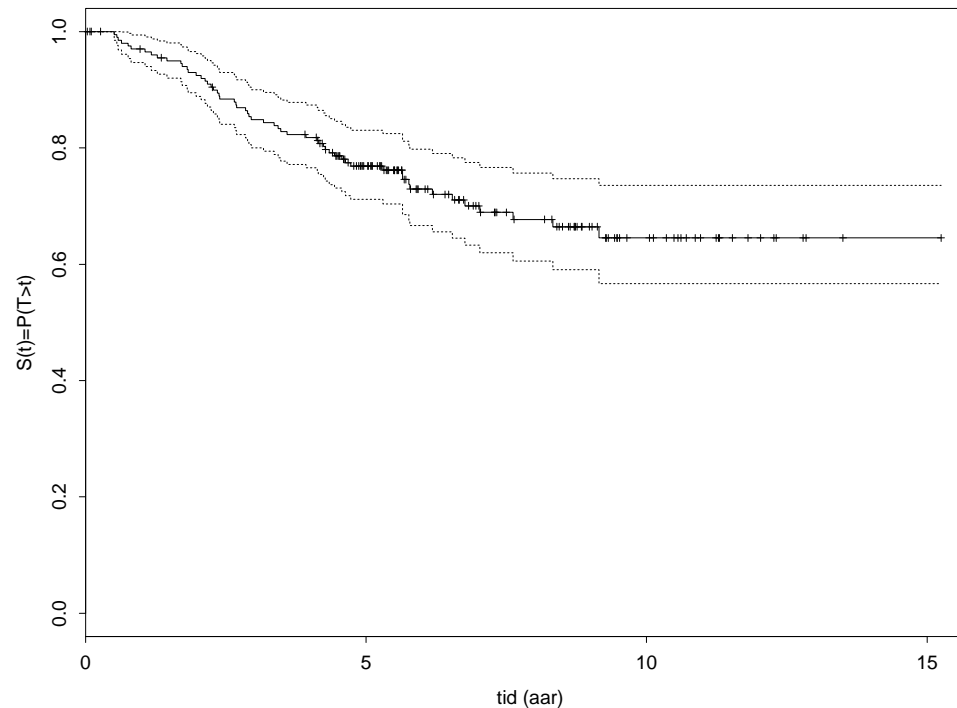
R-plot of Kaplan-Meier

```
> plot(survfit(Surv(tid,d)~1))
```
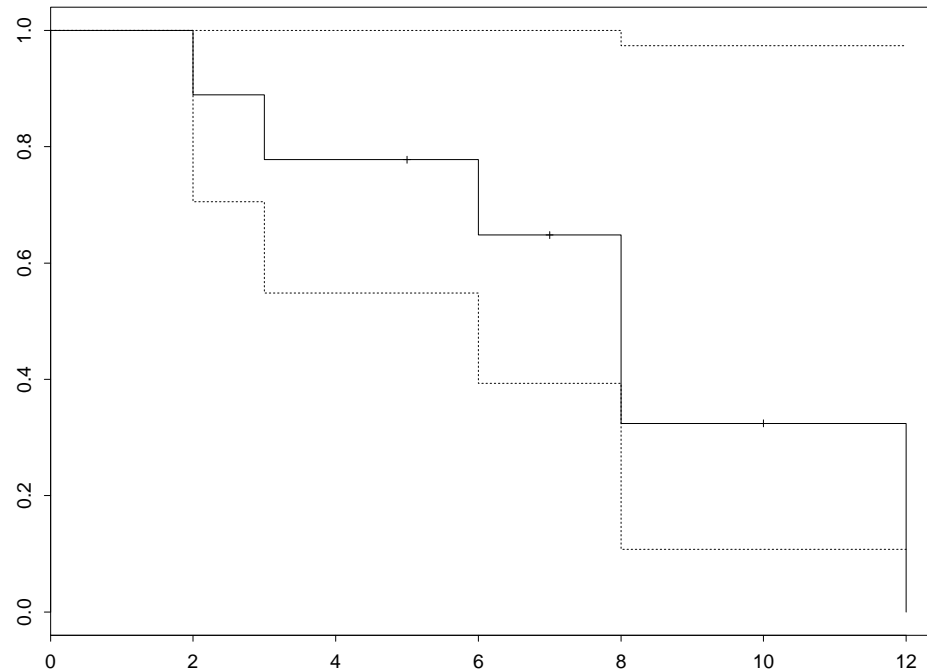
Example: 205 Danish melanoma patients

$$T = \text{Time to death from melanoma}$$

$$C = \text{Time to end of follow-up or death of other cause}$$


```
> survfit(Surv(time,dead)~1)
Call: survfit(formula = Surv(time, dead)~1)
      n   events   median 0.95LCL 0.95UCL
    205       57      Inf     Inf     Inf
```

Cumulative hazard: Nelson-Aalen estimator

Estimation of cumulative hazard: Nelson-Aalen estimator

$$\hat{A}(t) = \sum_{\widetilde{T}_i \leq t} \frac{D_i}{Y(\widetilde{T}_i)}$$

(May alternatively estimate $S(t)$ by $\exp(-\hat{A}(t))$, or $A(t)$ by $-\log(\hat{S}(t))$.)

Estimation of hazard $\alpha(t)$ and density $f(t)$:
Possible, but more difficult.

Comparison of two groups

Example: Is survival better with new therapy?

$$(\widetilde{T}_{i1}, D_{i1}); i = 1, ..., n_1 \quad \text{Survival data with trad. therapy}$$

$$(\widetilde{T}_{i2}, D_{i2}); i = 1, ..., n_2 \quad \text{Survival data with new therapy}$$

