Multi-state models: Rates, Risks, and Pseudo-Values

I Introduction to multi-state models

II Non-parametric estimation and regression models for intensities (Cox)

III Estimation of marginal parameters using plug-in

IV Direct regression models for marginal parameters (Cox, Fine-Gray, Ghosh-Lin)

V Pseudo-values

https://multi-state-book.github.io/Salerno25/

II: Non-parametric estimation and models for intensities

- Nelson-Aalen estimator
- Two-state model: Cox model for survival data
- Competing risks: Cox model for cause-specific hazard
- Recurrent events: PWP and Andersen-Gill model

II: Non-parametric estimation and models for intensities

- Nelson-Aalen estimator
- Two-state model: Cox model for survival data
- Competing risks: Cox model for cause-specific hazard
- Recurrent events: PWP and Andersen-Gill model

It's all about the rates $\alpha_{hj}(t)$

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 non-parametric estimator for

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

is

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

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

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

an increasing step function with steps at observed $h \to j$ transition times. The local slope estimates the intensity. The estimator has a maximum likelihood interpretation.

Likelihood

The Nelson-Aalen estimator has a maximum likelihood interpretation based on *the Jacod formula*: $L = \prod_i L_i$ with, Eq. (3.1),

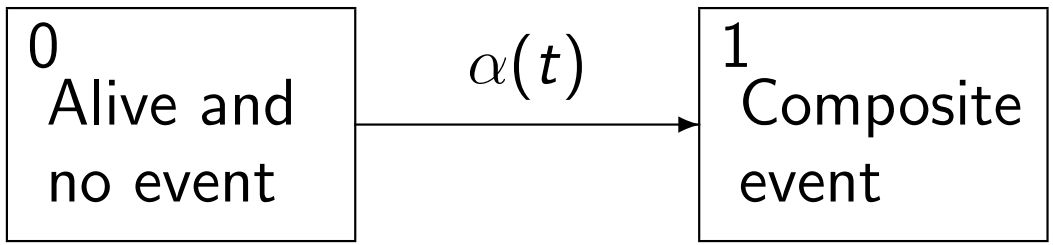
$$L_i = \prod_t \prod_\nu \left(\lambda_{\nu i}(t)^{dN_{\nu i}(t)} \right) \times \exp\left(- \sum_\nu \int_0^\infty \lambda_{\nu i}(u) du \right), \quad (*)$$

where $\lambda_{\nu i}(u) = \lambda_{hji}(u) = \alpha_{hji}(u) Y_{hi}(u)$, i.e., the event *types*, ν, corresponds to 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.

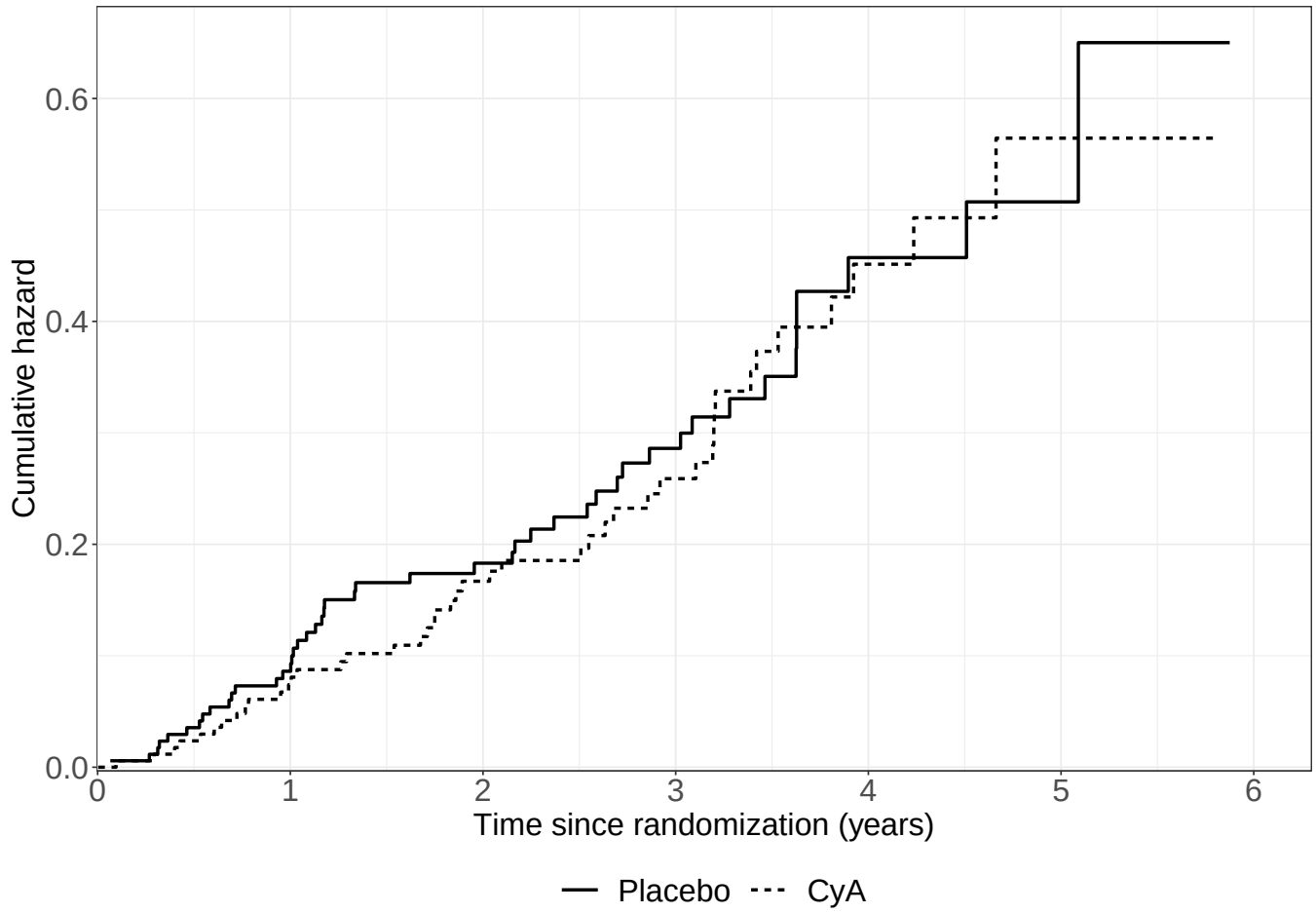
Treating 'jumps' $dA_\nu(t) = \alpha_\nu(t)dt$ as parameters, maximization of $(*)$ leads to the Nelson-Aalen estimator, and an estimate of $\mathrm{SD}(\widehat{A}_\nu(t))$ follows.

