Introduction to survival analysis

Per Kragh Andersen

Section of Biostatistics, University of Copenhagen

DSBS Course
Survival Analysis in Clinical Trials
2025

Overview

- Survival data: Definitions and examples
- Multi-state models: intensities and marginal parameters
- Non-parametric estimation of intensities
- Non-parametric tests
- Exercises - I
- Plug-in estimation of marginal parameters
- Exercises - II
- Parametric models
- Delayed entry

Textbook

'Models for Multi-State Survival Data: Rates, Risks, and Pseudo-Values'
PKA & HR (2023), Chapman & Hall/CRC.

Companion web pages:

https://multi-state-book.github.io/companion

include data sets for download and code for examples in both R and SAS.

Survival data: definitions and examples

Situations leading to survival data

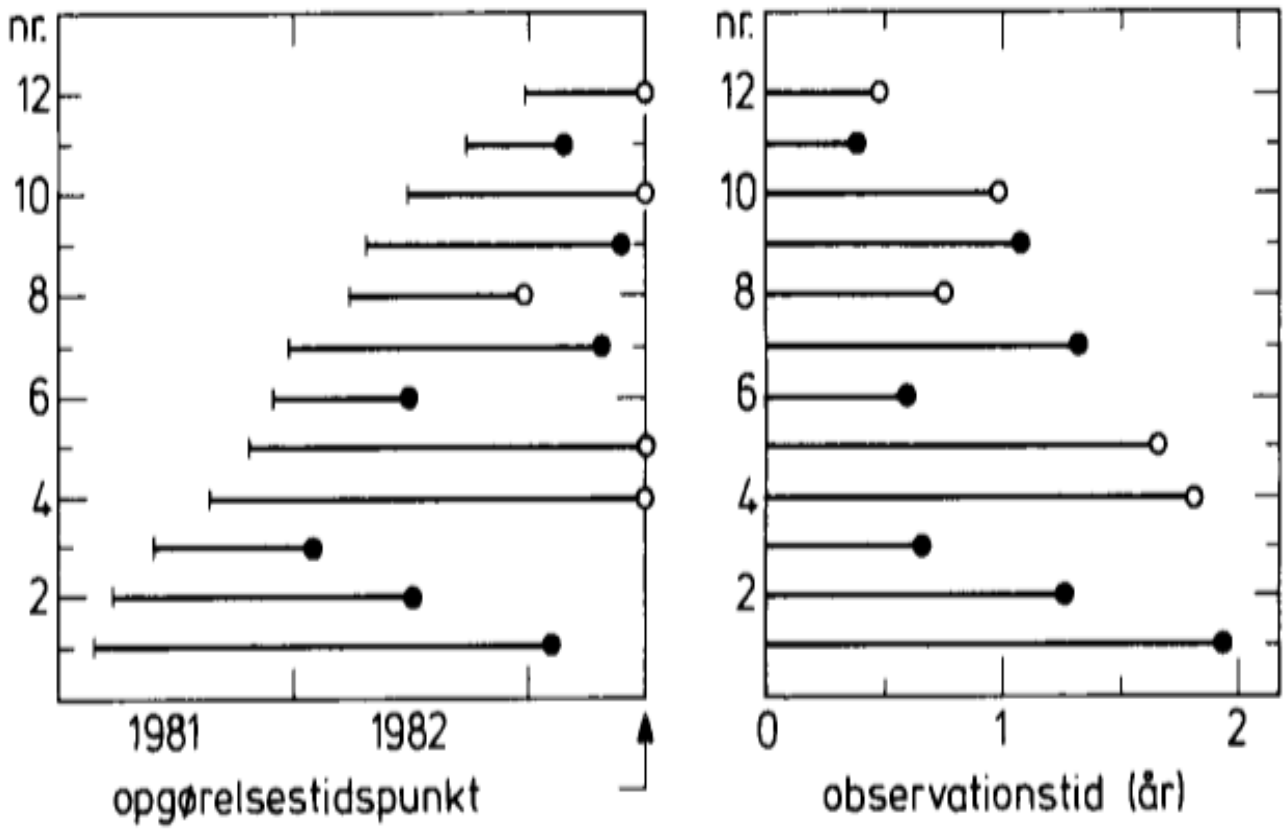
Time to *death* or other *event* of interest from a well-defined *time origin*:

- Time from start of randomized clinical trial to death
- or ... to some *composite end-point*
- Time from randomization to occurrence of side effect
- Time from birth to death
- Time from birth to first marriage
- Time from first employment to pension
- Time from filling a cavity in a tooth to filling falls out

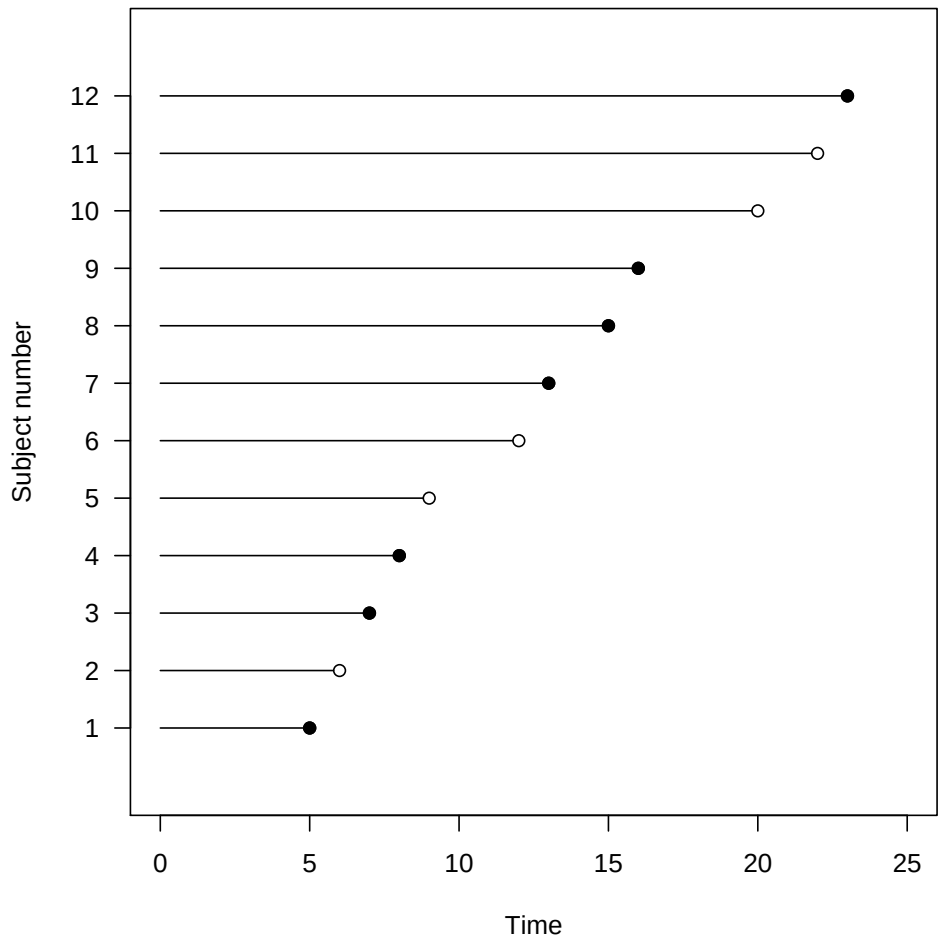
What is special about survival data?

- *(Right)-censoring*: For some subjects the event is not observed and we will only know an interval in which it did not occur.

A small data set



A small data set



A small data set

Ordered times: 5, 6*, 7 , 8, 9*, 12*, 13, 15, 16, 20*, 22*, 23,
* indicates censored survival times.
How to estimate the mean survival time?

$$\frac{5+6+7+8+9+12+13+15+16+20+22+23}{12} = \frac{156}{12} = 13.0?$$

$$\frac{5+7+8+13+15+16+23}{7} = \frac{87}{7} = 12.4?$$

Which fraction of patients survives past 12 months?

$$\frac{6}{12} = 0.5?$$

We need inference methods that are able to account for censoring.
This leads to a focus on other parameters than the mean value.

Survival and hazard functions

Let T be the *time* to the event of interest:

$$
\begin{aligned}
S(t) &= P(T > t) \\
&= \text{probability of survival beyond time } t \\
&= 1 - F(t), \quad F(t) \text{ is the failure } \textit{risk} \text{ before time } t. \\
\alpha(t) &= \textit{rate} \text{ or hazard function} \\
&= -\frac{d}{dt} \log(S(t)) \\
&= \frac{dF(t)}{S(t)}, \text{ i.e.,} \\
\alpha(t)dt &\approx P(T \leq t + dt \mid T > t) \\
&= \text{probability of failure before } t + dt \text{ given survival beyond } t.
\end{aligned}
$$

Math

Relationship between survival and hazard functions:

$$S(t) = \exp\left(-\int_0^t \alpha(s)ds\right) = \exp(-A(t));$$

$A(t)$ is the *integrated* or *cumulative* hazard function.

Note that this relationship between 'rate' ($\alpha(t)$) and 'risk' ($F(t) = 1 - S(t)$) requires that

- there are *no competing risks* (much more later),
- the distribution is absolutely continuous

Interpretation

The survival and distribution functions $S(t)$ and $F(t)$ are simple cumulative fractions of patients having survived until, or having failed by time t.

The hazard function $\alpha(t)$ describes the *instantaneous risk per time unit* of failing 'now' given alive.

The integrated hazard function:

- has a derivative ('slope') that is the hazard function
- is the expected number of 'renewals' before time t in a certain (strange?) experiment.

(When, in Part II of the course, we discuss recurrent events this experiment is much more natural.)

Other parameters

The *mean survival time* is:

$$E(T) = \varepsilon_0(\infty) = \int_0^\infty S(t)dt.$$

This depends critically on the right-hand tail of the distribution of T which we typically do not see because of censoring.
Instead, the τ-*restricted mean survival time* (RMST):

$$E(T \wedge \tau) = \varepsilon_0(\tau) = \int_0^\tau S(t)dt$$

may be studied. The interpretation is less nice (average time lived before time τ) and its value depends on the choice of τ.

Median: $\inf_t\{S(t) \leq 0.5\}$ (and other quantiles, that is, $1 - p$ instead of 0.5).

Math

Proofs by partial integration or, more simply:

$$T = \int_0^T 1 dt = \int_0^\infty I(T > t) dt$$

and take expectations ($E\big(I(T > t)\big) = S(t)$).

Similarly,

$$T \wedge \tau = \int_0^{T \wedge \tau} 1 dt = \int_0^\tau I(T > t) dt.$$

Population and sample

We are used to considering our data as a *sample* from some *population*, and the parameters refer to this population.

That is no different in survival analysis, however, it is important to realize that the target population is a *complete* population, i.e., *without censoring*.

Our ambition in survival analysis is therefore to draw inference on parameters like the survival function $S(t)$ or the hazard function $\alpha(t)$ from a potentially completely observed population based on incomplete (censored) data.

This is quite ambitious and requires certain assumptions.

Target population; censoring

For this ambition to be feasible:

1. the complete population should be well-defined
2. censoring should not leave us with a biased sample

Requirement 1 basically tells that the event under study should happen for every one in the population.

This means that we need to distinguish between situations where there are *no competing risks* and where there *are* competing risks (much more later).

Thus, censoring is an *avoidable* event while competing risks are *non-avoidable*.

Independent censoring

Requirement 2 is the assumption of *independent censoring* (by some denoted *non-informative* censoring).

This means that individuals censored at any given time t should not be a biased sample of those who are *at risk* at time t.

Stated in other words: the hazard function $\alpha(t)$ gives the event rate at time t, i.e., the failure rate given that the subject is still alive ($T > t$).

Independent censoring means that the extra information that the subject is not only alive, but also uncensored at time t does not change the failure rate.

Independent censoring

Typically, independent censoring cannot be tested from the available data - it is a matter of discussion.

Censoring caused by being alive at the end of study (so-called 'administrative censoring') can usually safely be taken to be 'independent'. One should be more suspicious to other kinds of loss to follow-up before end of study.

It is strongly advisable always to keep track of subjects who are lost to follow-up and to note the reasons for loss to follow-up (e.g., drop-out of follow-up schedule or emigration).

