INTENSITY MODELS
Cox Regression

(sections 2.2.1, 3.3, 4.1, 5.8.4)

Henrik Ravn

Novo Nordisk A/S

DSBS Course
Survival Analysis in Clinical Trials
2025

Overview

- Cox regression for two-state model
- Stratified Cox model
- Delayed entry
- Time-dependent covariates
- Goodness-of-fit I
- Plug-in estimation of marginal parameters
- Cox regression for competing risks
- Goodness-of-fit II

Cox regression for two-state model

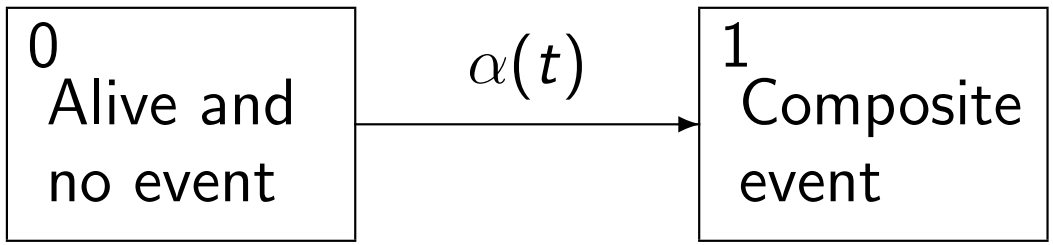
The PBC3 trial in liver cirrhosis, Ex. 1.1.1

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 Cyclosporin A (CyA) (immunosuppressive agent) (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

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.

On the log-scale the Cox model

$$\alpha_i(t) = \alpha_0(t) \exp(\mathsf{LP}_i)$$

becomes

$$\log(\alpha_i(t)) = \log(\alpha_0(t)) + \mathsf{LP}_i.$$

- The baseline hazard $\alpha_0(t)$ is non-parametric.
- The effects of covariates are additive and linear on the log-rate scale.

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

$$\alpha_i(t) \quad = \quad \alpha_0(t) \exp(\beta Z_i)$$

and we get

$$\alpha_i(t) = \begin{cases} \alpha_0(t) & \text{if patient } i \text{ was in the placebo group} \\ \alpha_0(t) \exp(\beta) & \text{if patient } i \text{ was in the active group} \end{cases}$$

Hazard Ratio

The *hazard ratio* (HR) between active and placebo is

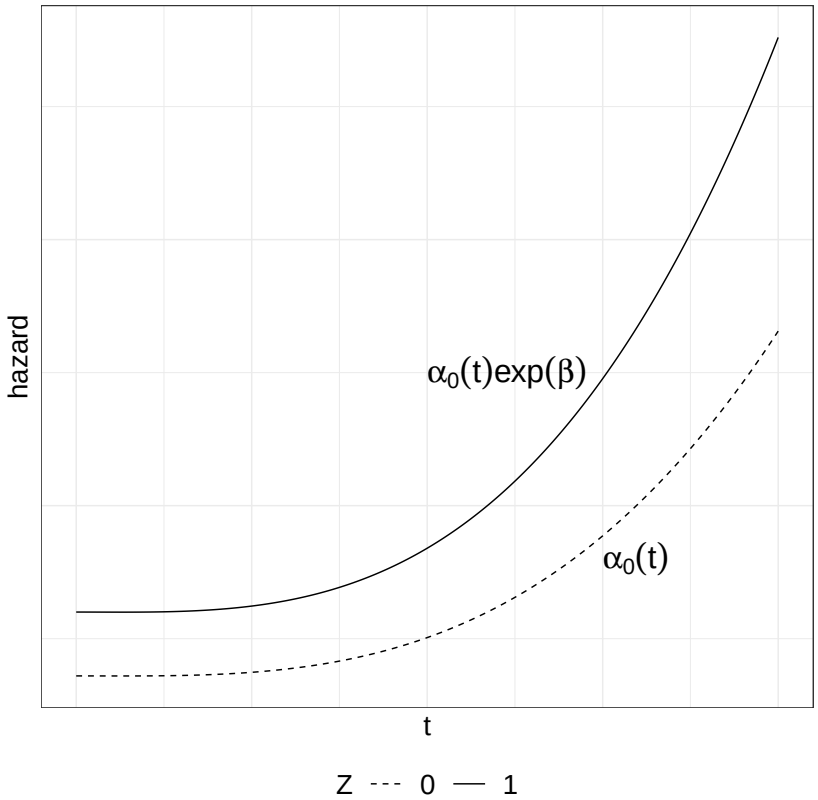
$$\text{HR} = \frac{\alpha_0(t)\exp(\beta)}{\alpha_0(t)} = \exp(\beta).$$