The PBC-3 trial, composite endpoint

Failure of medical treatment, i.e., death or transplantation.



Data set PBC-3 description:
https://multi-state-book.github.io/companion/Ch1.html




```
library(survival)
plot(survfit(Surv(days, status != 0) ~ tment, data=pbc3, cumhaz=T)
```

Logrank test

The standard non-parametric test for comparison of intensities is the *logrank test*.

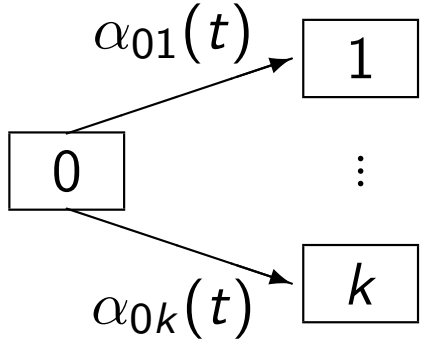
```
library(survival)
survdiff(formula = Surv(years, status != 0) ~ tment, data = pbc3)

          N Observed Expected (O-E)^2/E (O-E)^2/V
tment=0 173       46     44.7    0.0388    0.0771
tment=1 176       44     45.3    0.0382    0.0771

 Chisq= 0.1  on 1 degrees of freedom, p= 0.8
```

A stratified version is available

```
survdiff(Surv(days, status != 0) ~ tment+strata(sex), data = pbc3)
```

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

Competing risks



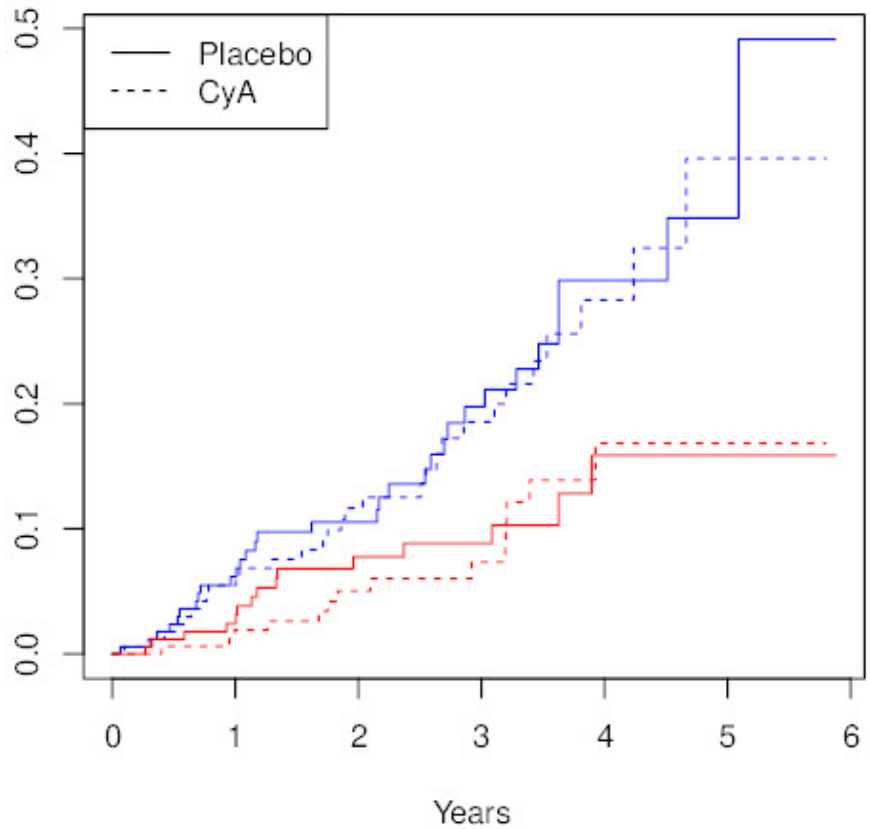
For the competing risks model, the possible transitions are $\nu = 0 \to h, \quad h = 1, \ldots, k$.