The above discussion of independent censoring should be thought of as 'for given covariates'. This means that censoring may depend on covariates as long as these covariates are accounted for in the hazard model (e.g., using the Cox regression model).

The PBC-3 trial in liver cirrhosis, Ex. 1.1.1

Lombard et al. (1993, *Gastroenterology*)

- Multi-centre randomized trial in patients with primary biliary cirrhosis.
- Patients ($n = 349$) recruited 1 Jan, 1983 - 1 Jan, 1987 from six European hospitals and randomized to CyA (176) or placebo (173).
- Followed until death or liver transplantation (no longer than 31 Dec, 1989); CyA: 30 died, 14 were transplanted; placebo: 31 died, 15 were transplanted; 4 patients were lost to follow-up before 1989.
- Primary outcome variable: time to death, incompletely observed (right-censoring), due to: liver transplantation, loss to follow-up, alive 31 Dec, 1989
- In some analyses, the outcome is defined as 'time to failure of medical treatment', i.e., to the composite end-point of either death or liver transplantation

LEADER trial, Ex. 1.1.6

Marso et al. (2016, *NEJM*)

- 9340 patients with Type 2 diabetes and high cardiovascular risk,
- randomized to liraglutide or placebo:
- 410 sites in 32 countries.
- Primary outcome: composite end-point including death from cardiovarcular cause, non-fatal infarction, non-fatal stroke – so-called 'MACE'.
- Minimum planned follow-up: 42 (+1) months, maximum 60 (+1) months.
- 4668 received liraglutide with 608 events (13.0%),
- 4672 received placebo with 694 events (14.9%).

Multi-state models

Multi-state models

We will view survival data as a special case of a data from a *multi-state model*.

A multi-state process is a stochastic process, $V(t)$ with values in a finite set

$$\mathcal{S} = \{0, 1, \ldots, k\}$$

indicating the *state occupied* at time t.

Observing $V(t)$ over time corresponds to observation of *transitions* between states in $\mathcal{S}$. We denote an observed transition an *event* and the resulting data *multi-state survival data* or *event history data*.

A state $h \in \mathcal{S}$ is *absorbing* if no transitions out of h are possible. A non-absorbing state is *transient*.

Parameters in multi-state models

Marginal parameters:

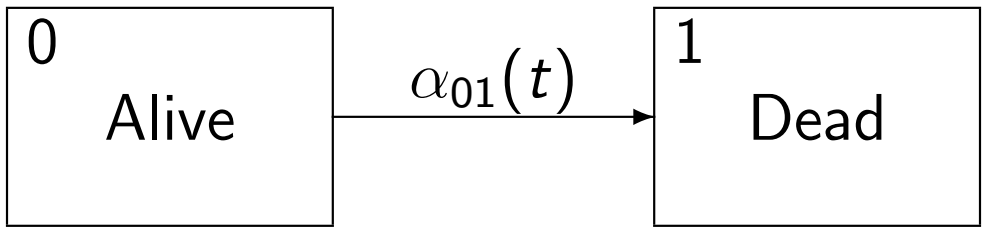
- $Q_h(t) = P(V(t) = h) = E(I(V(t) = h)), h \in \mathcal{S}$ state occupation probabilities
- $\varepsilon_h(\tau) = E(\int_0^\tau I(V(t) = h)dt) = \int_0^\tau Q_h(t)dt$ *expected length of stay* in state $h \in \mathcal{S}$ in $[0, \tau]$

Conditional parameters:

- $P_{hj}(s, t) = P(V(t) = j \mid V(s) = h, \text{past information in } [0, s))$ transition probabilities
- $\alpha_{hj}(t) = \lim_{dt \to 0} P_{hj}(t, t + dt)/dt$ transition intensities

The transition intensities (hazards, rates) are the basic building blocks for multi-state models, but marginal parameters often have more direct interpretations.

Two-state model for survival data

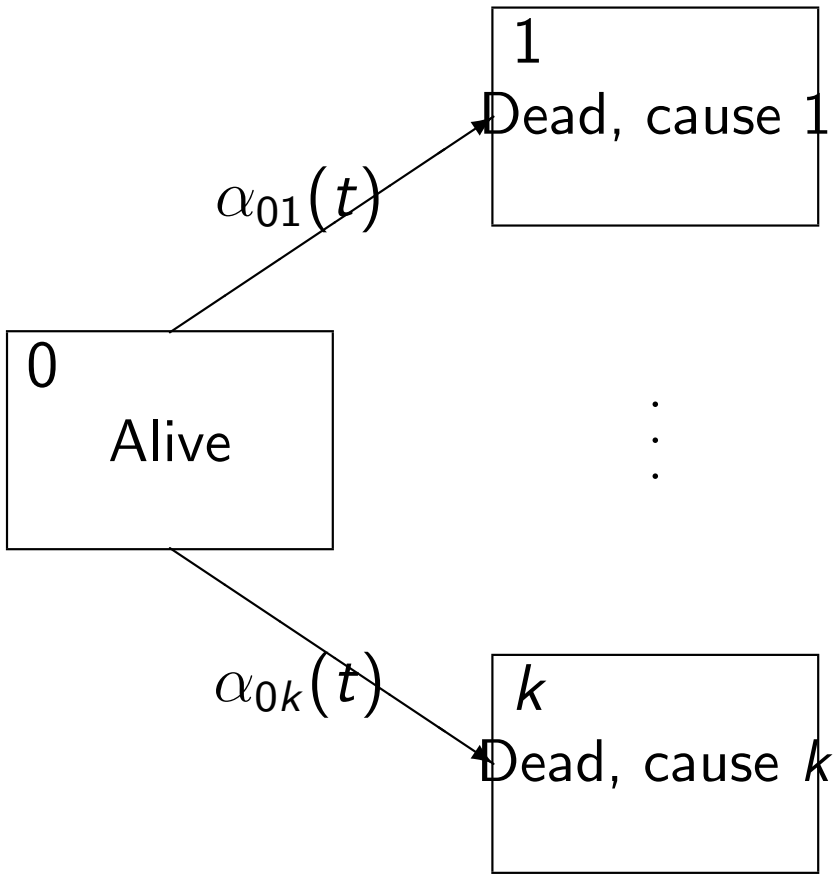


Transition intensity: *hazard function*
$\alpha(t) = \alpha_{01}(t) \approx P(\text{state 1 time } t + dt \mid \text{state 0 time } t)/dt$.
State occupation probabilities: *survival function*,
$Q_0(t) = S(t) = P(\text{state 0 time } t) = P(T > t)$,
and *cumulative probability of death before time t*, Eq. (1.2):
$Q_1(t) = 1 - Q_0(t) = F(t) = P(\text{state 1 time } t) =$
$1 - \exp(-\int_0^t \alpha_{01}(u)du)$.

Examples of survival data

- PBC-3 trial: composite end-point
- LEADER trial: first (composite) event, i.e., MACE or non-CV death

Competing risks model



Competing risks model

Transition intensities: *cause-specific hazards* $h = 1, \ldots, k$:

$\alpha_h(t) = \alpha_{0h}(t) \approx P(\text{state } h \text{ time } t + dt \mid \text{state 0 time } t)/dt.$

State occupation probabilities: *overall survival function*:

$Q_0(t) = S(t) = P(\text{alive time } t),$

$= \exp(-\int_0^t (\alpha_{01}(u) + \cdots + \alpha_{0k}(u))du)$

and *cumulative incidences* $h = 1, \ldots, k$, Eq. (1.3):

$Q_h(t) = F_h(t) = P(\text{dead from cause } h \text{ before time } t)$

$= \int_0^t S(u)\alpha_{0h}(u)du.$

For two groups, (e.g., treatment groups) $0, 1$, $Q_h^{(1)}(t_0) - Q_h^{(0)}(t_0)$ is the cause h *risk difference* at time t_0 (similarly for the two-state model and for the *risk ratio*).

Examples of competing risks data

- PBC-3 trial: death without liver transplantation and liver transplantation
- LEADER trial: first MACE and non-CV-death

Expected length of stay

In both models, the τ-restricted mean survival time (RMST), $\varepsilon_0(\tau) = E(T \wedge \tau)$, Eq. (1.10), is:

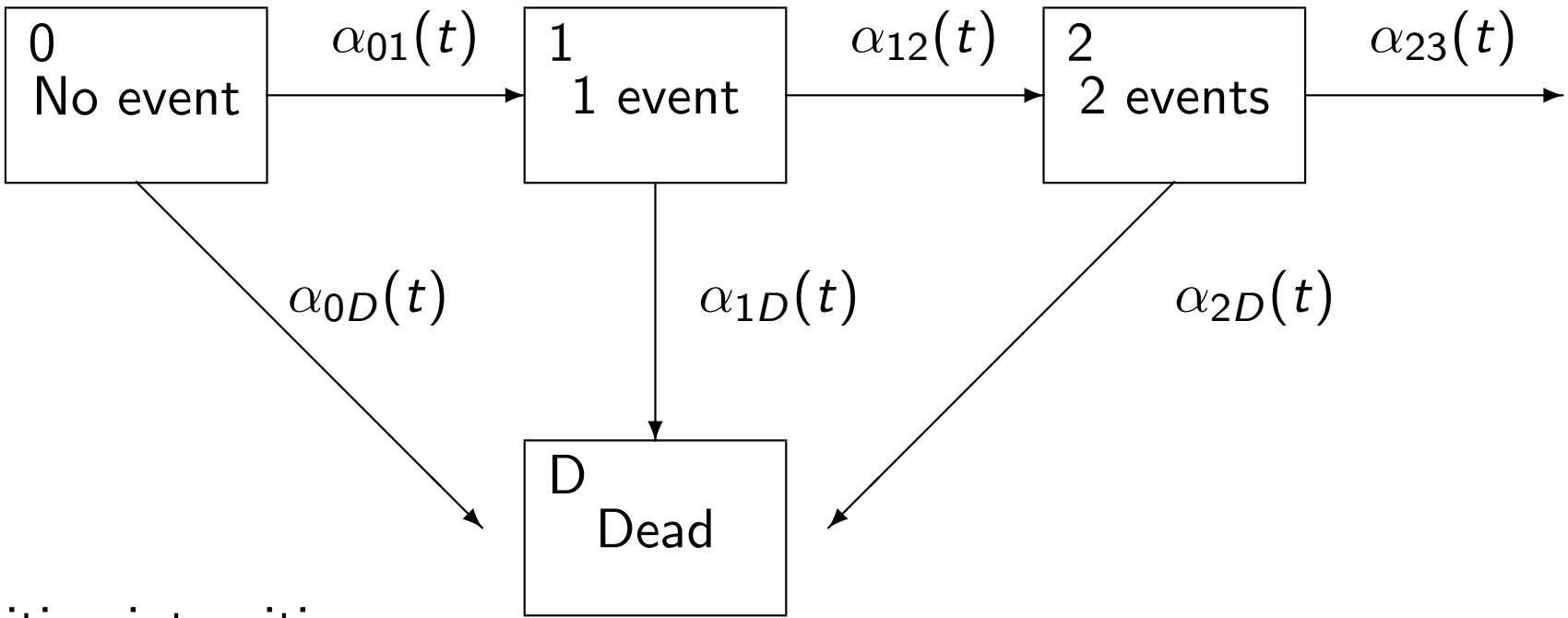
$$\varepsilon_0(\tau) = \int_0^\tau S(t)dt.$$

In the two-state survival model: $\varepsilon_1(\tau) = \int_0^\tau Q_1(t)dt$ is the expected time lost before time τ, i.e., $\tau - E(T \wedge \tau)$.
In the competing risks model,

$$\varepsilon_h(\tau) = \int_0^\tau Q_h(t)dt$$

is the expected time lost 'due to cause h' before time τ.

Recurrent events (with competing risks)



Transition intensities:

$\alpha_{hj}(t) \approx P(\text{state } j \text{ time } t + dt \mid \text{ state } h \text{ time } t, \text{ past at } t-)/dt.$

Recurrent events: marginal parameters

The most important marginal parameter is

$$\mu(t) = E(N(t)) = \int_0^t S(u)\alpha^*(u)du,$$

with $N(t)$ = number of events in $[0, t]$, $S(t) = P(T > t)$, and $\alpha^*(\cdot)$ the *marginal rate function given survival*

$$\alpha^*(t) \approx E(dN(t) \mid T > t)/dt$$

(T is the survival time, i.e., time of entry into state D). In the model without the final death state, D, $\mu(t) = \int_0^t \alpha^*(u)du$ with $\alpha^*(\cdot)$ now being the marginal rate function $\alpha^*(t) \approx E(dN(t))/dt$.