$$\hat{S}_k(t) = \text{ Kaplan-Meier estimator in group } k, k = 1, 2$$

Compare

- Graphically: Plot $\hat{S}_1(t)$ and $\hat{S}_2(t)$

- Hypothesis test: *Log-rank-test*

Graphical comparison in R

```
time<-c(2,3,4,7,10,22,28,29,32,37,40,41,54,61,63,71,127,140,146,158,
    167,182,2,6,12,54,56,68,89,96,96,125,128,131,140,141,143,145,
    146,148,162,168,173,181)
d<-c(rep(1,16),rep(0,6),c(1,1,1,1,0,1,1,1,1,0,0,0,0,0,1,0,1,0,0,1,0,0))
gr<-c(rep(1,22),rep(2,22))
plot(survfit(Surv(time,d)~gr),lty=1:2,xlab="Time (months)",ylab="Surviv
legend(1,0.2,c("Control","Treat"),lty=1:2,bty="n")
```
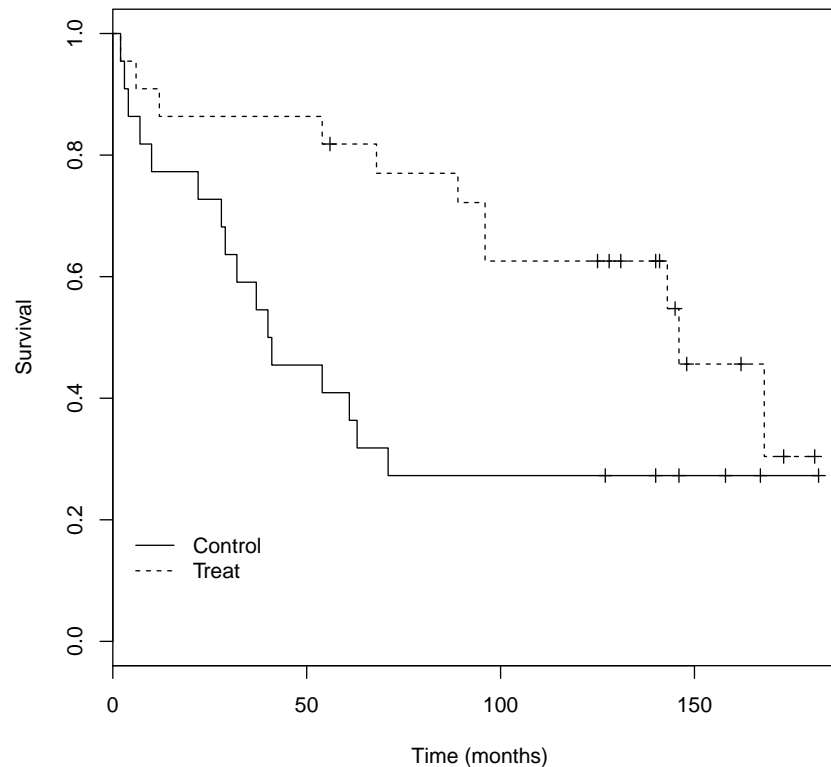
Log-rank test

$$O_1 = \text{No. of observed deaths in control group}$$

$$O_2 = \text{No. of observed deaths in treatment group}$$

$$E_k = \text{"Expected" no. deaths in group } k \text{ under } H_0:\text{Same mortality}$$

$$= \sum_j n_{kj} \frac{m_{1j} + m_{2j}}{n_{1j} + n_{2j}}$$

where $n_{kj} = $ "no. at risk" and m_{kj} no. deaths at time t_j i group k.
Tests the hypothesis by

$$Z = \frac{O_2 - E_2}{\sqrt{\text{Var}(O_2 - E_2)}} \sim \text{N}(0, 1) \text{ under } H_0$$

or equivalently

$$Z^2 = \frac{(O_2 - E_2)^2}{\text{Var}(O_2 - E_2)} \sim \chi_1^2 \text{ under } H_0$$

Log-rank test, contd.

A somewhat conservative test (too large p-values) is given by

$$X^2 = \frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)^2}{E_2} \sim \chi_1^2 \text{ under } H_0$$

```
survdiff(Surv(tid,d)~gr)
Call: survdiff(formula = Surv(tid, d) ~ gr)

        N Observed Expected (O-E)^2/E (O-E)^2/V
gr=1 22        16     10.6      2.73      4.66
gr=2 22        11     16.4      1.77      4.66

 Chisq= 4.7  on 1 degrees of freedom, p= 0.0309
```

The proportional hazards model: 1. One covariate

Hazard rate for subject with one covariate X:

$$\alpha_X(t) = \alpha_0(t)\exp(\beta X)$$

where baseline hazard $\alpha_0(t)$ is the hazard for subject with $X = 0$.

Interpretation: Hazard rate ratio (or loosely Relative Risk),

$$\text{HR} = \exp(\beta(X_1 - X_0)) = \frac{\alpha_{X_1}(t)}{\alpha_{X_0}(t)}$$
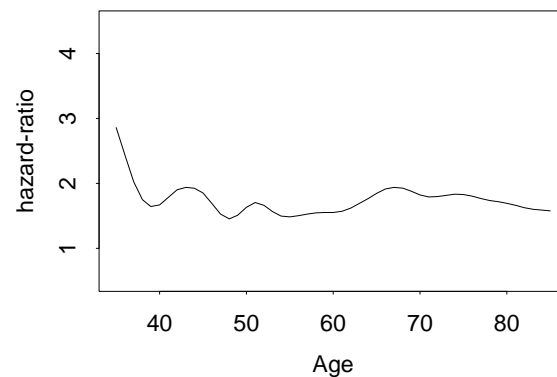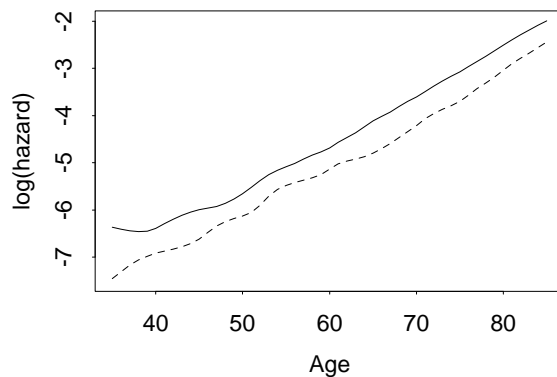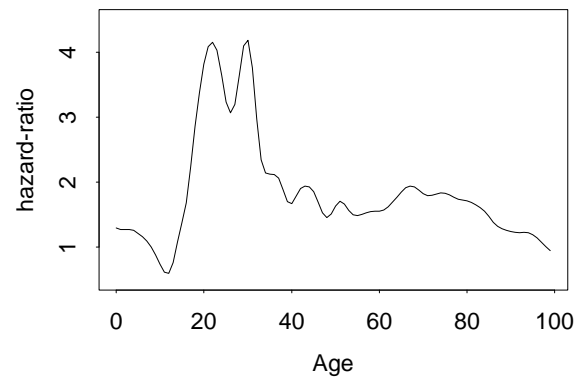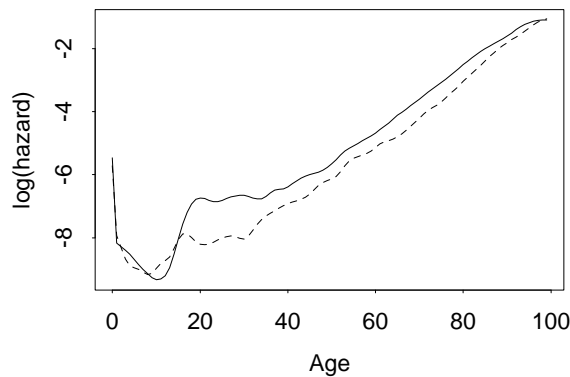
In particular with X binary

$$\text{HR} = \exp(\beta) = \frac{\alpha_1(t)}{\alpha_0(t)}$$

Example: Mortality rates among men and women, Statistics Norway, 2000, smoothed.