Since the Jacod formula (*) *factorizes* over transition types (i.e., causes of death), estimation and test 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, e.g., the `R` code, e.g., for death without transplantation in the PBC-3 study, both for Nelson-Aalen and logrank test:

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

PBC-3: cause-specific hazards

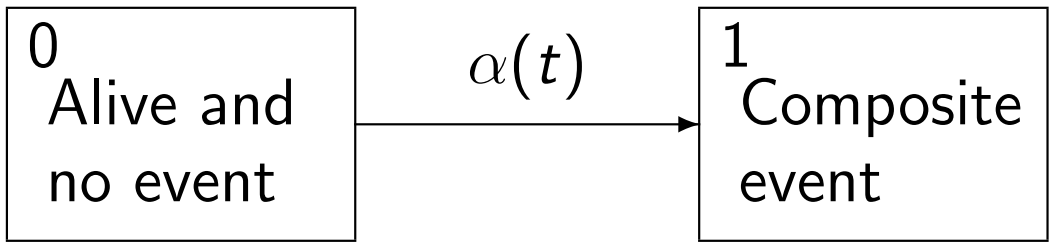


Transplantation (red), death without transplantation (blue)

Cox model

Two-state model for PBC3 composite endpoint

Composite endpoint 'failure of medical treatment', i.e., death or transplantation, as motivating example.



t = time since randomization

$N_i(t)$ = number of events observed in $[0, t]$ for patient $i = 1, \ldots, n$

$Y_i(t)$ = indictor for patient i being observed in state 0 at time $t-$

X_i = observation time (failure or censoring)

$D_i = I(T_i = X_i)$ the failure indicator

Cox regression model for two-state model

The model assumes that the hazard $\alpha(t)$ for the patient i is

$$\begin{aligned} \alpha_i(t) &= \alpha_0(t)\exp(\beta_1 Z_{i1} + \beta_2 Z_{i2} + \cdots + \beta_p Z_{ip}) \\ &= \alpha_0(t)\exp(\beta^{\mathsf{T}} Z_i) = \alpha_0(t)\exp(\mathrm{LP}_i) \end{aligned}$$

where $\beta_1, \beta_2, \ldots, \beta_p$ are regression parameters, Z_{i1} is the covariate value for covariate 1 for individual i, *etc.* and $\mathrm{LP}_i = \beta^{\mathsf{T}} Z_i$ is the *linear predictor*.

The *baseline hazard*, $\alpha_0(t)$, is the only part depending on time t and is the hazard of an individual having all covariates equal to zero.

The Cox model does not make any parametric assumptions for the baseline hazard. Focus is on the regression parameters.

Binary covariate

We study the effect of a single binary covariate, e.g., randomized treatment (placebo, active) on the rate of composite endpoint

$$Z_i = \begin{cases} 0 & \text{if patient } i \text{ was in the placebo group} \\ 1 & \text{if patient } i \text{ was in the active group} \end{cases}$$

The Cox model is $\alpha_i(t) = \alpha_0(t)\exp(\beta Z_i)$ and the *hazard ratio* (HR) between active and placebo is

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

HR is assumed independent of time, i.e., *proportional hazards* over time. On the log-scale

$$\log\left[\alpha_0(t)\exp(\beta)\right] - \log\left[\alpha_0(t)\right] = \beta.$$

Cox's partial likelihood function

With the notation $X_1, ..., X_n$ for observation times and $D_i = I(T_i = X_i)$ for failure indicators, Cox's partial likelihood function is

$$\mathrm{PL}(\beta) = \prod_{i=1}^{n} \prod_{t} \left(\frac{Y_i(t) \exp(\beta Z_i)}{\sum_j Y_j(t) \exp(\beta Z_j)} \right)^{dN_i(t)} = \prod_{i=1}^{n} \left(\frac{\exp(\beta Z_i)}{\sum_{j \in R(X_i)} \exp(\beta Z_j)} \right)^{D_i}$$

where $R(t) = \{j : Y_j(t) = 1\}$ is the *risk set* at time t.

Math:
The partial likelihood function may be obtained from the general likelihood function (Jacod formula) by profiling out the baseline hazard function $\alpha_0(t)$. (MSB p. 76)

Math

From Cox's partial likelihood

$$\mathrm{PL}(\beta) = \prod_{i=1}^{n} \prod_{t} \left(\frac{Y_i(t) \exp(\beta Z_i)}{\sum_j Y_j(t) \exp(\beta Z_j)} \right)^{dN_i(t)}$$

we get the Cox *score*

$$U(\beta) = \frac{d}{d\beta} \log(\mathrm{PL}(\beta)) = \sum_{i=1}^{n} \int_0^{\infty} \left(Z_i - \frac{\sum_j Y_j(t) Z_j \exp(\beta Z_j)}{\sum_j Y_j(t) \exp(\beta Z_j)} \right) dN_i(t).$$

When evaluated at the true parameter value (β_0) and considered as a process in t (i.e., integrating to t instead of ∞), $U_t(\beta_0)$ is a *martingale*. The martingale CLT gives asymptotic normality of the score and we can get asymptotic normality of $\widehat{\beta}$.