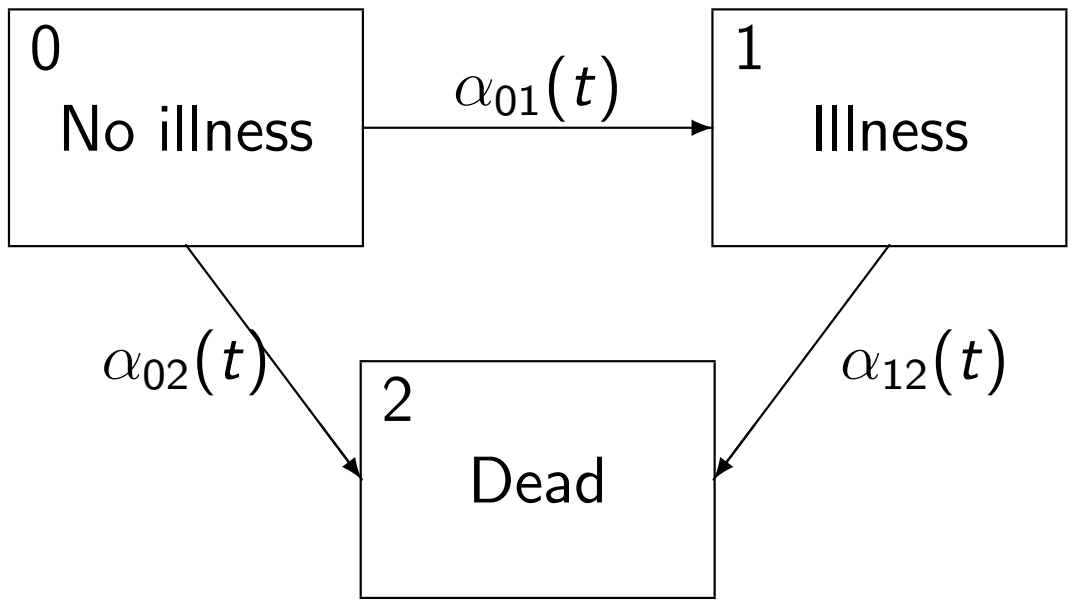
Math

$$
\begin{aligned}
E(N(t)) &= E\left(\int_0^t dN(u)\right) \\
&= \int_0^t E(dN(u)) \\
&= \int_0^t P(dN(u) = 1) \\
&= \int_0^t P(dN(u) = 1 \mid T > u) P(T > u)
\end{aligned}
$$

Examples of recurrent events data

- LEADER trial: recurrent MACE (competing event: non-CV death).
- LEADER trial: recurrent myocardial infarctions (competing event: all-cause death).
- Ex. 1.1.5: Recurrent episodes in affective disorder (competing event: all-cause death – more in Part II); Kessing, Hansen, Andersen, Angst (2004, *Acta Psych. Scand.*)

The irreversible illness-death model



Examples of illness-death model data

- PBC-3 trial: no event, liver transplantation, death with or without liver transplantation ('in principle' – information after liver transplantation is not available).
- LEADER trial: no event, first myocardial infarction, death with or without myocardial infarction.
- Example 1.1.4: The PROVA trial in liver cirrhosis (more in Part II)
- Example 1.1.7: Bone marrow transplantations (more in Part II)

Non-parametric estimation of cumulative intensities

Observations

Observation of

$$(V_i(t), t \in [0, \tau_i], i = 1, \ldots, n),$$

(where τ_i is either the time when $V_i(\cdot)$ reaches an *absorbing state*, or a time C_i of *right-censoring*) can be represented by *counting processes*:

$$N_{hji}(t) = \text{ number of direct } h \to j \text{ transitions } (h \neq j)$$

$$\text{observed in } [0, t] \text{ for subject } i = 1, \ldots, n,$$

and *at risk processes*

$$Y_{hi}(t) = \text{ indictor for } i \text{ being observed in state } h \text{ at time } t-$$

leading to the *intensity process* for $N_{hji}(t)$ being $\lambda_{hji}(t) = \alpha_{hji}(t) Y_{hi}(t)$ (under independent censoring).

The Nelson-Aalen estimator

If we assume that the intensity $\alpha_{hj}(t)$ is the same for all subjects and independent on the past (Markov assumption), then a natural estimator for

$$\alpha_{hj}(t)dt \approx P(V(t+dt) = j \mid V(t) = h)$$

is

$$\widehat{\alpha_{hj}(t)dt} = \frac{\sum_i dN_{hji}(t)}{\sum_i Y_{hi}(t)} = \frac{dN_{hj}(t)}{Y_h(t)},$$

leading to the *Nelson-Aalen* estimator, Eq. (3.10), for the cumulative intensity $A_{hj}(t) = \int_0^t \alpha_{hj}(u)du$:

$$\widehat{A}_{hj}(t) = \int_0^t \frac{\sum_i dN_{hji}(u)}{\sum_i Y_{hi}(u)} = \int_0^t \frac{dN_{hj}(u)}{Y_h(u)},$$

an increasing step function with steps at observed $h \to j$ transition times. The local slope estimates the intensity.

Likelihood

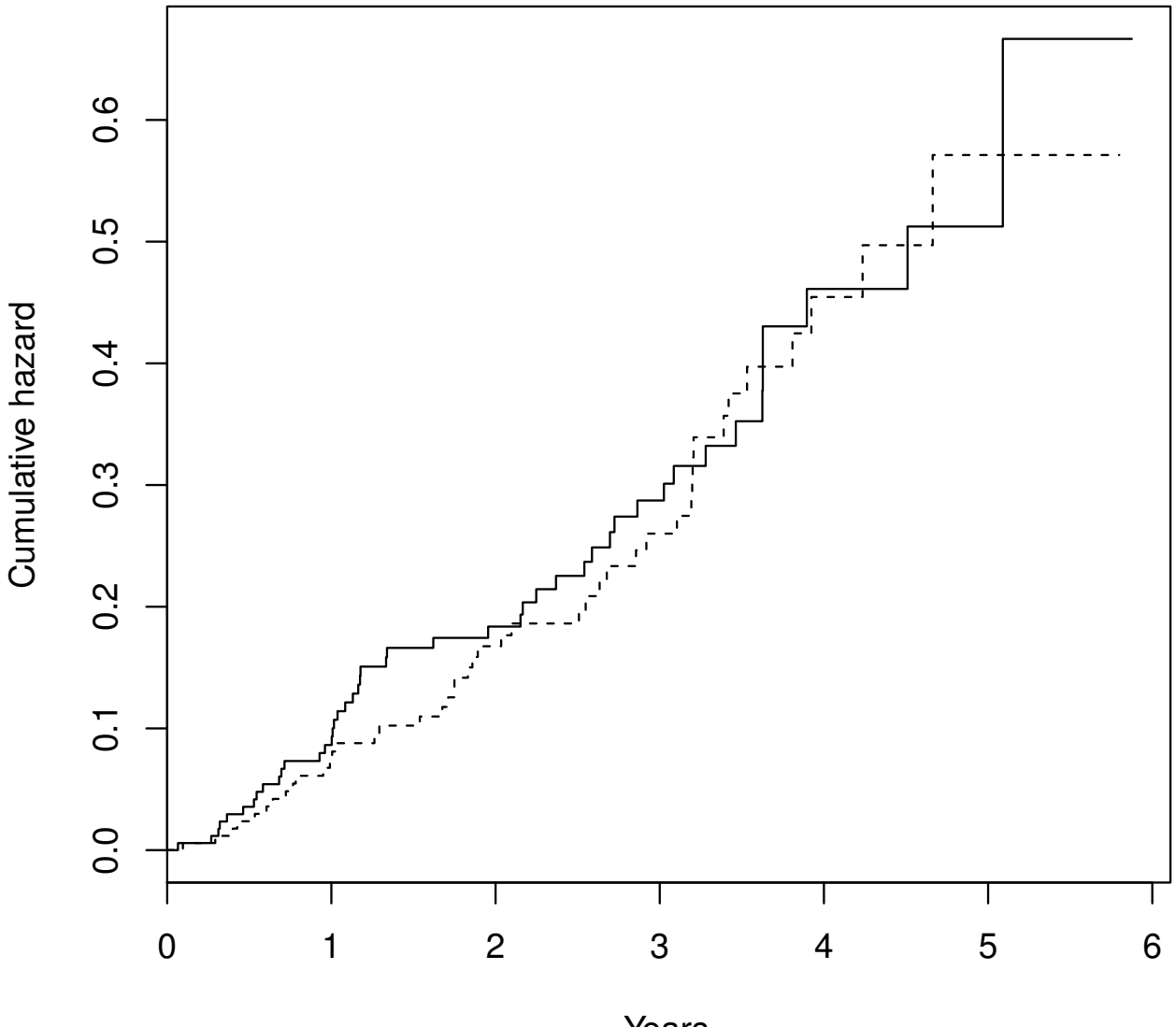
The Nelson-Aalen estimator has a maximum likelihood interpretation based on *the Jacod formula*: $L = \prod_i L_i$ with

$$L_i = \prod_t \prod_\nu \left(\lambda_{\nu i}(t)^{dN_{\nu i}(t)} \right) \times \exp\left(- \sum_\nu \int_0^\infty \lambda_{\nu i}(u) du \right). \quad (*)$$

Here, the event *types*, ν, correspond the transitions, $h \to j$, that are possible. For survival data, this is only a $0 \to 1$ transition corresponding to an observed time of death.

Treating 'jumps' $dA_\nu(t) = \alpha_\nu(t)dt$ as parameters, maximization of $(*)$ leads to the Nelson-Aalen estimator.

The PBC-3 trial, composite end-point



R code, cumulative hazards

```
library(survival)

pbcna <- survfit(Surv(days, status != 0) ~ tment,
         data = pbc3)

plot(pbcna,fun="cumhaz")

# Alternatively:

plot(pbcna,cumhaz=TRUE)
```

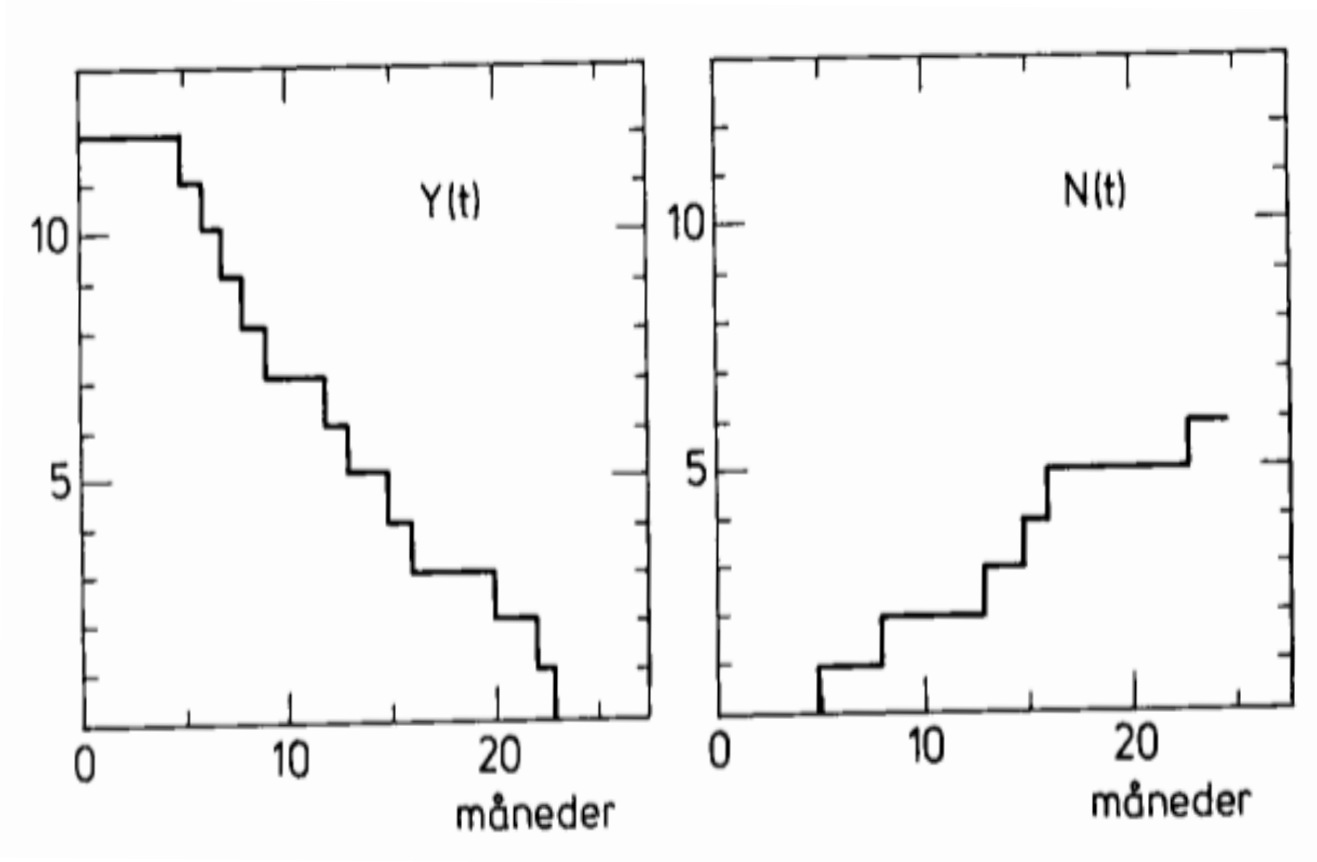
This code gives a crude version of the figure that may be improved upon by adding options to `plot` etc.