Binary covariate X indicator of men.

Prop. hazard model **not** valid in age interval 0-100 years

Prop. hazard model roughly valid in interval 40-85 years with $HR \approx 1.8$.

Example 1: Melanomadata

$T =$ time to death from melanoma

hazard $\alpha_X(t) = \alpha_0(t) \exp(\beta X)$

$X =$ indicator of ulceration,

$HR = \frac{\alpha_1(t)}{\alpha_0(t)} = \exp(\beta) =$ hazard ratio between those with and without ulceration.

$X_1 =$ tumor thickness (mm) subject 1,

$X_2 =$ thickness (mm) subject 2$= X_1 + 1$ mm,

$HR = \exp(\beta) =$ rate ratio w. 1 mm difference.

Proportional hazards model: 2. Several covariates

Hazard rate for individual with covariate vector
$X = (X_1, X_2, ..., X_p)$

$$\alpha_X(t) = \alpha_0(t) \exp\{\beta_1 X_1 + \beta_2 X_2 + ... + \beta_p X_p\}$$

where baseline hazard $\alpha_0(t)$ is hazard function for individual
with all $X_1 = X_2 = ... = X_p = 0$.

Interpretation: Hazard rate ratio (HR)

Another subject with $X' = (X'_1, X'_2, ..., X'_p)$ where $X'_1 = 1$,
$X_1 = 0$ and $X'_j = X_j$ otherwise:

$$\text{HR}_1 = \exp\{\beta_1\} = \frac{\alpha_{X'}(t)}{\alpha_X(t)}$$

Example 1: Melanomadata

$$\alpha_X(t) = \alpha_0(t) \exp(\beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4)$$

X_1 = sex (M=1, F=0)

X_2 = indicator of ulceration,

X_3 = age,

X_4 = thickness (mm)

$X = (X_1, 0, X_3, X_4)$

$X' = (X_1, 1, X_3, X_4)$

$HR = \frac{\alpha_X(t)}{\alpha_{X'}(t)} = \exp(\beta_2)$ = hazard ratio between those with and without ulceration **adjusted** for sex, age and thickness.
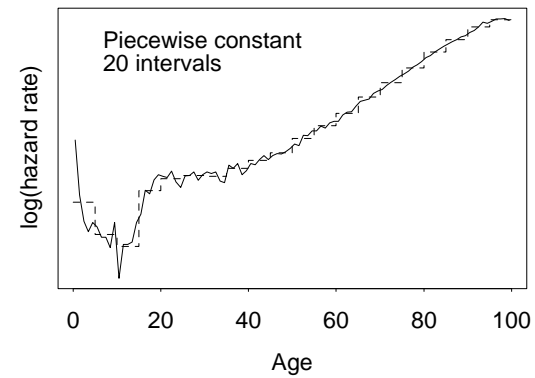
Estimation in the proportional hazards model

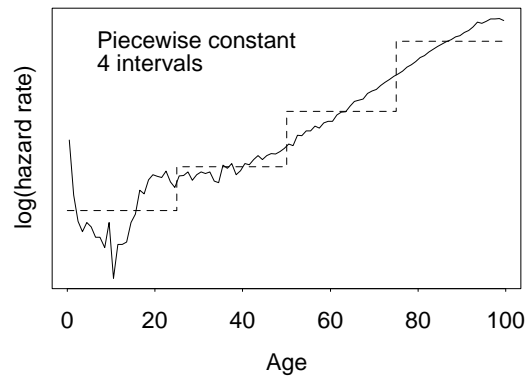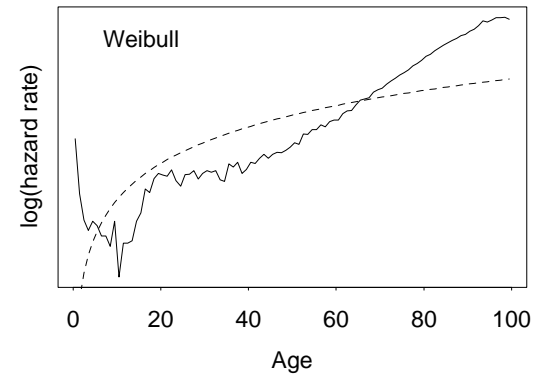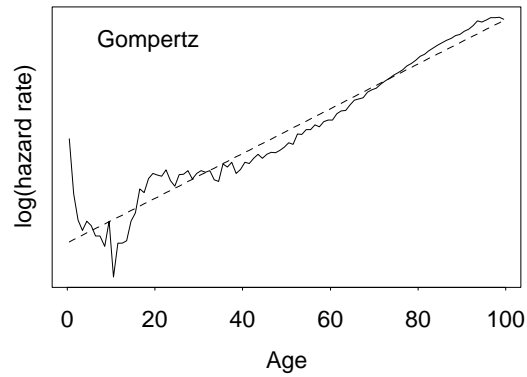- With baseline hazard $\alpha_0(t) = \alpha_0(t, \theta)$ parametrically specified
$\rightarrow$ by likelihood for censored data.

$$\text{Gompertz: } \alpha_0(t, \theta = (\gamma, \lambda)) = \lambda \gamma^t$$

$$\text{Weibull: } \alpha_0(t, \theta = (\gamma, \lambda)) = \lambda^\gamma t^{\gamma-1}$$

- With baseline $\alpha_0(t) = \alpha_{0j}$ piecewise constant on $(t_{j-1}, t_j]$
$\rightarrow$ by Poisson regression.

- With baseline hazard $\alpha_0(t)$ arbitrary function
$\rightarrow$ by Cox-regression.