Score test and logrank test

The variance of $\widehat{\beta}$ is estimated from the second derivative and the distribution of standard likelihood-based tests are also obtained.

One such test is the *score test* and, for a binary (and categorical) covariate the score test is the *logrank test*. Thus, the logrank test is closely related to the Cox model and, therefore, has certain optimality properties against proportional hazards alternatives.

Cox in R

```
pbc3 <- read.csv("data/pbc3.csv")
library(survival)

# Breslow ties handling
coxph(Surv(days,status!=0)~tment, data=pbc3, method="breslow")

  n= 349, number of events= 90

          coef exp(coef) se(coef)      z Pr(>|z|)
tment -0.05854   0.94314  0.21092 -0.278    0.781

      exp(coef) exp(-coef) lower .95 upper .95
tment    0.9431       1.06    0.6238     1.426

Likelihood ratio test= 0.08  on 1 df,   p=0.8
Wald test            = 0.08  on 1 df,   p=0.8
Score (logrank) test = 0.08  on 1 df,   p=0.8
```

The Breslow estimator

The cumulative baseline hazard $A_0(t) = \int_0^t \alpha_0(s)ds$ from the Cox model $\alpha_i(t) = \alpha_0(t)\exp(\beta Z_i)$ can be estimated by the Breslow estimator

$$\widehat{A}_0(t) = \int_0^t \frac{\sum_i dN_i(u)}{\sum_i Y_i(u)\exp(\widehat{\boldsymbol{\beta}}^{\mathsf{T}}\boldsymbol{Z}_i)} \qquad \text{Eq. (3.18)}$$

where $\widehat{\beta}$ is the maximum likelihood estimate of β.

Having no covariates, the Breslow estimator is the Nelson-Aalen estimator

$$\widehat{A_0}(t) = \int_0^t \frac{\sum_i dN_i(u)}{\sum_i Y_i(u)}.$$

The Breslow estimator in R

```
fitb<-coxph(Surv(days,status!=0)~tment, data=pbc3, method="breslow")

## centered=FALSE sets all covariates equal to 0, if TRUE returns data
## from a predicted curve for the covariate values fitb$means

plot(survfit(fitb, centered = F), cumhaz = T)

plot(survfit(fitb, newdata = data.frame(tment=0:1)),
     lty=1:2, cumhaz = T)
```

Multiple Cox regression in R

```
coxph(formula = Surv(days, status != 0) ~ tment + alb + bili,
    data = pbc3, method = "breslow")

             coef  exp(coef)   se(coef)      z        p
tment -0.4964995  0.6086576  0.2256244 -2.201   0.0278
alb   -0.1156850  0.8907558  0.0212814 -5.436 5.45e-08
bili   0.0089491  1.0089893  0.0009801  9.130  < 2e-16

Likelihood ratio test=99.06  on 3 df, p=< 2.2e-16
n= 343, number of events= 88
   (6 observations deleted due to missingness)
```

Stratified Cox model

If the assumption of proportional hazards is questionable for a *categorical* covariate with, say, m categories it is possible to expand the Cox model to include different baseline hazards for each category

$$\alpha_i(t) \quad = \quad \alpha_{j0}(t)\exp(\beta^{\mathsf{T}} Z_i), \text{ when } i \text{ is in stratum } j$$

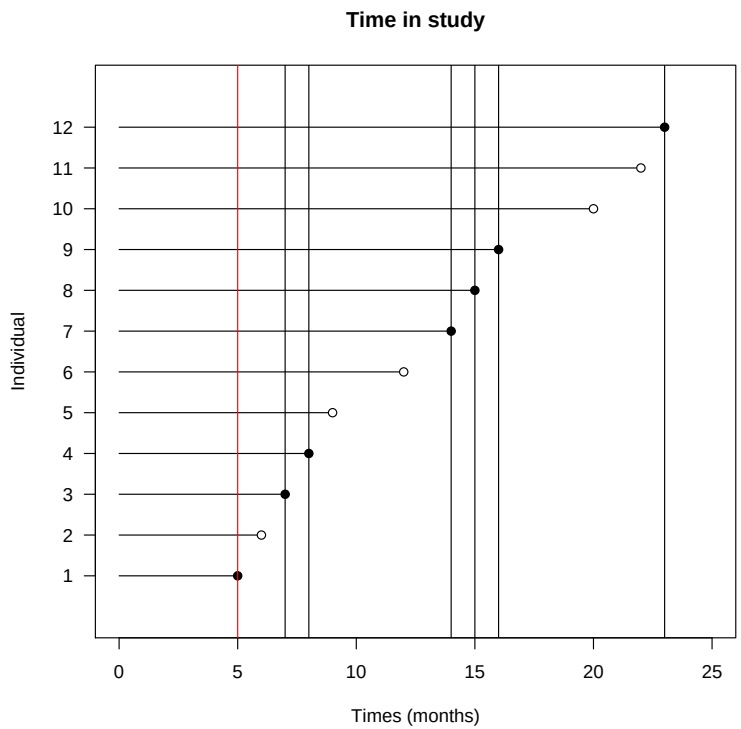
where $\alpha_{j0}(t)$ is the baseline hazard in stratum $j = 1, \ldots, m$.