Math – survival data

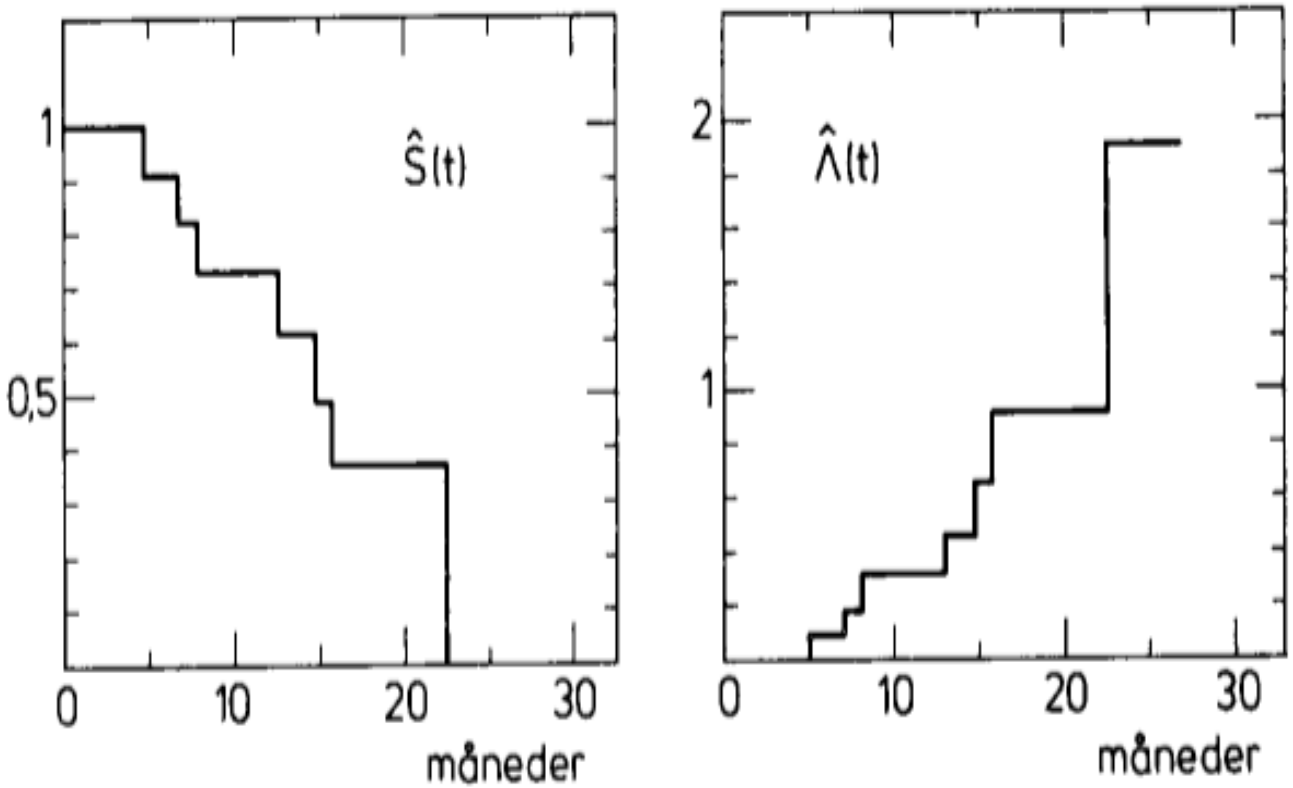
Properties of the estimators based on *counting processes*:

- $N(t) =$ number of observed failures in $[0, t]$, $Y(t)$ number observed to be at risk at time $t-$, the size of the *risk set* at time t.
- $N(t)$ has *intensity process* $\lambda(t) = Y(t)\alpha(t)$, i.e., $E(dN(t) \mid \text{past}) \approx Y(t)\alpha(t)dt$,
- By the *Doob-Meyer decomposition* $M(t) = N(t) - \int_0^t Y(u)\alpha(u)du$ is a *martingale*
- $\widehat{A}(t) = \int_0^t \frac{I(Y(u)>0)}{Y(u)} dN(u)$
- $\widehat{A}(t) - \int_0^t I(Y(u) > 0)\alpha(u)du = \int_0^t \frac{I(Y(u)>0)}{Y(u)} dM(u)$ is also a martingale
- From this, (approximate) unbiasedness, consistency, asymptotic normality and variance formula follow: $\text{SD}(\widehat{A}(t)) = \sqrt{\int_0^t dN(u)/(Y(u))^2}$.

A small data set: $Y(t)$ and $N(t)$



A small data set: $\widehat{A}(t)$ (and $\widehat{S}(t)$ – more later)



Competing risks

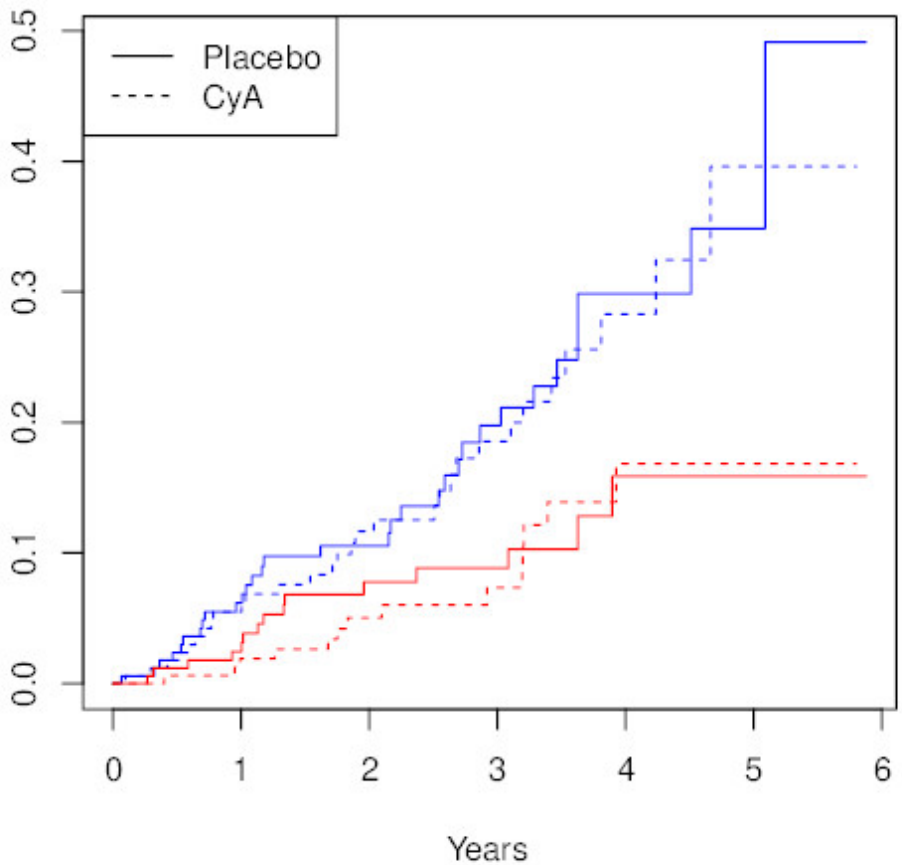
For the competing risks model, the possible transitions are $\nu = 0 \to h, \quad h = 1, \ldots, k$.

Since the Jacod formula (*) *factorizes* over transition types (i.e., causes of death), estimation of (cumulative) cause-specific hazards $A_h(t) = \int_0^t \alpha_h(u)du$ can be performed *one cause at a time*, formally, treating other causes in the same way as censored observations.

This is utilized in the `R` code, e.g., for death without transplantation in the PBC-3 study, the `survfit` object is:

```
survfit(Surv(days, status == 2) ~ tment, data = pbc3)
```

PBC-3: cause-specific hazards



Transplantation (red), death without transplantation (blue)

Non-parametric testing

A general test statistic

We want to compare hazard functions $\alpha_1(t)$ and $\alpha_0(t)$ in two groups.

Counting process notation: In group j we have: $N_j(t) =$ number of observed events in $[0, t]$, $Y_j(t) =$ number at risk just before time t. Nelson-Aalen estimators for $A_j(t) = \int_0^t \alpha_j(u)du$:

$$\widehat{A}_j(t) = \int_0^t \frac{I(Y_j(u) > 0)}{Y_j(u)} dN_j(u), \quad j = 0, 1.$$

Idea in general test statistic: Look at K-weighted differences between increments in Nelson-Aalen estimators:

$$U(t) = \int_0^t K(u)\big(d\widehat{A}_1(u) - d\widehat{A}_0(u)\big).$$

The logrank test

Different choices of $K(\cdot)$ provide different tests with different properties.

The most common choice is

$$K(t) = \frac{Y_0(t)Y_1(t)}{Y_0(t) + Y_1(t)}$$

leading to

$$U(t) = N_1(t) - \int_0^t \frac{Y_1(u)}{Y_0(u) + Y_1(u)}(dN_0(u) + dN_1(u)).$$

Evaluated at $t = \infty$, we get the *logrank test*:

$$U(\infty) = \text{‘Observed’ - ‘Expected’ (in group 1).}$$

The logrank test

From each 2 by 2 table at a failure point, say X:

Group	Died	Survived	Alive before
0	$dN_0(X)$	$Y_0(X) - dN_0(X)$	$Y_0(X)$
1	$dN_1(X)$	$Y_1(X) - dN_1(X)$	$Y_1(X)$
	$dN_0(X) + dN_1(X)$		$Y_0(X) + Y_1(X)$

we add the observed $dN_1(X)$ and expected

$$\frac{Y_1(X)}{Y_0(X) + Y_1(X)}(dN_0(X) + dN_1(X))$$

numbers of failures from one group (here group 1).

The logrank test

The logrank test (as we shall see later) has optimality properties against *proportional hazards* alternatives:

$$\alpha_1(t) = \theta\alpha_0(t).$$

Using instead weights given by $K(t) = Y_0(t)Y_1(t)$, a test statistic is obtained where values of 'observed - expected' at earlier time points are given larger weight.
This test statistic is the two-sample Wilcoxon (Mann-Whitney) test when there are no censored observations.

For either choice of $K(\cdot)$, the statistic $(U(\infty))^2$, properly normalized, is referred to the χ^2_1-distribution.

The logrank test has developed into the test of choice, and any paper using a different test will be looked upon with suspicion.