Comparison of different types of baseline hazards

Cox' Regression:

Death at t_i. Let

$$
\begin{aligned}
L_i(\beta) &= \mathrm{P}(\text{Subject } i \text{ died at } t_i | i \in \mathcal{R}(t_i), \text{ death at } t_i) \\
&= \frac{\alpha_i(t_i)}{\sum_{k \in \mathcal{R}(t_i)} \alpha_k(t_i)} \\
&= \frac{\exp(\beta X_i)\alpha_0(t_i)}{\sum_{k \in \mathcal{R}(t_i)} \exp(\beta X_k)\alpha_0(t_i)} \\
&= \frac{\exp(\beta X_i)}{\sum_{k \in \mathcal{R}(t_i)} \exp(\beta X_k)}
\end{aligned}
$$

where

- $\alpha_i(t) = \alpha_0(t)\exp(\beta X_i) = $ hazard of subject i at t
- $\mathcal{R}(t) = $ subjects under observation at t = riskset at t.

Note $L_i(\beta)$ depend on β only,

not on the baseline hazard $\alpha_0(t)$.

Cox' Partial likelihood:

Assume subject i died at t_i, $i = 1, \ldots, d$.
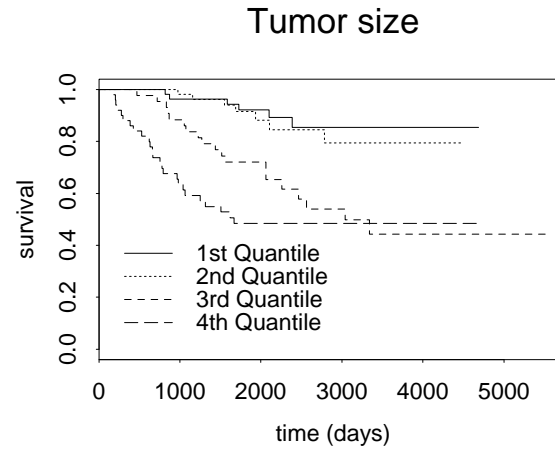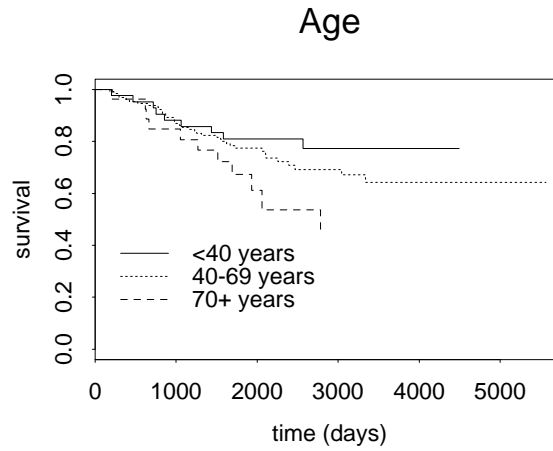
Estimate β by maximizing (Cox, 1972)

$$
\begin{aligned}
L(\beta) &= \prod_{i=1}^{d} L_i(\beta) \\
&= L_1(\beta) L_2(\beta) \ldots L_d(\beta)
\end{aligned}
$$

Note: We may estimate β and $HR = \exp(\beta)$ without saying anything about the baseline $\alpha_0(t)$.

The partial likelihood behaves as a usual likelihood.

In particular standard errors of Cox-estimator $\hat{\beta}$ and confidence intervals for $\hat{HR} = \exp(\hat{\beta})$ are produced "automatically".

Example 1: Melanomadata
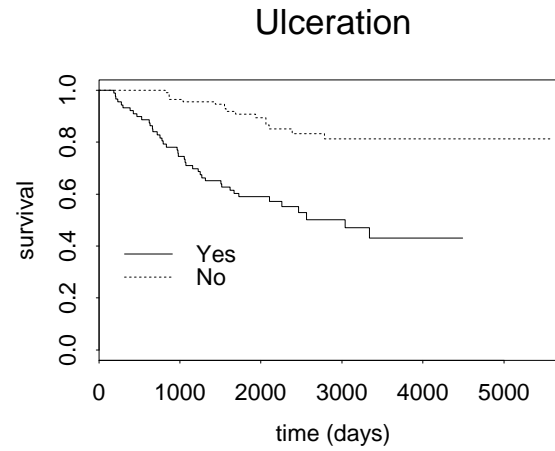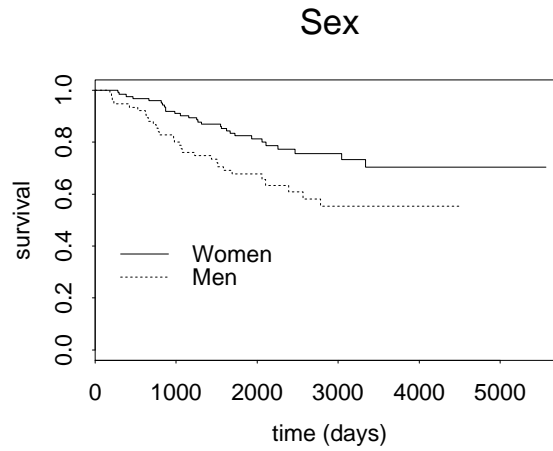
Example 1: Melanomadata

Variable	$\hat{\beta}$	$\text{se}(\hat{\beta})$	Z-value	p-value
tumorsize (mm)	0.11	0.04	2.89	0.004
ulceration	1.16	0.31	3.76	0.0002
sex (F=0,M=1)	0.43	0.27	1.62	0.11
age (years/10)	0.12	0.08	1.47	0.14

Variable	$\hat{\text{HR}} = \exp(\hat{\beta})$	$\hat{\text{HR}}_L$	$\hat{\text{HR}}_U$
tumorsize (mm)	1.12	1.04	1.20
ulceration	3.20	1.75	5.88
sex (F=0,M=1)	1.54	0.91	2.60
age (years/10)	1.13	0.96	1.33

R-code and print-out:

```
> coxph(Surv(time,dead)~sex+ulcer+age+thickn,data=mel)
Call:
coxph(formula = Surv(time, dead) ~ sex + ulcer + age + thickn, data = m

          coef exp(coef) se(coef)      z       p
   sex  0.4328     1.542   0.2674  1.62 0.11000
 ulcer -1.1645     0.312   0.3098 -3.76 0.00017
   age  0.0122     1.012   0.0083  1.47 0.14000
thickn  0.1089     1.115   0.0377  2.89 0.00390

Likelihood ratio test=41.6  on 4 df, p=2e-08  n= 205
```

More R-code and print-out:

```
> summary(coxph(Surv(time,dead)~sex+ulcer+age+thickn,data=mel))
coxph(formula = Surv(time, dead) ~ sex + ulcer + age + thickn, data = me

  n= 205
          coef exp(coef) se(coef)      z        p
    sex  0.4328     1.542   0.2674  1.62 0.11000
  ulcer -1.1645     0.312   0.3098 -3.76 0.00017
    age  0.0122     1.012   0.0083  1.47 0.14000
 thickn  0.1089     1.115   0.0377  2.89 0.00390


        exp(coef) exp(-coef) lower .95 upper .95
    sex     1.542      0.649     0.913     2.604
  ulcer     0.312      3.204     0.170     0.573
    age     1.012      0.988     0.996     1.029
 thickn     1.115      0.897     1.036     1.201


Rsquare= 0.184   (max possible= 0.937 )
Likelihood ratio test= 41.6  on 4 df,    p=2e-08
Wald test            = 39.4  on 4 df,    p=5.72e-08
Score (logrank) test = 46.7  on 4 df,    p=1.79e-09
```

Likelihood for right-censored data

Assume that lifetimes T_i stem from a distribution with density $f(t; \theta)$, survival function $S(t; \theta)$ and hazard $\alpha(t; \theta)$.