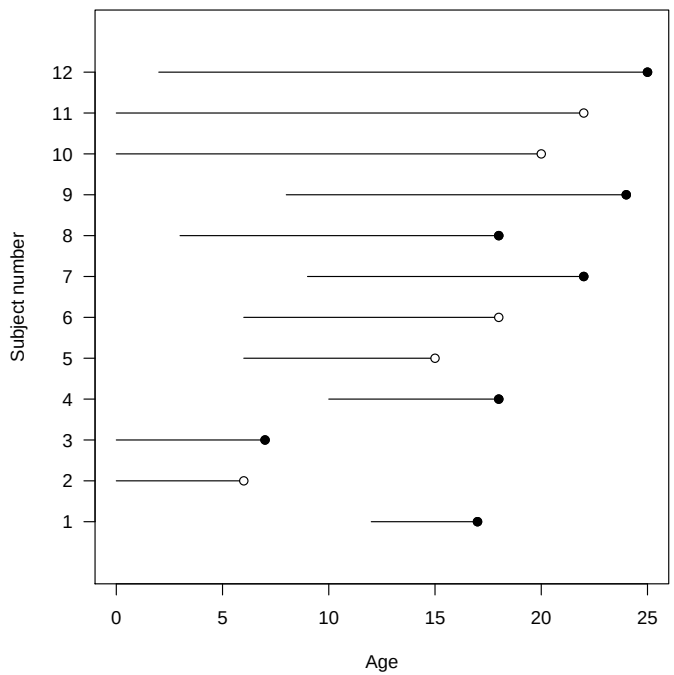
The effect of the stratified covariate is not estimated directly, but other covariates are adjusted for the stratified covariate. The effect of the covariates are assumed equal over strata.

```
coxph(Surv(days,status!=0)~ tment + strata(sex),
      data=pbc3, method="breslow")
```

Delayed entry

Not often in randomized trials, but often so in epidemiological studies, subjects are only becoming at risk at a certain age or time. To be included in the sample, a subject must survive until the date that the sample is identified. This type of incomplete observation is denoted *left-truncation* or *delayed entry*.





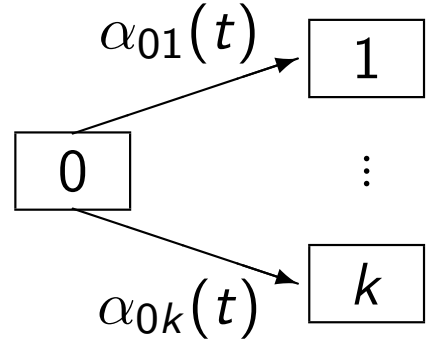
If the truncation mechanism satisfies a condition of "independent truncation" similar to that of "independent censoring" then left-truncation is done by careful control of the risk sets in

$$\mathrm{PL}(\beta) = \prod_{i=1}^{n} \Big(\frac{\exp(\beta Z_i)}{\sum_{j \in R(X_i)} \exp(\beta Z_j)} \Big)^{D_i}$$

Only individuals at risk and under observation are included in the risk set $R(t)$ at time t. Additionally, a time of entry (e.g., age at entry) into the study will be needed in the data.

```
coxph(Surv(inage,outage,dead) ~ z, data=epidata)
```

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*:

$$\begin{aligned} Q_0(t) = S(t) &= P(\text{alive time } t) \\ &= \exp(-\int_0^t (\alpha_{01}(u) + \cdots + \alpha_{0k}(u))du) \end{aligned}$$

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

$$\begin{aligned} Q_h(t) = F_h(t) &= P(\text{dead from cause } h \text{ before time } t) \\ &= \int_0^t S(u)\alpha_{0h}(u)du. \end{aligned}$$