Math

Doob-Meyer decomposition for each $j = 0, 1$ (M_j is a martingale):

$$N_j(t) = \int_0^t Y_j(u)\alpha_j(u)du + M_j(t).$$

$$U(t) = \int_0^t K(u)\big(dN_1(u)/Y_1(u) - dN_0(u)/Y_0(u)\big)$$

Using the decomposition under $H_0 : \alpha_1(t) = \alpha_0(t)$ we see that

$$U(t) = \int_0^t K(u)\big(dM_1(u)/Y_1(u) - dM_0(u)/Y_0(u)\big)$$

is a *martingale*, i.e., $E(U(t)) = 0$ and the asymptotic distribution (a normal distribution) together with the normalizing variance can be found by a martingale CLT.

The stratified logrank test

Comparison of two groups (say, $Z = 1$ and $Z = 0$) *after adjustment* for a categorical variable (Z_0) can be performed using the *stratified* logrank test.
Here, observed and expected numbers of failures (e.g., for $Z = 1$) are first computed within strata given by values of Z_0 and, subsequently, added across strata.

The PBC-3 study

		Placebo	CyA	Logrank
Total	*n*	173	176	
	OBS	46	44	
	EXP	44.7	45.3	0.08
Unit 1	*n*	11	12	
	OBS	1	3	
	EXP	2.23	1.77	
Unit 2	*n*	74	76	
	OBS	27	20	
	EXP	22.96	24.04	
Unit 3	*n*	22	24	
	OBS	3	6	
	EXP	4.35	4.65	
Unit 4	*n*	40	39	
	OBS	11	11	
	EXP	10.99	11.01	
Unit 5	*n*	12	11	
	OBS	2	3	
	EXP	1.90	3.10	
Unit 6	*n*	14	14	
	OBS	2	1	
	EXP	1.12	1.88	0.3

Doing it in R

```
library(survival)

survdiff(Surv(days,status!=0) ~ tment, data=pbc3)

survdiff(Surv(days,status!=0) ~ tment + strata(unit),

data=pbc3)
```

Exercises - I

Estimation of marginal parameters: plug-in

Two-state model

The intensities are the basic parameters in multi-state models and marginal parameters, such as state occupation probabilities $Q_h(t)$, may be estimated by *plug-in*.
For the two-state model, this leads to the Kaplan-Meier estimator, Eq. (4.3), for $S(t) = Q_0(t)$:

$$\widehat{S}(t) = \prod_{u \leq t} \left(1 - \frac{\sum_i dN_i(u)}{\sum_i Y_i(u)}\right).$$

Median: $\inf_t\{\widehat{S}(t) \leq 0.5\}$, time point at which the K-M estimator goes below 0.5 – not always estimable.

Note how censored observations are used for both K-M and N-Aa: a subject censored at X_j gives rise to *no jump* in the estimator but contributes to the size, $Y(t)$ of the risk set for $t \leq X_j$.

Math

Why not estimate $A(t)$ by $-\log(\widehat{S}(t))$ or $S(t)$ by $\exp(-\widehat{A}(t))$?
This is because the relation $S(t) = \exp(-A(t))$ holds for *absolutely continuous distributions* and our estimators are discrete distributions.
For discrete distributions, the relationship between cumulative hazard (measure) and survival function is given by the *product-integral*:

$$S(t) = \prod_{u<t} \big(1 - dA(u)\big)$$

and $\widehat{S}(t)$ is, indeed, the product-integral of $\widehat{A}(t)$.
Properties of the K-M estimator, including asymptotic normality and SD, follow from those of N-Aa via this relationship (the product-integral is a continuous and differentiable mapping).
In practice, it makes little difference using $\widehat{S}(t)$ or $\exp(-\widehat{A}(t))$.

Confidence limits

The standard error (SD) of the Kaplan-Meier estimator may be estimated by *Greenwood's formula*:

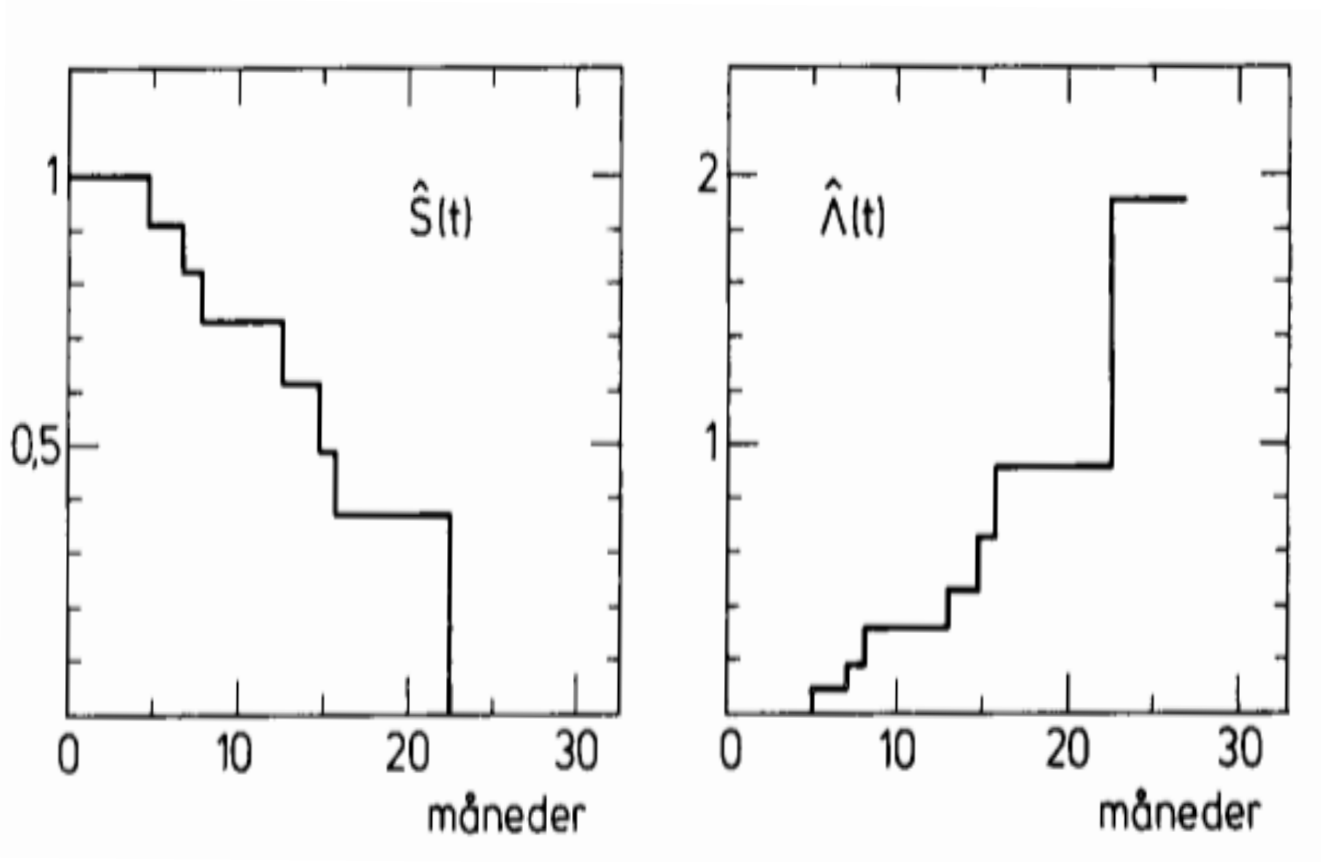
$$SD(\widehat{S}(t)) = \widehat{S}(t)\sqrt{\int_0^t \frac{dN(u)}{Y(u)(Y(u) - \Delta N(u) + 1)}}.$$

To get an approximate 95% confidence interval for $S(t)$, one may use simple linear limits $\widehat{S}(t) \pm 1.96 \cdot SD(\widehat{S}(t))$.
To eliminate problems with range restrictions when $S(t)$ is close to 0 or 1, transformations (i.e., using the delta-method) may be used, e.g., the $\log(-\log)$ transformation, which leads to the interval

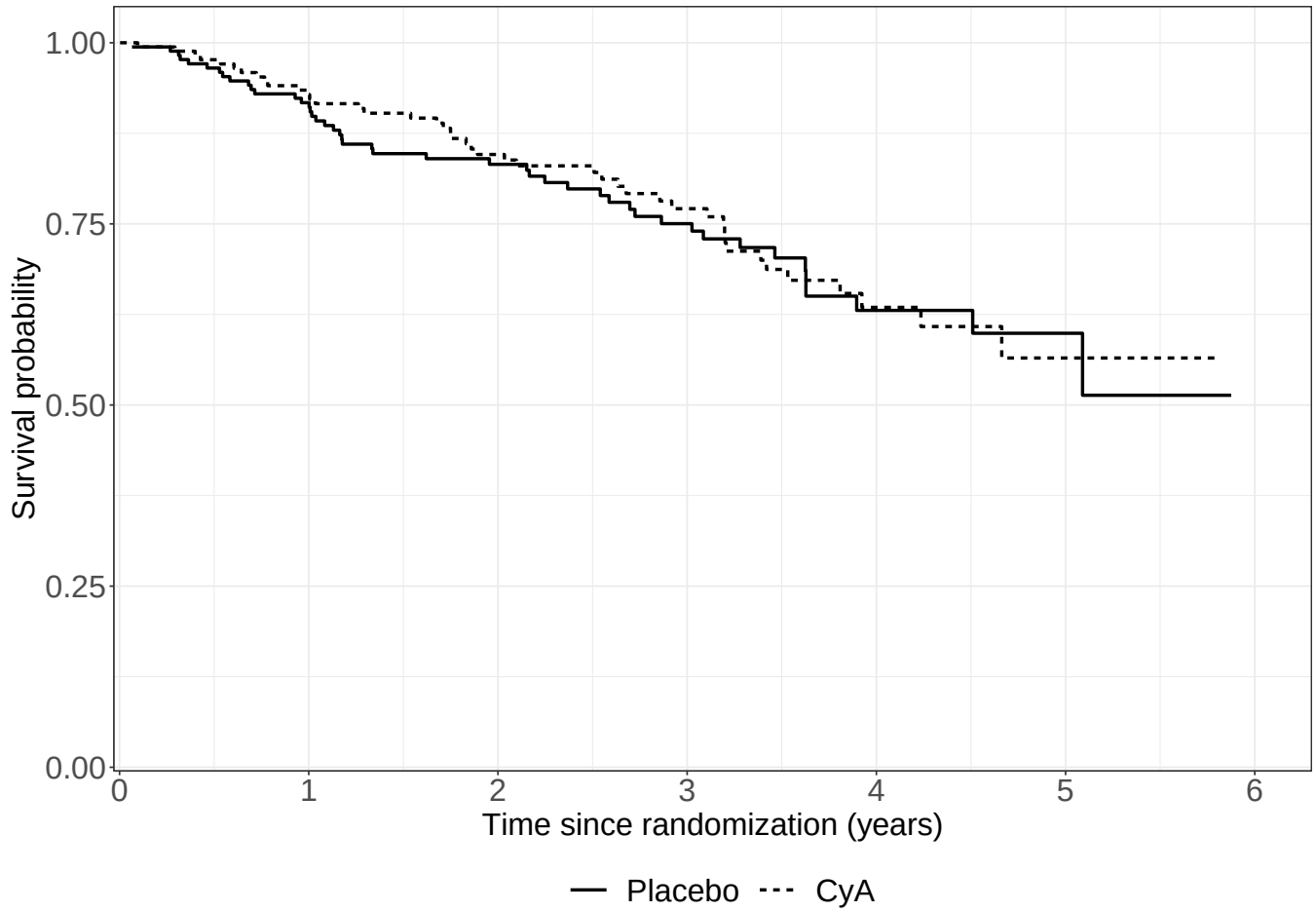
$$(\widehat{S}(t))^a \leq S(t) \leq (\widehat{S}(t))^b,$$

where $b = 1/a$ and $a = \exp(1.96 \cdot SD(\widehat{S}(t))/(-\log(\widehat{S}(t))))$.