Right-censored obs: $\widetilde{T}_i = \min(C_i, T_i)$ and $D_i = I(T_i = \widetilde{T}_i)$.

Likelihood

$$L(\theta) = \prod_{i=1}^{n} L_i(\theta)$$

where the the likelihood contribution $L_i(\theta)$ is given by

Exact observed $(D_i = 1): \quad L_i(\theta) = f(\widetilde{T}_i; \theta) = \alpha(\widetilde{T}_i; \theta) S(\widetilde{T}_i; \theta)$

Right censored $(D_i = 0): \quad L_i(\theta) = \mathbf{P}(T_i > \widetilde{T}_i) = S(\widetilde{T}_i; \theta)$

Thus we can summarize the likelihood contribution as

$$L_i(\theta) = \alpha(\widetilde{T}_i; \theta)^{D_i} S(\widetilde{T}_i; \theta)$$

Example: Exponential distribution

Hazard: $\alpha(t) = \nu$ (constant in time)

Survival function $S(t) = \exp(-\nu t)$

Likelihood contribution: $L_i(\nu) = \nu^{D_i} \exp(-\nu \widetilde{T}_i)$

Likelihood $L(\nu) = \prod_{i=1}^{n} \nu^{D_i} \exp(-\nu \widetilde{T}_i) = \nu^{D_\bullet} \exp(-\nu T_\bullet)$
where

$$D_\bullet = \sum_{i=1}^{n} D_i = \text{ Total no. of deaths}$$

and

$$T_\bullet = \sum_{i=1}^{n} \widetilde{T}_i = \text{ Total observation time}$$

The likelihood is maximized for the occurrence / exposure rate

$$\hat{\nu} = \frac{D_\bullet}{T_\bullet} = \frac{\text{"Occurrence"}}{\text{"Exposure"}}$$

Inference in STK4080

For some estimator we may use maximum likelihood (ML) theory in this course.

An advantages of ML is general large sample results in finite parameter models.

However, we will work with estimators of functions: Survival function $S(t)$, cumulative hazard functions $A(t)$ for simple samples and cumulative baseline function $A_0(t) = \int_0^t \alpha_0(s)ds$ in proportional hazards models.

To deal with estimators of functions it will be convenient to introduce the counting process and martingale framework (Aalen, Borgan & Gjessing, 2008).
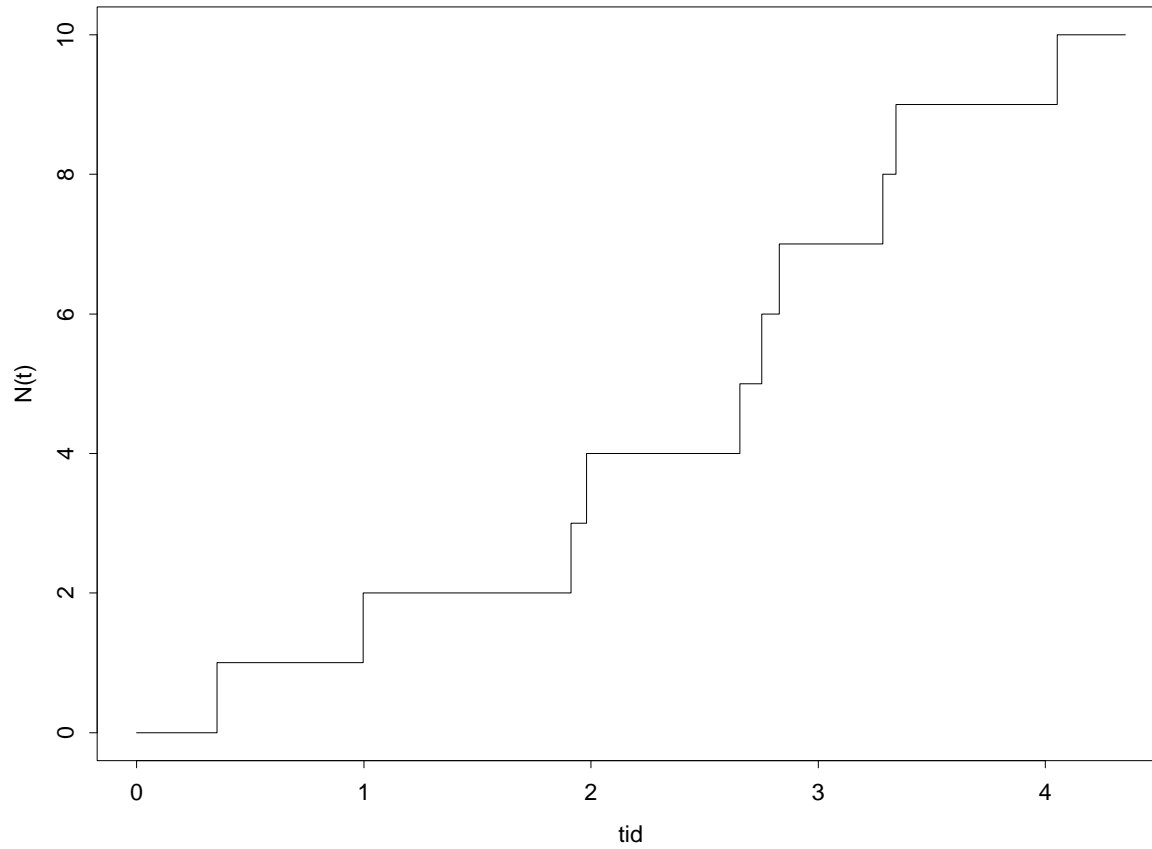
A counting process is a process

- that counts events over time

- has steps of size 1

- at random times

Examples

- Poisson process (STK2130)

- $N_i(t) = I(\widetilde{T}_i \leq t, D_i = 1)$

- $N(t) = \sum_{i=1}^{n} N_i(t)$

Example: Counting process

Nelson-Aalen / Kaplan-Meier with counting processes

We write the Nelson-Aalen estimator of the cumulative hazard $A(t)$ as

$$\hat{A}(t) = \sum_{\widetilde{T}_i \leq t} \frac{D_i}{Y(\widetilde{T}_i)} = \int_0^t \frac{dN(s)}{Y(s)},$$

that is as an integral with respect to the counting process.