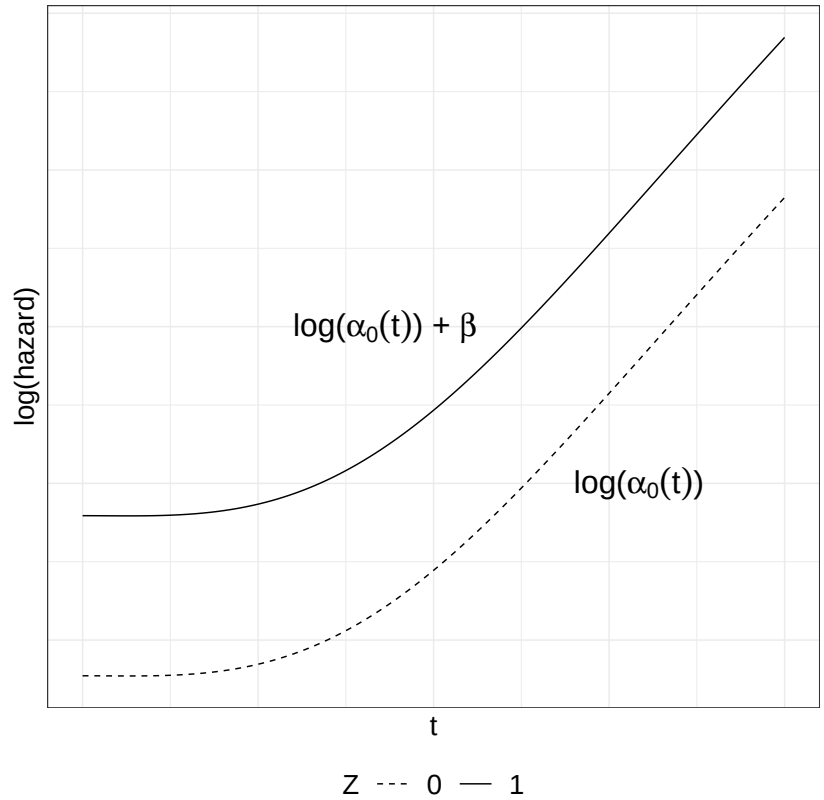
The ratio is independent of time, i.e., *proportional hazards* over time. The placebo group is the *reference*. On the log-scale

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

- HR < 1 ($\beta < 0$), active lower rate than placebo
- HR $= 1$ ($\beta = 0$), active and placebo have the same rate
- HR > 1 ($\beta > 0$), active higher rate than placebo

Proportional hazards





Cox's partial likelihood function

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, Cox's partial likelihood function is