A small data set: estimates



The PBC-3 trial, composite end-point



Estimation of marginal parameters: competing risks

The cause-h cumulative incidence is $F_h(t) = Q_h(t) = \int_0^t S(u)\alpha_{0h}(u)du$, and the plug-in estimator is the *Aalen-Johansen* estimator, Eq. (4.9):

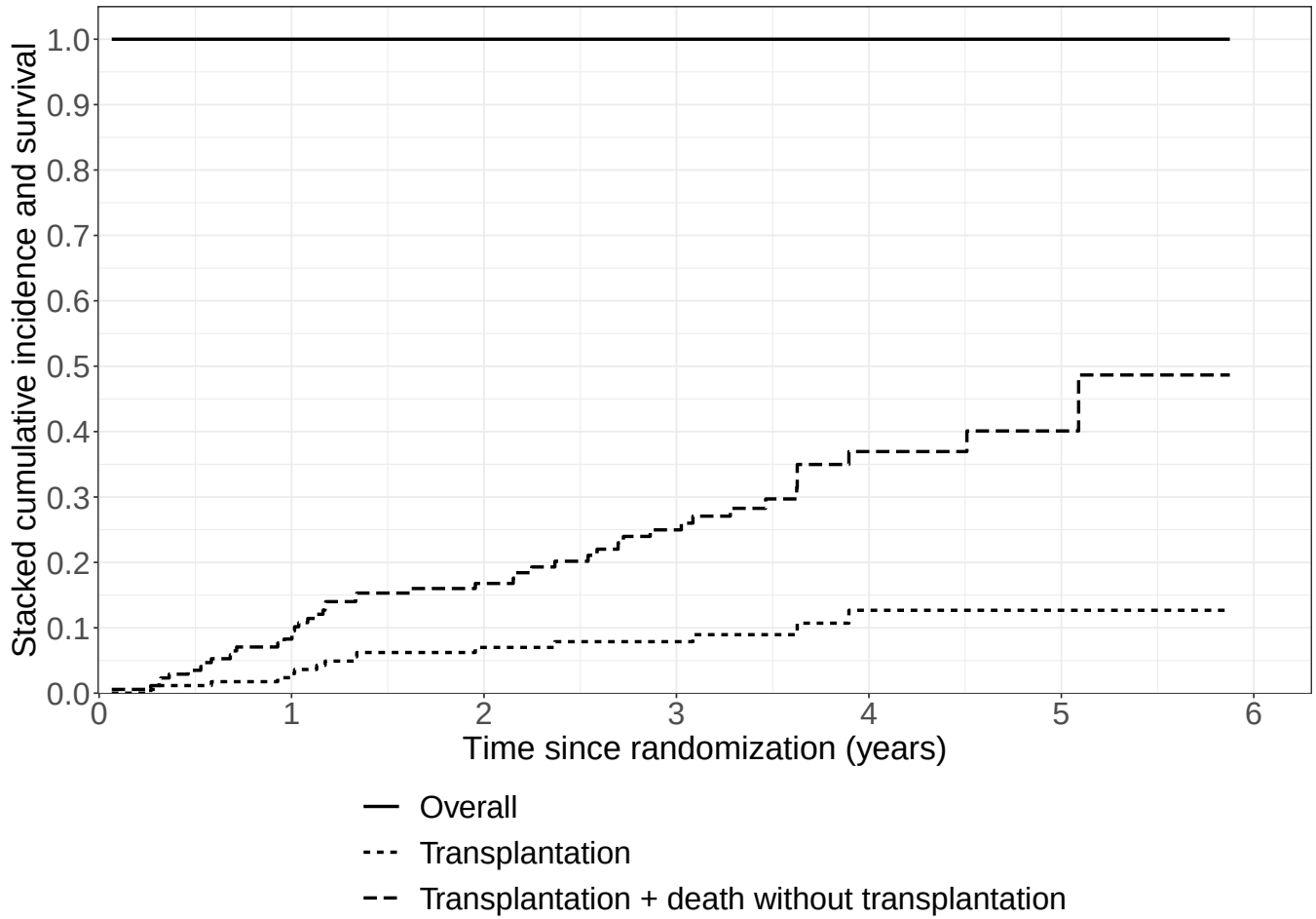
$$\widehat{Q}_h(t) = \int_0^t \widehat{S}(u-)d\widehat{A}_{0h}(u),$$

with $\widehat{S}$, the Kaplan-Meier estimator for overall survival (i.e., counting all deaths) and $\widehat{A}_{0h}$ the Nelson-Aalen estimator for the cumulative cause-h-specific hazard.

Asymptotic results, including an estimated SD, are available.

The quantity $Q_h(t_0)$ is the t_0-year *risk* of cause h from which t_0-year *risk differences* or *risk ratios* may be estimated. We will study inference for these contrasts later in the course using *pseudo-values*.

The PBC-3 trial, death and transplantation



Aalen-Johansen estimators (placebo) – correct!

Rates and risks

Recall that the risk for cause h depends on the rates for *all* causes – also for $j \neq h$.

For the two-state model for (overall) survival, hazard function α and failure function F contain *equivalent* information and one may be obtained from the other.

This one-to-one correspondence is lost for competing risks

All of the cause-specific hazards, $\alpha_1(t), \ldots, \alpha_k(t)$, are needed when computing each of the cumulative incidences, $F_h(t)$ (and vice versa).

Using the Kaplan-Meier estimator on a single cause

We have the relation:

$$F_h(t) = P(\text{dead from cause 1 before time } t) = \int_0^t S(u)\alpha_h(u)du.$$

If $\alpha_j(t) = 0, \quad j \neq h$, i.e., when the competing events are not present, then

$$F_h^0(t) = 1 - \exp(-\int_0^t \alpha_h(u)du) = 1 - S_h(t), \text{ say.}$$

That is, '1–KM for cause h', $1 - \widehat{S}_h(t)$, estimates

$$P(\text{dead from cause h before time } t) \quad \textbf{IF} \quad \alpha_j(t) = 0, \quad j \neq h,$$

i.e., if the competing risks do not exist.

1-Kaplan-Meier vs. Aalen-Johansen

We always have:

$$F_h(t) \leq F_h^0(t) = 1 - S_h(t).$$

Thus, the risk is over-estimated by using 1–KM instead of Aalen-Johansen.

The degree of bias depends on the magnitude of the competing causes: if there are no competing risks ($\alpha_j(t) = 0, \quad j \neq h$), then they are identical, and the difference between the two increases with the magnitude of the rates for competing events.

At best, the simple 1–KM estimator can be considered an *approximation* to the cumulative incidence that may be used if the competing risks are *small*. However, the best advice is *never* to use 1-KM in the presence of competing risks.

Kaplan-Meier vs. Nelson-Aalen

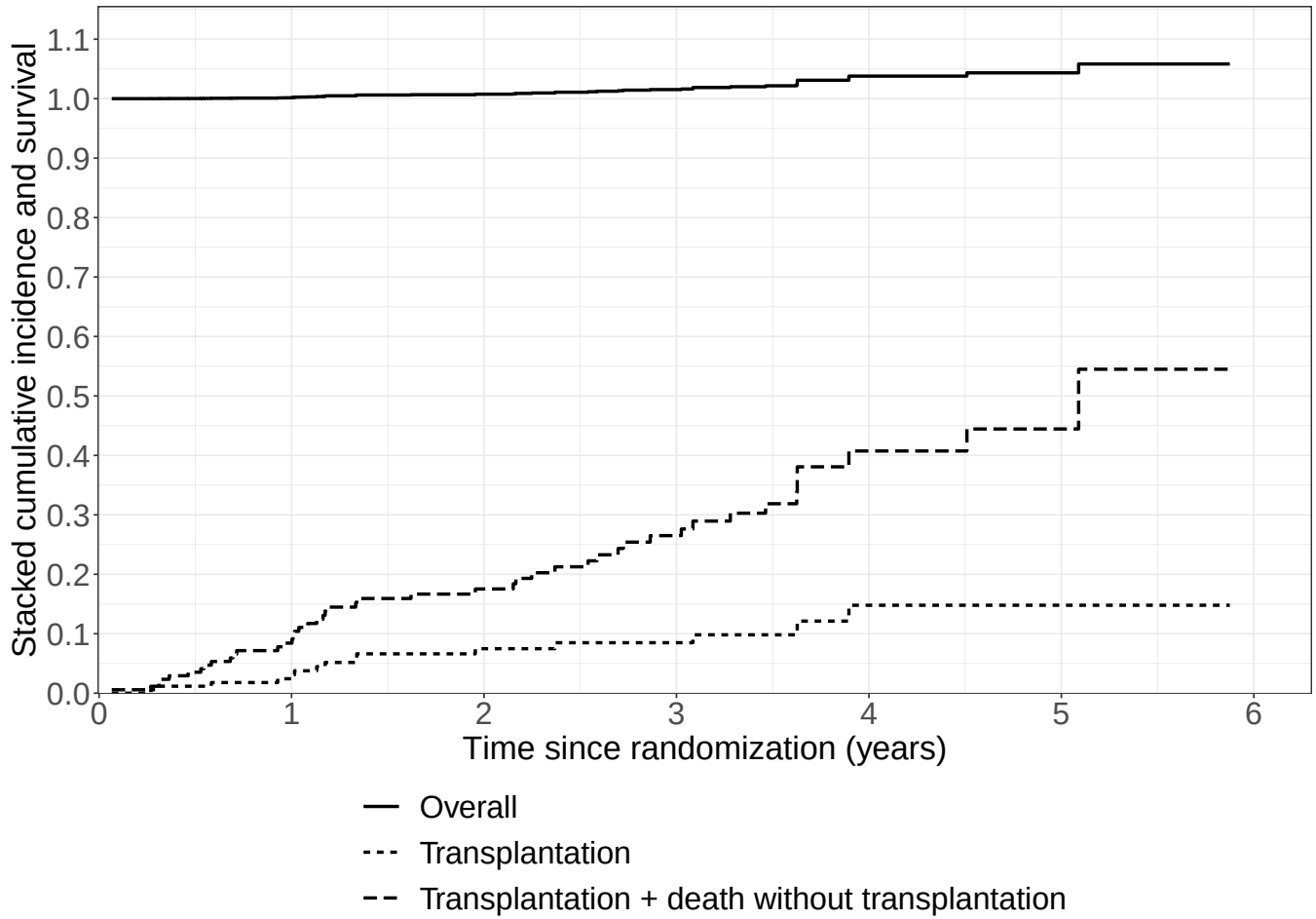
Why does Nelson-Aalen work with competing risks when Kaplan-Meier doesn't?

This has to do with the fact that the Nelson-Aalen estimates the (cumulative) *rate*, and rates describe the 'local' (in time) behavior of the failure process for a given cause. 'Therefore', when assessing the local strength of cause h, causes $j \neq h$ need not be taken into account.

Risks, however, cumulate over (longer) time periods and the impact of competing causes must be accounted for when assessing the strength of cause h.

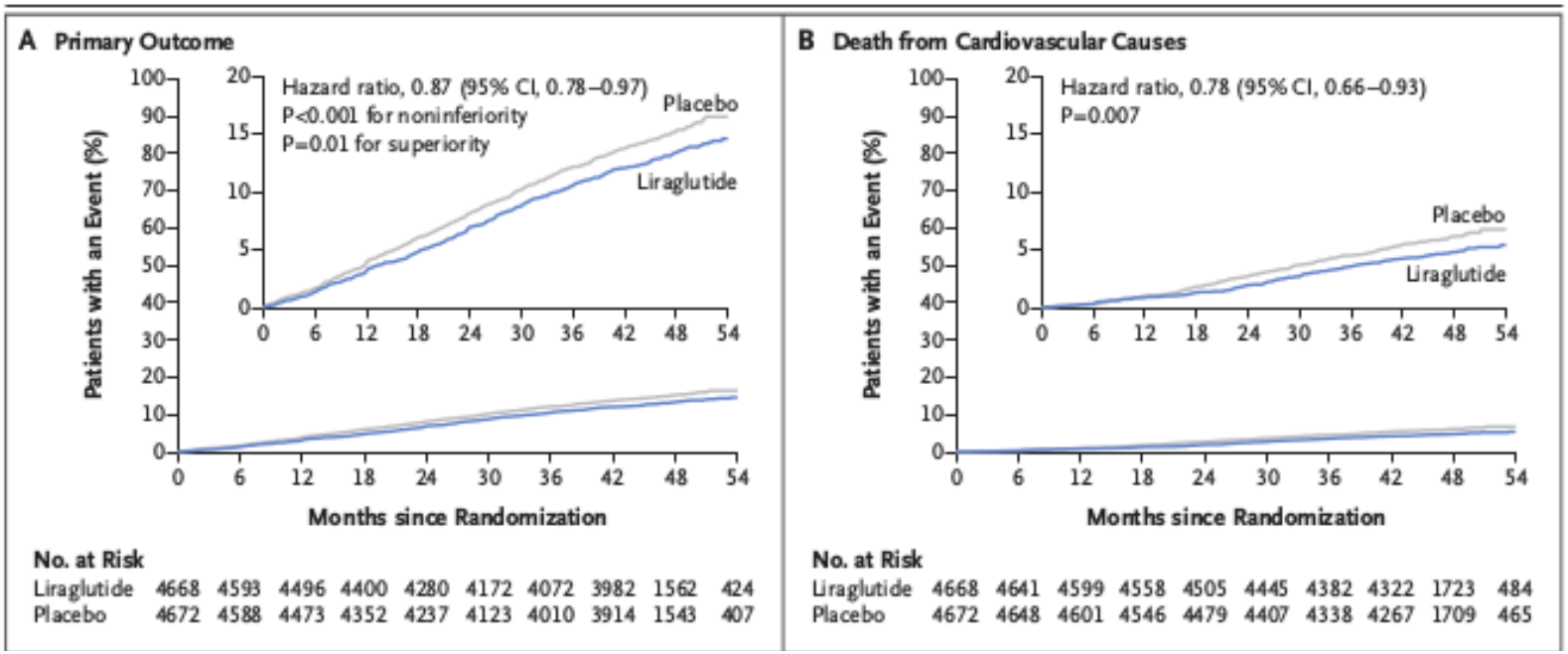
The mathematical argument builds of the likelihood factorization (the Jacod formula).