Cox models for cause-specific hazards

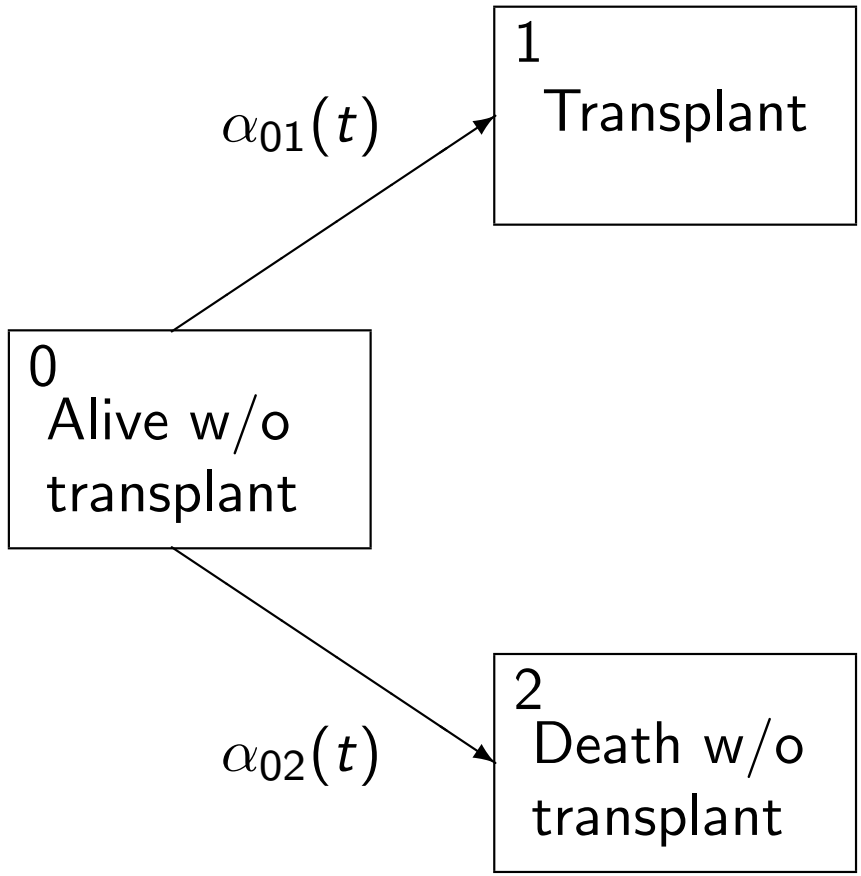
Model for cause h:

$$\alpha_h(t \mid Z) = \alpha_{0h}(t) \exp(\beta_h^{\mathsf{T}} Z),$$

that is, separate baseline hazards and separate regression coefficients for each cause.

As argued previously, the Jacod formula (slide 5) *factorizes* over transition types (i.e., causes of death), and cause-specific hazard models can be analysed *one cause at a time*, formally, treating other causes in the same way as censored observations.

PBC3 as competing risks



Cox models for cause-specific hazards

Table 2.13 *PBC3 trial in liver cirrhosis: Estimated coefficients (and SD) from Cox models for death without transplantation, transplantation, and failure of medical treatment, respectively.*

Event type	Covariate		$\widehat{\beta}$	SD
Death without transplantation	Treatment	CyA vs. placebo	-0.420	0.268
	Albumin	per 1 g/L	-0.070	0.029
	$\log_2$(Bilirubin)	per doubling	0.692	0.093
	Sex	male vs. female	-0.486	0.319
	Age	per year	0.073	0.016
Transplantation	Treatment	CyA vs. placebo	-0.673	0.413
	Albumin	per 1 g/L	-0.094	0.039
	$\log_2$(Bilirubin)	per doubling	0.832	0.147
	Sex	male vs. female	-0.204	0.563
	Age	per year	-0.048	0.021
Failure of medical treatment	Treatment	CyA vs. placebo	-0.510	0.223
	Albumin	per 1 g/L	-0.071	0.023
	$\log_2$(Bilirubin)	per doubling	0.738	0.078
	Sex	male vs. female	-0.585	0.267
	Age	per year	0.031	0.012

Cox models for cause-specific hazards in R

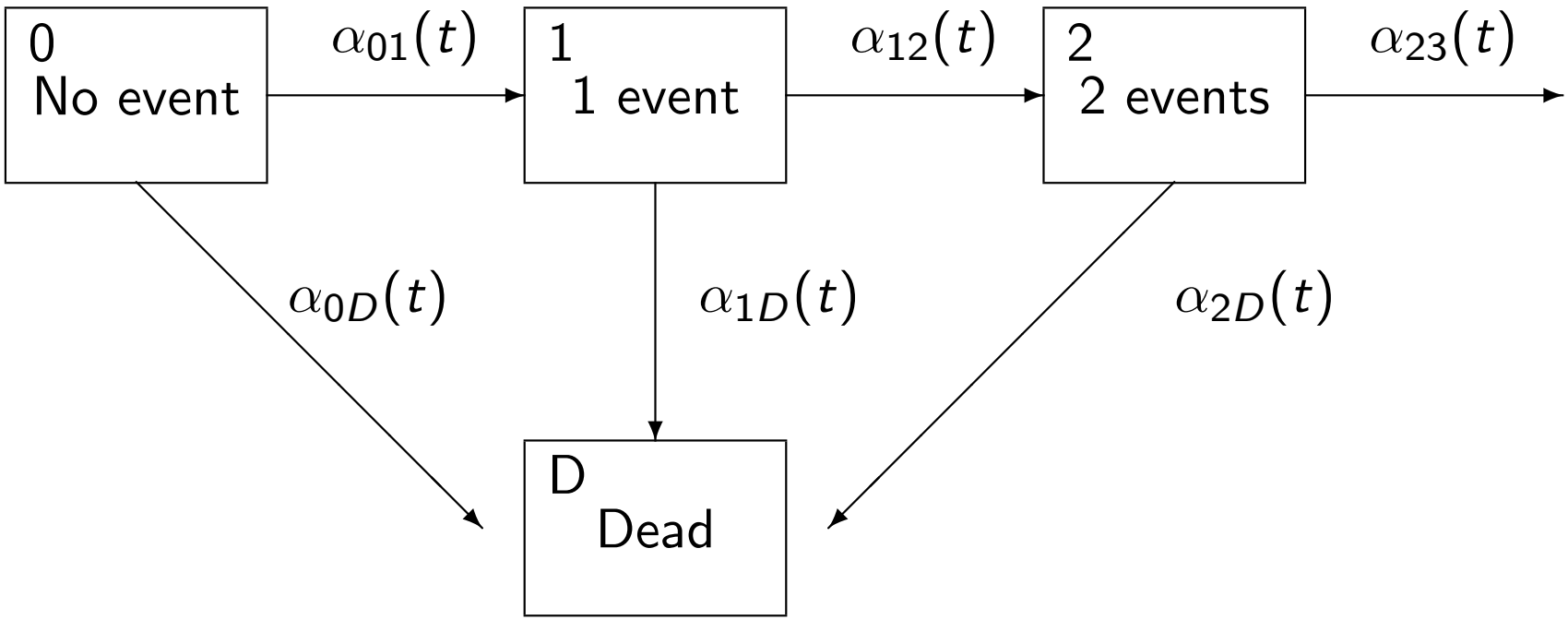
```
# Death without transplantation
coxph(Surv(days, status == 2) ~ tment,
      method = "breslow", data = pbc3)

# Both transitions: id variable needed
coxph(Surv(days, factor(status)) ~ tment,
      method = "breslow", data = pbc3, id=id)

# different models for each cause
# sex only in model for death transition
coxph(list(Surv(days,factor(status))~ tment,
           1:3 ~ sex),
           method = "breslow", data = pbc3, id=id)
```