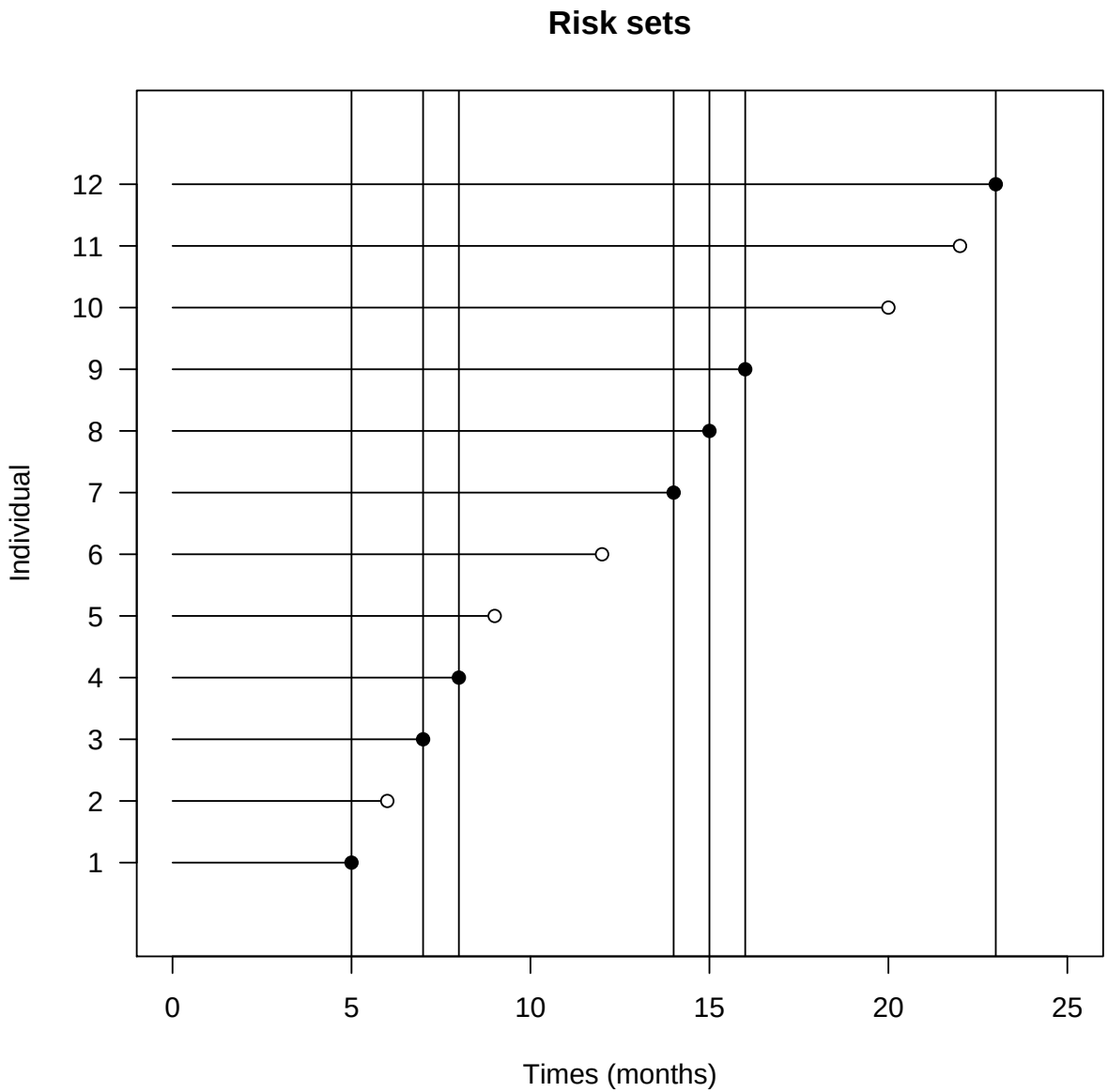
$$\mathrm{PL}(\beta) = \prod_{i=1}^{n} \prod_{t} \left(\frac{Y_i(t) \exp(\beta Z_i)}{\sum_j Y_j(t) \exp(\beta Z_j)} \right)^{dN_i(t)} = \prod_{i=1}^{n} \left(\frac{\exp(\beta Z_i)}{\sum_{j \in R(X_i)} \exp(\beta Z_j)} \right)^{D_i}$$

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

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

Estimates of the parameters are obtained by maximizing $\mathrm{PL}(\beta)$ and the usual type of large-sample likelihood methods also apply to partial likelihoods when censoring is independent and certain regularity assumptions are satisfied.

Risk sets


Risk sets
Individual
12
11
10
9
8
7
6
5
4
3
2
1
0
5
10
15
20
25
Times (months)

Time

The time-variable t is adjusted for by comparing individuals at the same time t – think about the risk sets. However, we don't get an estimate of the effect of the time-variable on the event, but may model interactions with time and covariates.

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

- Time is "automatically" adjusted for.

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 $\widehat{\beta}$ may be 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.

Ties

Now, let $t_i, i = 1, \ldots, d$ index the d distinct event times and d_i no of events at time t_i and $\mathcal{D}_i$ the set of individuals that fail at t_i. For no ties ($d_i = 1$) the partial likelihood function is

$$\text{PL}(\beta) = \prod_{i=1}^{d} \frac{\exp(\beta Z_i)}{\sum_{j \in R(t_i)} \exp(\beta Z_j)}$$

Breslow's method for ties:

$$\text{PL}(\beta) = \prod_{i=1}^{d} \frac{\exp(\beta \sum_{j \in \mathcal{D}_i} Z_j)}{(\sum_{j \in R(t_i)} \exp(\beta Z_j))^{d_i}}$$

Efron's method for ties:

$$\text{PL}(\beta) = \prod_{i=1}^{d} \frac{\exp(\beta \sum_{j \in \mathcal{D}_i} Z_j)}{\prod_{l=1}^{d_i} (\sum_{j \in R(t_i)} \exp(\beta Z_j) - \frac{l-1}{d_i} \sum_{j \in \mathcal{D}_i(t_i)} \exp(\beta Z_j))}$$