The PBC-3 trial, death and transplantation



1 minus Kaplan-Meier estimators (placebo) – biased!

The LEADER trial



Estimation of general state occupation probabilities

Both the Kaplan-Meier and Aalen-Johansen estimators are special cases of a general plug-in estimator of *transition probabilities* in Markov processes due to Aalen and Johansen (1978) – more in Part II.

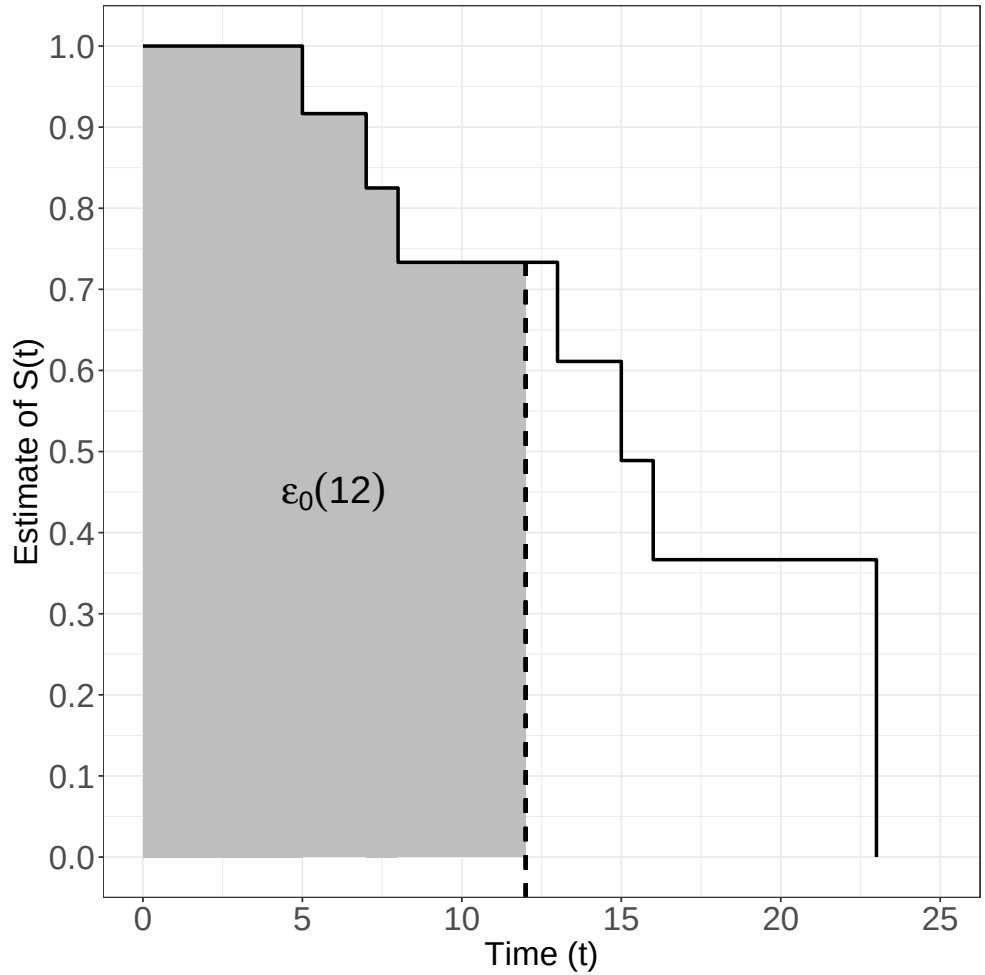
Datta and Satten (2001) showed that for general multi-state processes (i.e., also non-Markov), this estimator is consistent for *state occupation probabilities*.

ELOS

$$\varepsilon_h(\tau) = \int_0^\tau Q_h(t)dt$$

may also be estimated by plug-in.

A small data set



The PBC-3 trial, death and transplantation

	Placebo	CyA
RMST, 3 years	2.606	2.678
Years lost, 3 years		
Transplantation	0.143	0.086
Death without transpl.	0.251	0.236
	3.000	3.000

Estimates are areas under Kaplan-Meier, resp. Aalen-Johansen estimates at τ =3 years.

R code, K-M, Aa-J, ELOS

```
library(survival)
pbcna <- survfit(Surv(days, status != 0) ~ tment,
data = pbc3)
#NB: same as for N-Aa

plot(pbcna)

ajfit <- survfit(Surv(days, factor(status)) ~ tment,
data = pbc3)

plot(ajfit)
print(ajfit,rmean=3*365)
```

Again, the code gives a crude (non-stacked) version of the correct figure that may be improved upon.

Exercises - II

Parametric models

Examples

Non-parametric inference (including the Cox model - more later) has become the standard method in survival analysis.

Useful parametric models do exist:

- The exponential distribution with *constant* hazard: $\alpha(t) = \alpha$ for all t. This is a restrictive assumption which is often not justified. This is the model underlying the calculation of simple 'occurrence/exposure' rates.
- Piecewise exponential models have piecewise constant hazards: $\alpha(t) = \alpha_j$ when $s_{j-1} \leq t < s_j$ for pre-specified intervals, $0 = s_0 < s_1 < ... < s_J = \infty$. This leads to interval-specific occurrence/exposure rates and provides the basis for *Poisson* regression models (more later).
- Another simple extension of the exponential model is the Weibull model with $\alpha(t) = \alpha\gamma t^{\gamma-1}$. Mathematically simple, rather flexible (e.g., both increasing, constant, and decreasing hazard functions, however: value at $t = 0$), but rarely used in practice.
- Models using the log-normal distribution also exist (no simple hazard function).

Likelihood

With the notation $X_1, ..., X_n$ for observation times and $D_i = I(T_i = X_i), \quad i = 1, \ldots, n$ for failure indicators, and when the hazard function is $\alpha_\theta(t)$, the Jacod likelihood formula (*) becomes (density $f_\theta(t)$, survival function $S_\theta(t)$):

$$L(\theta) = \prod_{i=1}^{n} \big(\alpha_\theta(X_i)\big)^{D_i} \exp\Big(-\int_0^{X_i} \alpha_\theta(t)dt\Big) = \prod_{i=1}^{n} f_\theta(X_i)^{D_i} S_\theta(X_i)^{1-D_i}.$$

Standard inference via score function, observed information etc.
Martingale-based proof of 'standard' asymptotic properties: the score $D \log L(\theta_0)$ is a martingale at the true parameter value θ_0.
Also: LR and Wald tests.

When the full distribution of T is parametrically specified via θ, parameters like mean and median are also functions of θ.
Since the right-hand tail of the distribution is not observed because of censoring, one is reluctant to quote the mean.

Piecewise constant hazard

The hazard function is $\alpha(t) = \alpha_j$ when $s_{j-1} \leq t < s_j$ for pre-specified intervals, $0 = s_0 < s_1 < \ldots < s_J = \infty$.
The maximum likelihood estimator is most easily expressed in counting process notation:

$$N(t) = \sum_i I(X_i \leq t, D_i = 1), \quad Y(t) = \sum_i I(X_i \geq t).$$

Then

$$\widehat{\alpha}_j = \frac{N(s_j) - N(s_{j-1})}{\int_{s_{j-1}}^{s_j} Y(t)dt},$$

i.e., number of failures in interval j divided by the total time at risk in interval j. Further, from the observed information:

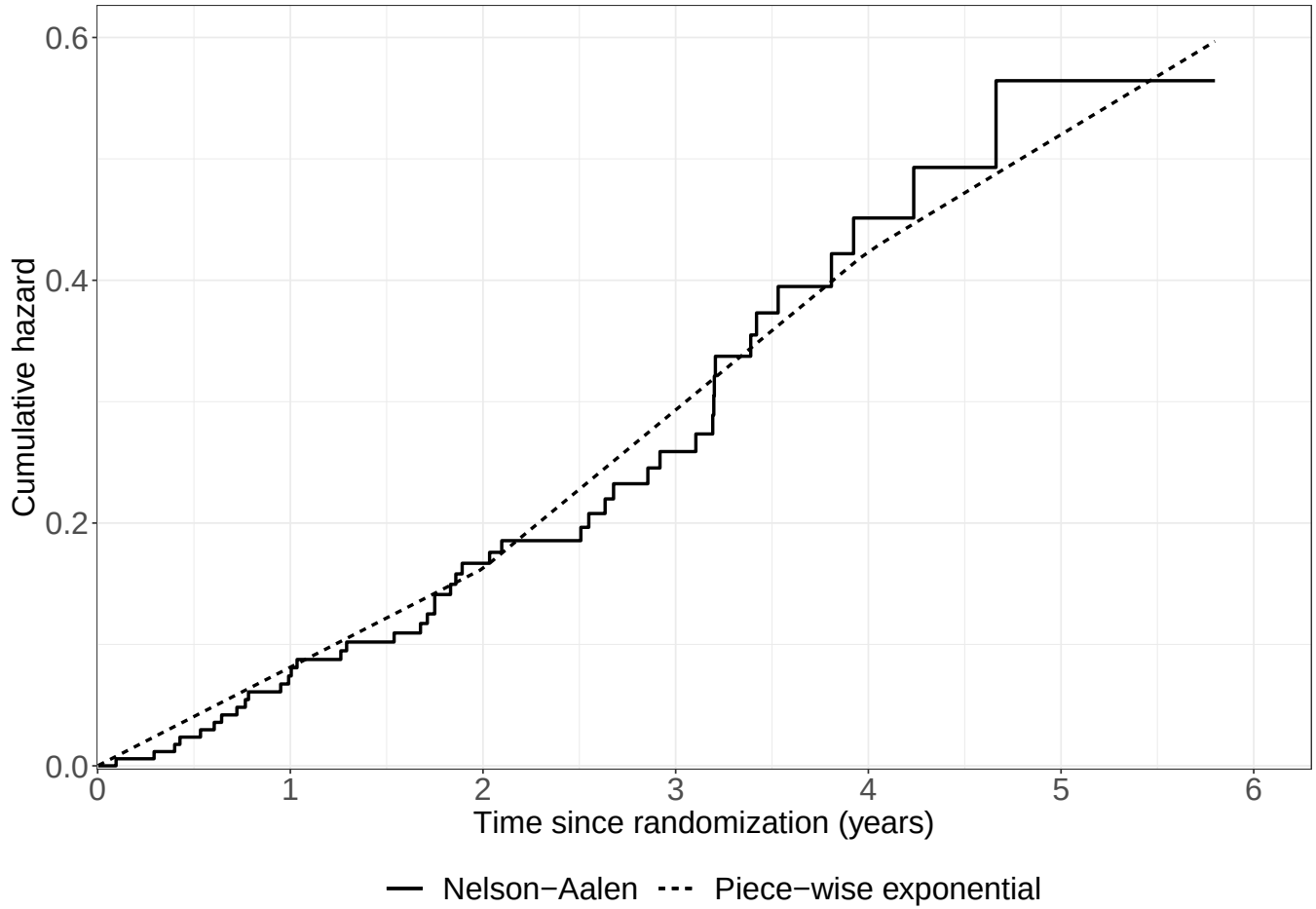
$$SD(\widehat{\alpha}_j) = \frac{\sqrt{N(s_j) - N(s_{j-1})}}{\int_{s_{j-1}}^{s_j} Y(t)dt}.$$

PBC-3 trial

Treatment	Interval ℓ (year)	Events D_ℓ	Risk Time (in years) Y_ℓ	Hazard (per 100 years) $\widehat{\alpha}_\ell$	SD
CyA	0-1	24	295.50	8.1	1.7
	2-3	18	137.67	13.1	3.1
	4-5	2	20.80	9.6	6.8
Placebo	0-1	27	287.08	9.4	1.8
	2-3	17	136.00	12.5	3.0
	4-5	2	23.66	8.5	6.0

PBC3 trial in liver cirrhosis: Events, risk time, and estimated hazards in a piece-wise exponential model by treatment group.