Similarly we will write the Kaplan-Meier estimator of the survival function as

$$\hat{S}(t) = \prod_{\widetilde{T}_i \leq t} [1 - \frac{D_i}{Y(\widetilde{T}_i)}] = \prod_{s \leq t} [1 - \frac{dN(s)}{Y(s)}],$$

i.e. a so-called *product integral* with respect to the counting process.

A martingale is a process $M(t)$ with properties

- $\mathrm{E}[M(t)|\mathcal{F}_s] = M(s)$ for $s < t$

where $\mathcal{F}_s$ is the **history** of the process up to time s.

Examples:

- $M(t) = N(t) - \lambda t$ for Poisson process $N(t)$ with rate λ

- $M(t) = N(t) - \int_0^t Y(s)\alpha(s)ds$
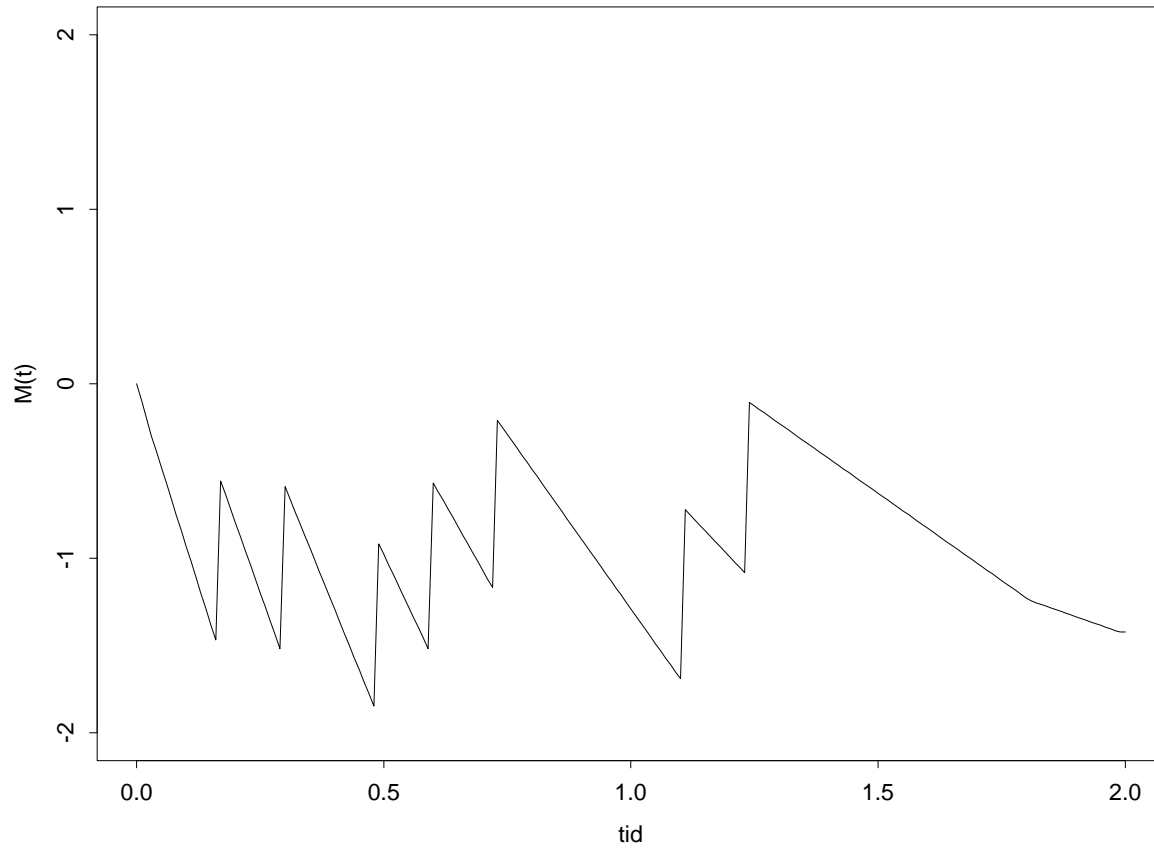 for $N(t) = \sum_{i=1}^n N_i(t)$

- For $Y(t) > 0$ we have the difference between the
  Nelson-Aalen and the cumulative hazard becomes
  $\hat{A}(t) - A(t) = \int_0^t \frac{1}{Y(s)}[dN(s) - Y(s)\alpha(s)ds] = \int_0^t \frac{dM(s)}{Y(s)}$
 which is a martingale!

Example: Martingale

Left truncation - delayed entry data

Previously in these slides we presented the following plot in connection with time on a clinical study when patient arrive over calendar time.

The plot may also be used to exemplify left truncated data where individuals are included in study from a certain age and where age is the underlying time scale.

Observations
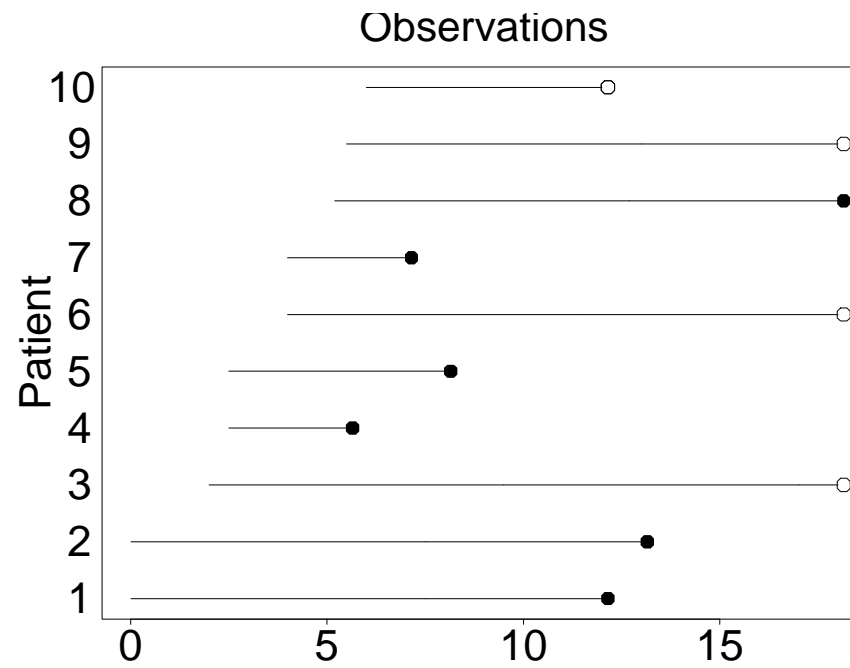
Left truncation and counting process framework