Cox in R

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

## binary - Efron ties handling is R default
coxph(Surv(days,status!=0)~tment, data=pbc3)

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

```
coxph(formula = 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.

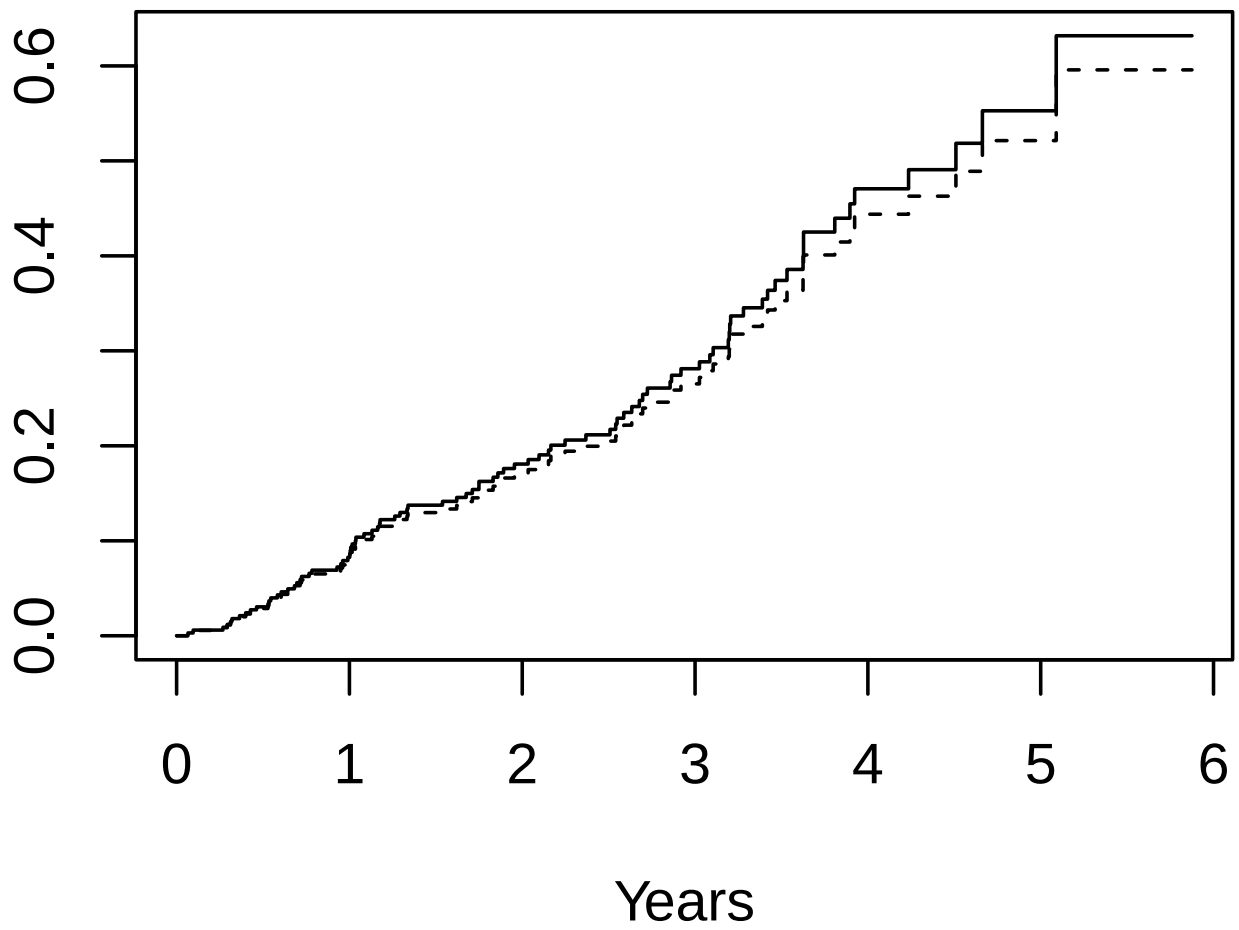
Having no covariates, the Breslow estimator is the Nelson-Aalen estimator

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

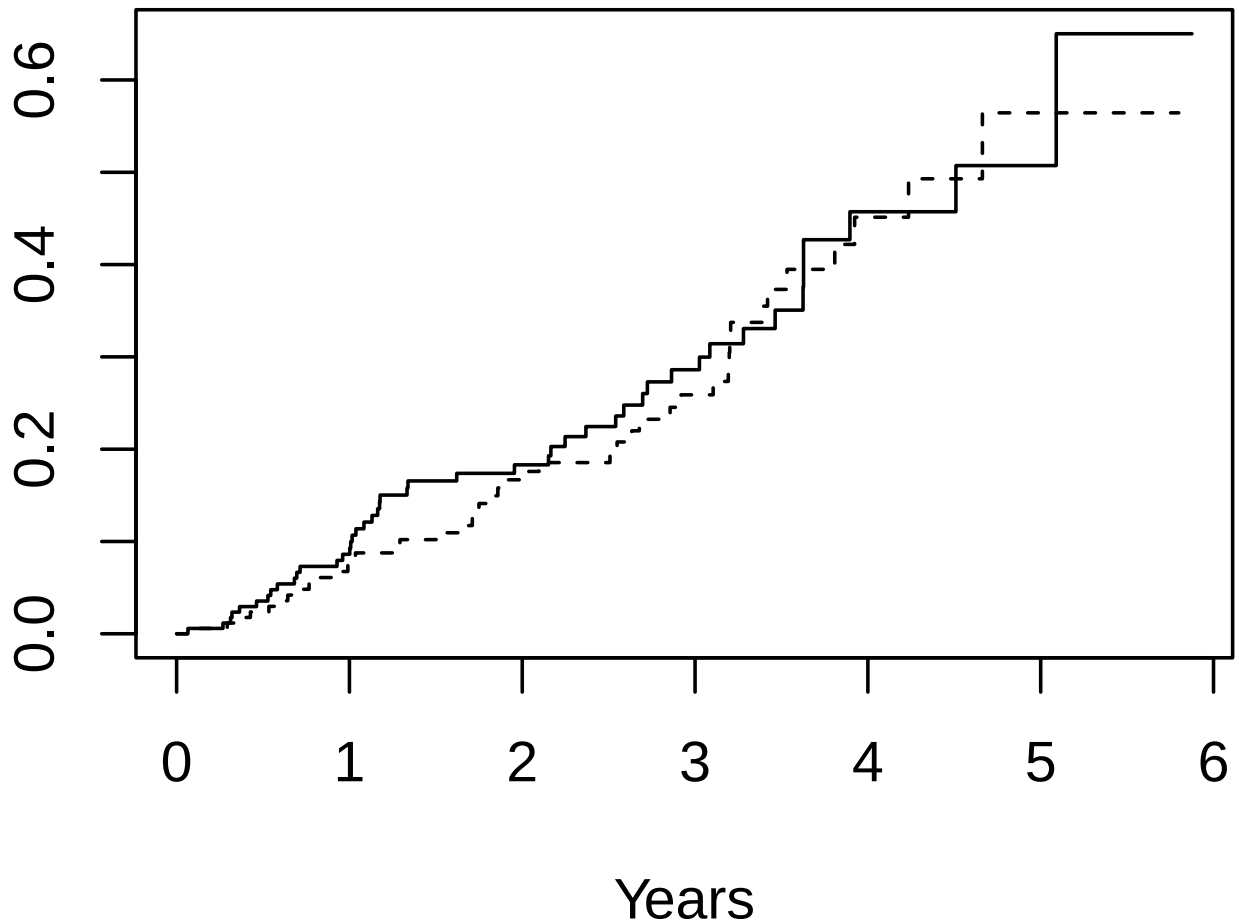
The Breslow estimator in R