PBC-3 trial



PBC3 trial in liver cirrhosis: Estimated cumulative hazards for the placebo group.

LEADER trial

Marso et al. (2016) quote a single rate for each type of outcome:

	Liraglutide		Placebo	
	Events	Rate per 100 years	Events	Rate per 100 years
Primary end-point	608	3.4	694	3.9
Death from any cause	381	2.1	447	2.5

Delayed entry

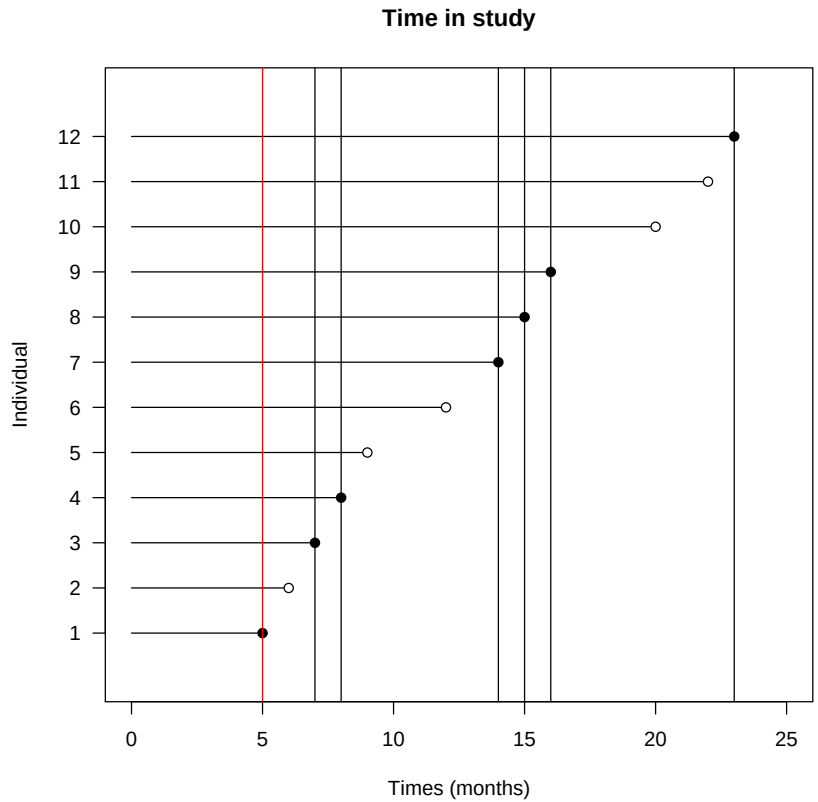
Delayed entry

Some times, subjects are not observed from time 0 but only from a later entry time, B_i, that is, subject i is only observed conditionally on having survived until B_i.

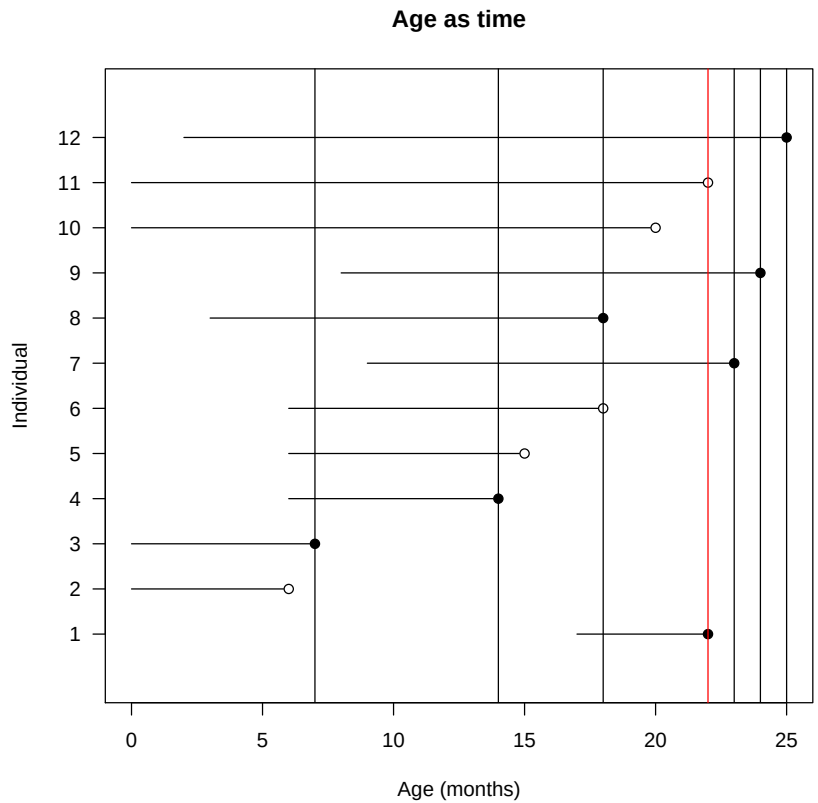
This is denoted as *delayed entry* or *left truncation* and is often present if age is the primary time variable, i.e., quite common in epidemiological studies, less so in trials.

A change of time variable causing delayed entry changes how *risk sets* are composed - see graph for the small data set.

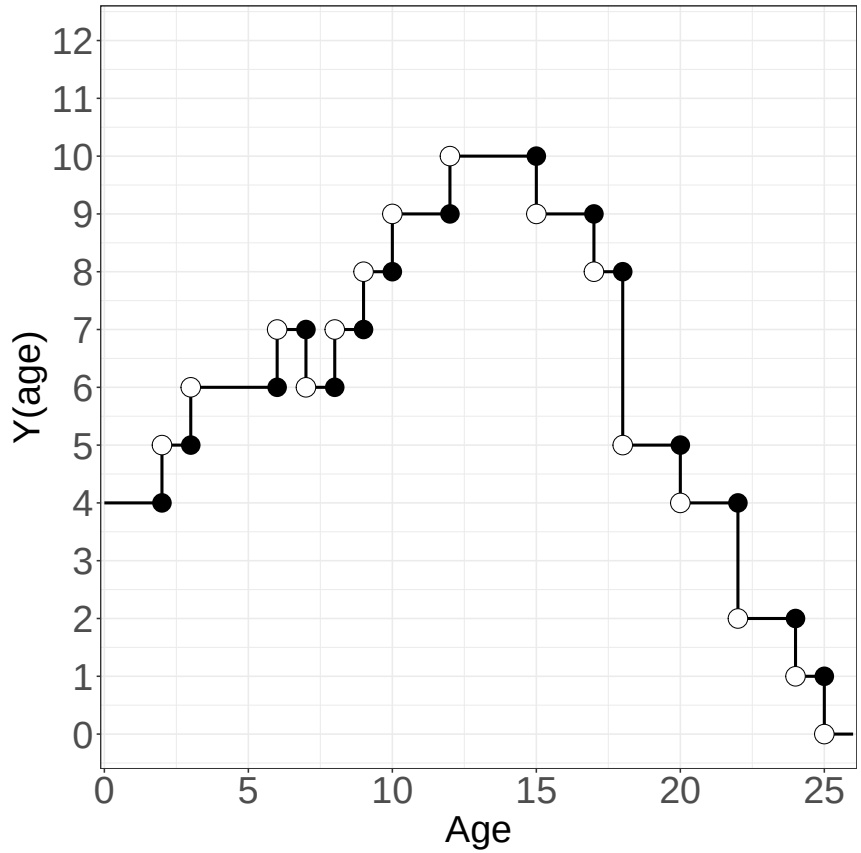
A small data set - risk sets in time



A small data set - risk sets in age



A small data set - risk set size in age



Handling delayed entry

The methods discussed so far immediately generalize to data with delayed entry:

- Re-define the individual at-risk indicator as $I(B_i < t \leq X_i)$,
- re-define risk set size as $Y(t) = \sum_i I(B_i < t \leq X_i)$.
- Then, formulations of estimators etc. via counting process notation $(N(t), Y(t))$ still apply.
- An assumption of 'independent delayed entry' is needed.
- In `R`, the survival object simply has an additional argument corresponding to the time of entry:

  `'Surv(entrytime,exittime,status==...)'`

Summary

- Survival data are a special case of data from a multi-state process.
- Basic parameters in a multi-state model are the transition intensities; however, marginal parameters often have more direct interpretations.
- (Cumulative) intensities may be estimated non-parametrically using the Nelson-Aalen estimator and compared using the logrank test.
- Marginal parameters may be estimated using plug-in, leading to the Kaplan-Meier estimator for the survival function, and the Aalen-Johansen estimator for the competing risks cumulative incidence.
- Parametric models exist but are rarely used (however: piece-wise constant intensity models).
- All methods build on an assumption of independent censoring.

R exercises

We will use the PBC3 data. Description of variables is on the book's web companion https://multi-state-book.github.io/companion/Ch1.html.

First part is on the **two-state model** with composite endpoint of transplantation or death (failure of medical treatment) and second part the **competing risks model** with transplantation and death without transplantation as competing causes.

Each part will first analyze intensities (hazards) and then marginal parameters.

Exercises can be solved using the `survival` package, see several vignettes at https://cran.r-project.org/web/packages/survival/index.html.

Please, add the variable (`years`) to pbc3 data, to use as time in the `Surv()` object.

```
pbc3$years<-pbc3$days/365.25
```

R exercises – two state-model

Intensities

1. Investigate if treatment (`tment`) impacts the hazard using the Nelson-Aalen estimator and the logrank test.
2. Investigate if bilirubin (`bili`) impacts the hazard using the Nelson-Aalen estimator and the logrank test by categorizing bilirubin into quartiles. Also make a binary version of bilirubin with a cut point at the upper quartile and do the same.
3. Use stratified logrank tests to investigate treatment effect adjusted for bilirubin in 4 categories or 2 categories. Interpretation?

Marginals

4. Estimate and plot separate survival curves (using the Kaplan-Meier estimator) for treatment and bilirubin in 2 categories and a plot of treatment within each bilirubin category.
5. Based on the Kaplan-Meier estimates for treatment, estimate the risk difference at year 3, and test if any risk difference between treatment groups.
6. Based on the Kaplan-Meier estimates for treatment, estimate the RMST difference at year 3, and test if any difference between treatment groups.

R exercises – competing risks

Intensities

1. Investigate if treatment impacts the two cause-specific hazards (transplantation and death without transplantation) using the Nelson-Aalen estimator and the logrank test.
2. Similarly, as for the two-state model, investigate if bilirubin impacts the cause-specific hazards.
3. Use stratified logrank tests to investigate treatment effect on the cause-specific hazards, adjusted for bilirubin in 2 categories.

Marginals

1. For each transition, estimate and plot separate cumulative incidences (using the Aalen-Johansen estimator) for treatment and bilirubin (2 categories) for each transition.
2. For each transition and based on the Aalen-Johansen estimates for treatment, estimate the risk differences at year 3, and test if any risk difference between treatment groups.
3. For each transition and based on the Aalen-Johansen estimates, estimate the ELOS (years lost) differences at year 3, and test if any risk difference between treatment groups.