For such data we assume that

- time t is age and hazard is $\alpha(t)$

- individual enter at time $V_i = $ left truncation time

- individuals have events or are censored $D_i = 1/0$ at time $\tilde{T}_i$

Such data can be represented through counting process notation

- $N_i(t) = I(\tilde{T}_i \leq t) = $ counting events

- $Y_i(t) = I(V_i < t \leq \tilde{T}_i) = $ at risk indicator function

Nelson-Aalen and Kaplan-Meier estimators can then be calculated on left truncated data using the formulas given a few pages back.

Also Cox-regression allows for left truncation.

Event history analysis, Sec. 1.5 and 3.4

So far we have only considered simple survival analysis situations. There are, however, more complicated observational scheme with several types of events that can occur:

Examples:

- Competing risk situations
- Illness-death situation

Methods for survival data may be modified to handle such situations.
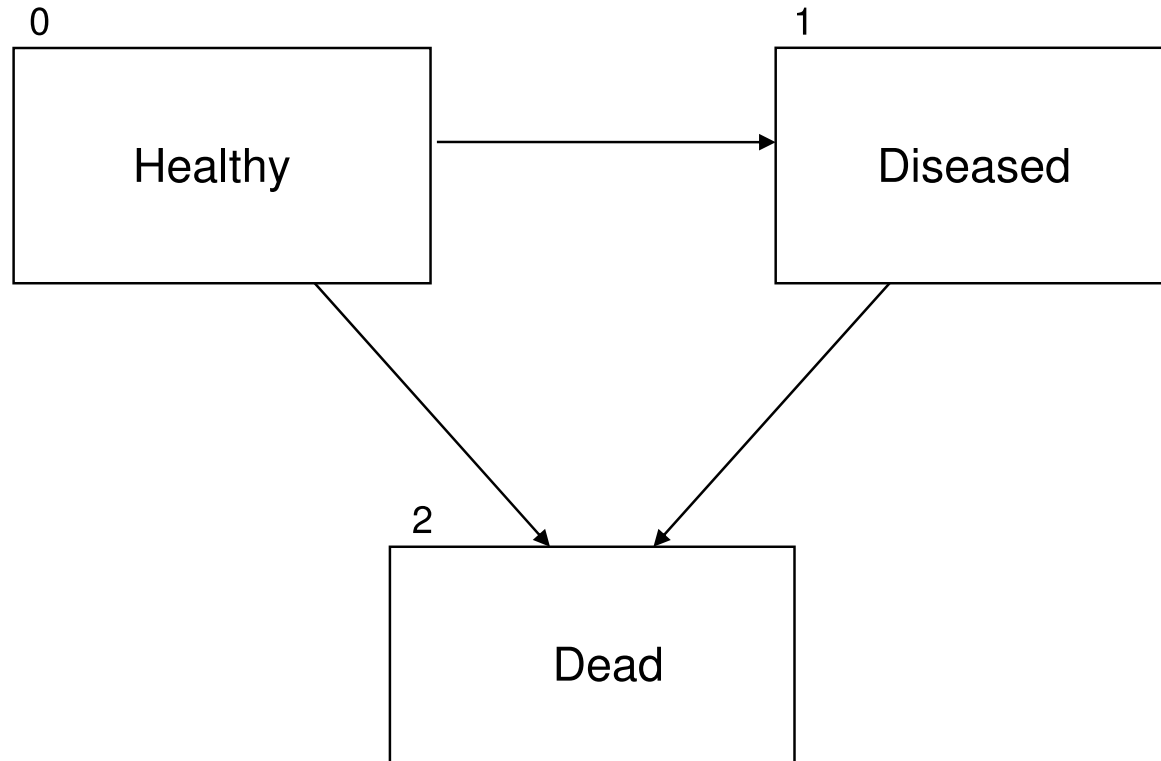
Competing risks, Sec. 3.4.1
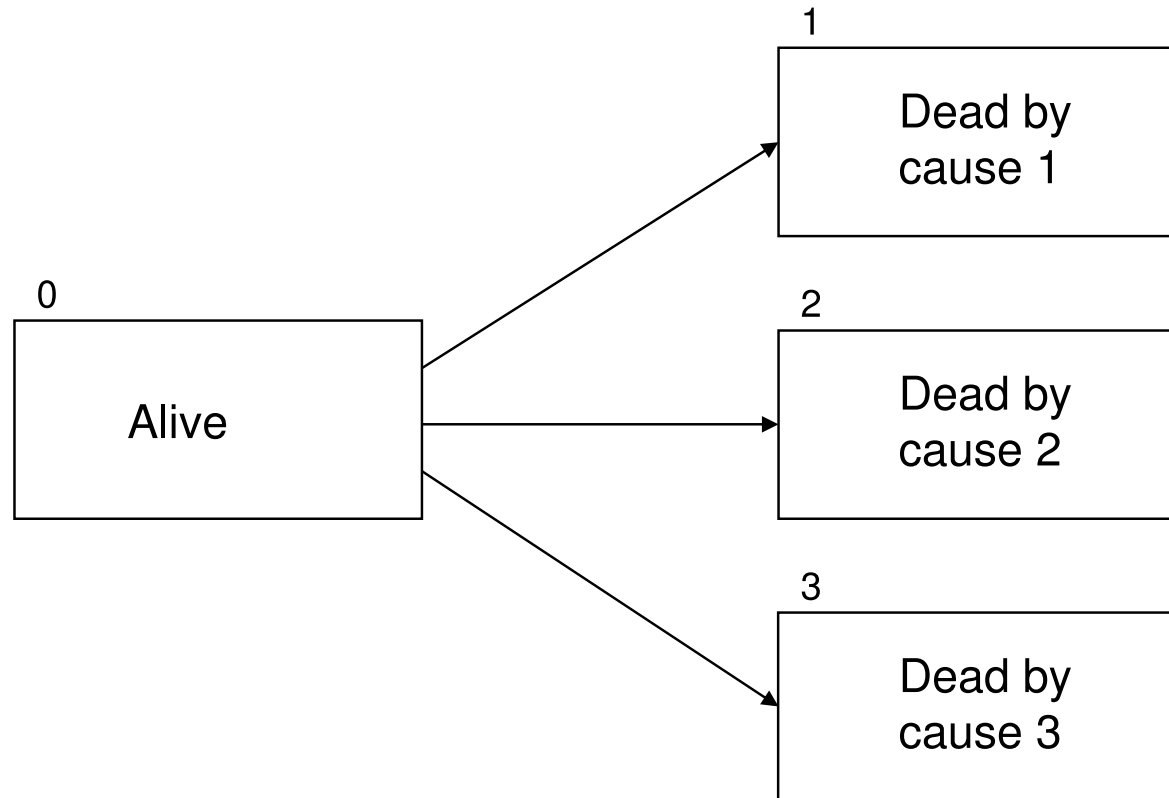
With $X(t) = j, j = 0, 1, 2, 3$ denoting the state of an individual at time t we will estimate the *cumulative incidence function*