Checking proportional hazards

This is something special for the Cox model.

- Graphical methods based on the stratified model, e.g., by plotting $\log(\widehat{A}_{0s}(t))$ against t (or $\log(t)$) for each stratum s and see if curves have constant vertical distance
- Modeling time-dependent effects via interactions with functions of time, e.g. add $Z \cdot I(t > \tau)$ or $Z \cdot \log(t)$ to a model including Z
- Schoenfeld (score) residuals

Recurrent events (with competing risks)



LEADER trial, Ex. 1.1.6

Marso et al. (2016, *NEJM*)

- 9340 patients with Type 2 diabetes and high cardiovascular risk,
- randomized 1:1 to liraglutide or placebo
- Global trial (410 sites in 32 countries)
- Primary outcome: composite end-point including death from cardiovarcular cause, non-fatal myocardial infarction (MI), or non-fatal stroke – so-called 'MACE'.
- 4668 received liraglutide with 608 events (13.0%),
- 4672 received placebo with 694 events (14.9%).
- Duration 5 years

Example of recurrent events in LEADER:

- Recurrent MI and all-cause mortality

Recurrent events, the PWP model

In all of the recurrent events scenarios studied (i.e., with or without duration and with or without mortality) intensity models may be used.
One option is to use separate models (possibly with shared covariate effects) for $\alpha_{h,h+1}(t), h = 0, 1, ...$

This is known as the *PWP model* (Prentice, Williams and Peterson, *Biometrika*, 1981):

$$\alpha_{h,h+1}(t \mid Z) = \alpha_{h0}(t) \exp(\beta^{\mathsf{T}} Z), h = 0, 1, ...$$

This is a Cox model with *time-dependent strata*.
The effect of Z might vary with the number of previous events, i.e., β_h instead of β.

Note that a PWP model may not be obvious for randomized studies because 'randomization is lost' after first event.

Recurrent events, the AG model

Assuming proportionality between the different $\alpha_{h0}(t)$, a special case of the *AG* model (Andersen and Gill, *Ann. Statist*, 1982) is obtained.