$$\mathrm{P}_{0j}(t) = \mathrm{P}(X(t) = j | X(0) = 0),$$

i.e. the probability that an individual died of cause j before (or at) time t.
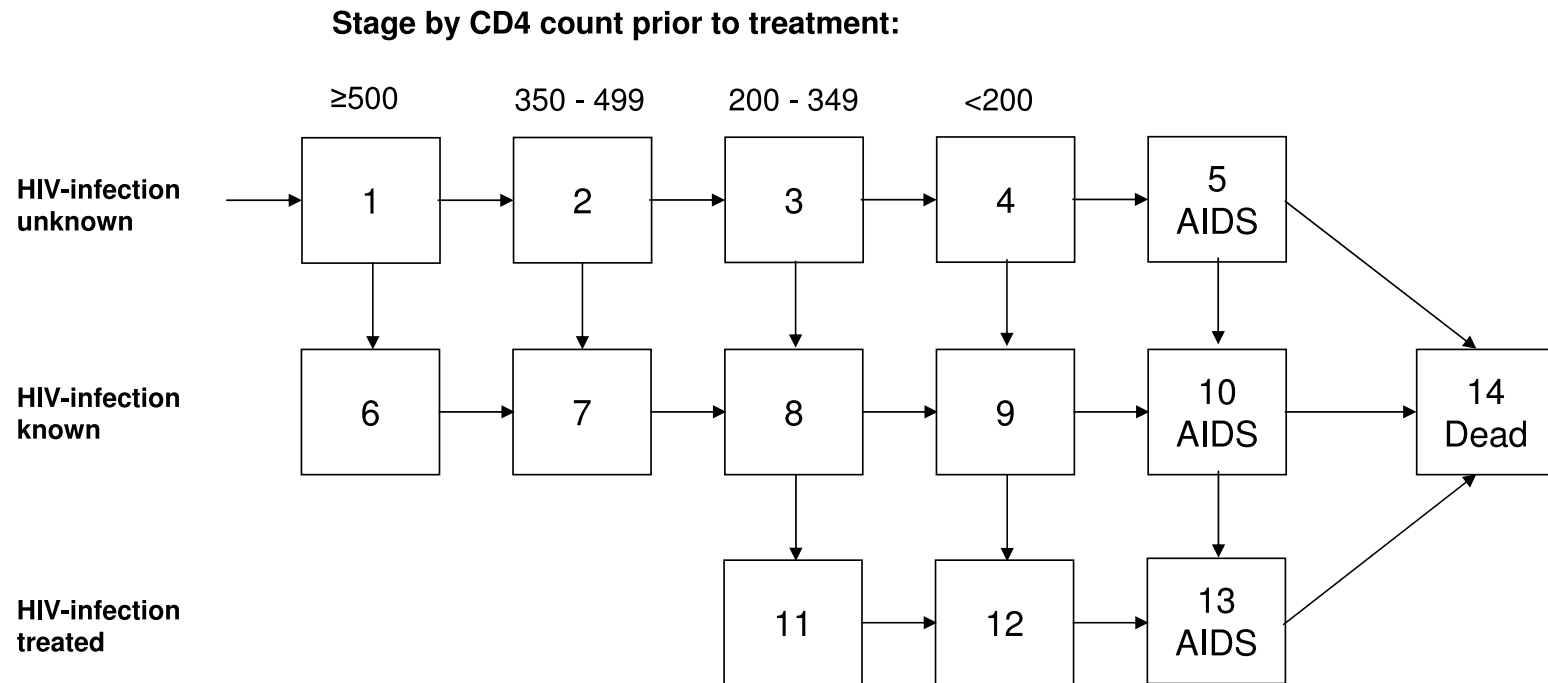
Healthy-Illness-Death, Sec. 3.4.2

0

Healthy

1

Diseased

2

Dead

Again denoting $X(t) = j$ the state of an individual at time t we are interested in estimating

$$\mathbf{P}_{jk}(t) = \mathbf{P}(X(t) = k | X(0) = j),$$

particularly for $j = 0$.

The general case, Sec. 3.4.3, Aalen-Johansen estimator

Stage by CD4 count prior to treatment:

| ≥500 | 350 - 499 | 200 - 349 | <200 |

Again, interest is on

$$\mathbf{P}_{jk}(t) = \mathbf{P}(X(t) = k | X(0) = j),$$

Frailty concept, Ch. 6

Individual hazards may depend on latent - nonobserved - variables Z referred to as *frailties* through

$$\alpha(t|Z) = Z\alpha(t)$$

where $\alpha(t)$ is some given hazard.

Observed hazards over populations without taking the frailty into account may appear strikingly different from $\alpha(t)$.

Example: Divorce rates over the population appears to have a maximum at 7 years ("7 year itch"). Under a frailty specification we can have e.g. increasing rates for each couple, the but frail marriages fail early - leaving only stable marriages after 7 years.

Curriculum will only consider Section 6.1 and 6.2

Multivatiate survival, Frailty Ch. 7 and Marginal Ch. 8

Example twin data: Hazards for twin j in pair i could be given

$$\alpha_{ij}(t) = Z_i \alpha_{ij}(t)$$

where Z_i is the frailty of the twin pair, $\alpha_{ij} = \exp(\beta' x_{ij})\alpha(t)$ and x_{ij} covariates of twin j in pair i.

Purpose:

- Estimate relative risks $\exp(\beta)$

- Investigate degree of dependence (i.e. frailty distribtion)

Curriculum will only consider Sections 7.1, 7.2, 8.1 and 8.2