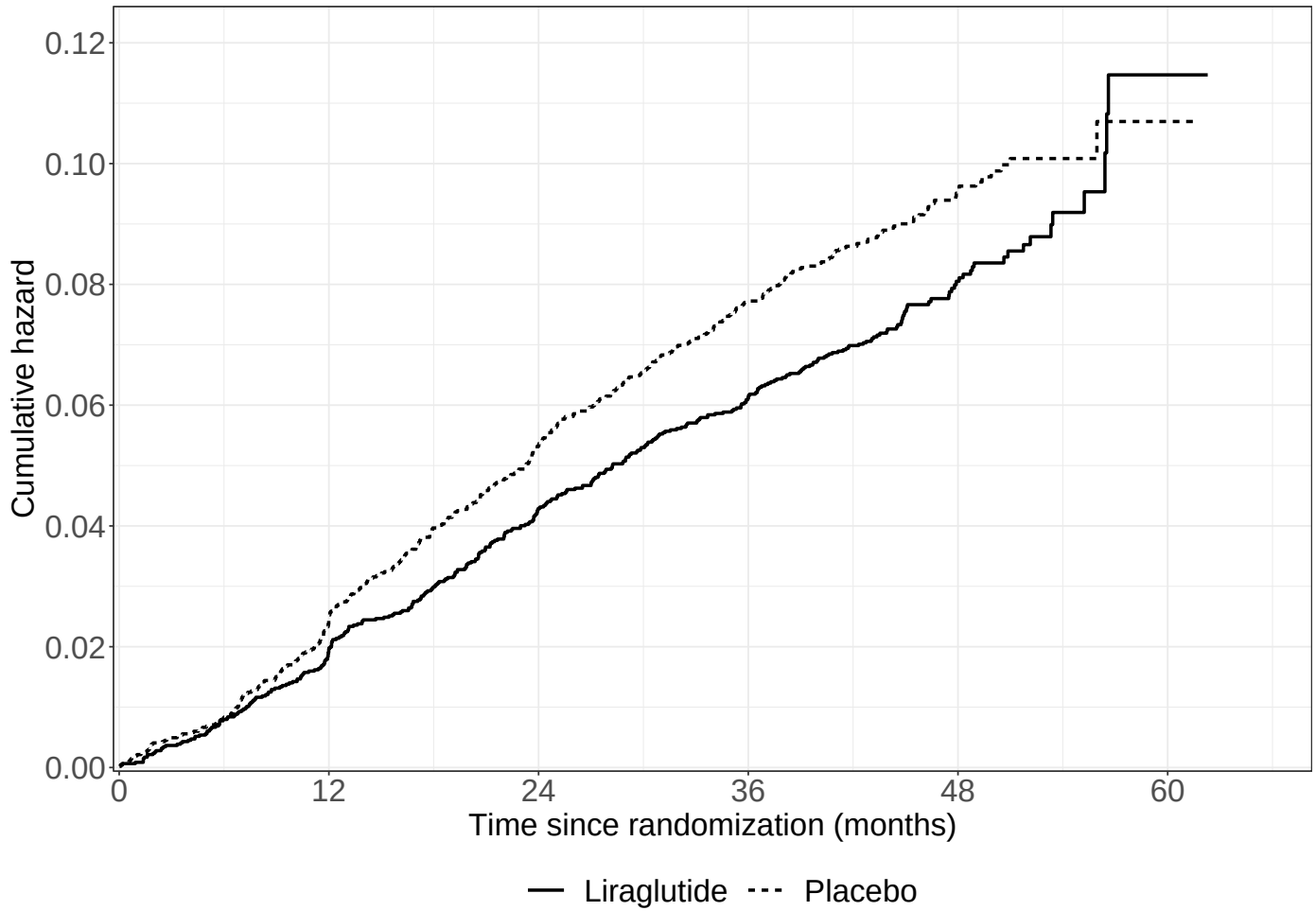
In this model there is a single event intensity:

$$\alpha(t \mid Z) = \alpha_0(t) \exp(\beta^{\mathsf{T}} Z(t))$$

that is allowed to depend on the past (e.g., number of previous events $N(t-)$) via *time-dependent covariates* which may also interact with other covariates.

The intensity model can be fitted using delayed entry for later events. Similar to the Breslow estimator, the integrated baseline intensity can be estimated and with no covariates this equals the Nelson-Aalen estimator (again using delayed entry for later events).

The LEADER trial, recurrent MI



AG model: $\widehat{\beta} = -0.164\,(SD = 0.072)$

R code (examples)

```
library(survival)
NAafit <- survfit(Surv(start, stop, mistatus == 1) ~
treat, data = leader_mi)

plot(NAafit, cumhaz=TRUE)

agfit <- coxph(Surv(start, stop, mistatus == 1) ~ treat,
data = leader_mi)
```

R exercises

We will first 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 for the **two-state model** with composite endpoint of transplantation or death (failure of medical treatment) and second part for the **competing risks model** with transplantation and death without transplantation as competing causes.

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. Also, add a binary version of bilirubin cut at upper quartile (42.3 micromoles/L):

```
pbc3$years<-pbc3$days/365.25
pbc3$bili2<-cut(pbc3$bili,c(0,42.3,454),na.rm=T)
```

R exercises

Two-state model

1. Investigate (as in the lectures) if treatment (`tment`) impacts the hazard using the Nelson-Aalen estimator and the logrank test.
2. Use the binary version of bilirubin (see previous slide) to investigate (using Nelson-Aalen and stratified logrank test), if treatment adjusted for bilirubin (in the binary version) impacts the endpoint. Interpretation?
3. Investigate if treatment impacts the hazard using a Cox model (with treatment as only covariate). Compare the score and logrank test from before.
4. Estimate treatment effect adjusted for bilirubin (as binary covariate) using a Cox model.
5. Estimate the treatment effect adjusted for bilirubin as binary, now in a stratified Cox model. Compare the score test to the stratified logrank test.

R exercises

Competing risks

6. Reproduce the graph for cause-specific Nelson-Aalen estimators from slide 10
7. Reproduce Table 2.13, slide 28.

Recurrent events

Use data set `rr.csv` and see variables on
`https://multi-state-book.github.io/Salerno25/`

8. Plot Nelson-Aalen estimates for the recurrent event rate by treatment arm.
9. Fit an Andersen-Gill model with treatment as covariate.
10. Add a binary time-dependent variable defined as 'had a previous event' (yes/no). What happened to the treatment effect?