```
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)
```

Breslow estimator `tment`



Nelson-Aalen estimator `tment`



Categorical covariates

Disease stage

$$\text{stage} = \begin{cases} 2 & \text{histological stage} = \text{I-II} \\ 3 & \text{histological stage} = \text{III} \\ 4 & \text{histological stage} = \text{IV} \end{cases}$$

Aim is to estimate the HR between the three groups. The Cox model will need two indicator functions, e.g.,

$$\text{stage3} = \begin{cases} 1 & \text{histological stage} = \text{III} \\ 0 & \text{otherwise} \end{cases}$$

and

$$\text{stage4} = \begin{cases} 1 & \text{histological stage} = \text{IV} \\ 0 & \text{otherwise} \end{cases}$$

The Cox model becomes

$$\alpha(t) \quad = \quad \alpha_0(t)\exp(\beta_1 \cdot \text{stage3} + \beta_2 \cdot \text{stage4}),$$

where index i is removed for readiness. The Cox model assumes

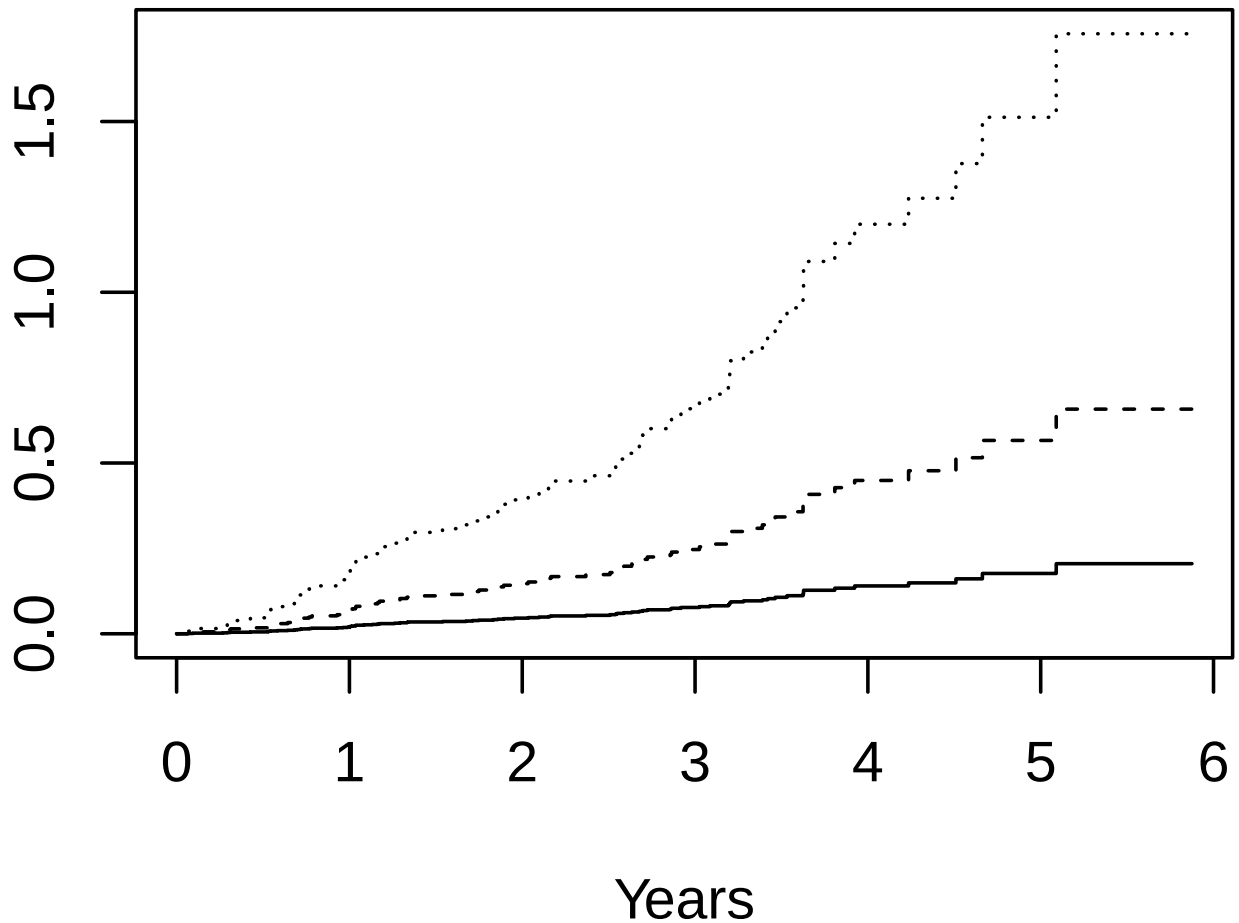
$$\alpha(t) = \begin{cases} \alpha_0(t) & \text{if histological stage} = \text{I-II} \\ \alpha_0(t)\exp(\beta_1) & \text{if histological stage} = \text{III} \\ \alpha_0(t)\exp(\beta_2) & \text{if histological stage} = \text{IV} \end{cases}$$

```
## categorical
fitstage<-coxph(Surv(days,status!=0)~factor(stage),
    data=pbc3, method="breslow")

plot(survfit(fitstage,
     newdata = data.frame(stage=factor(2:4))),
     cumhaz=T)
```

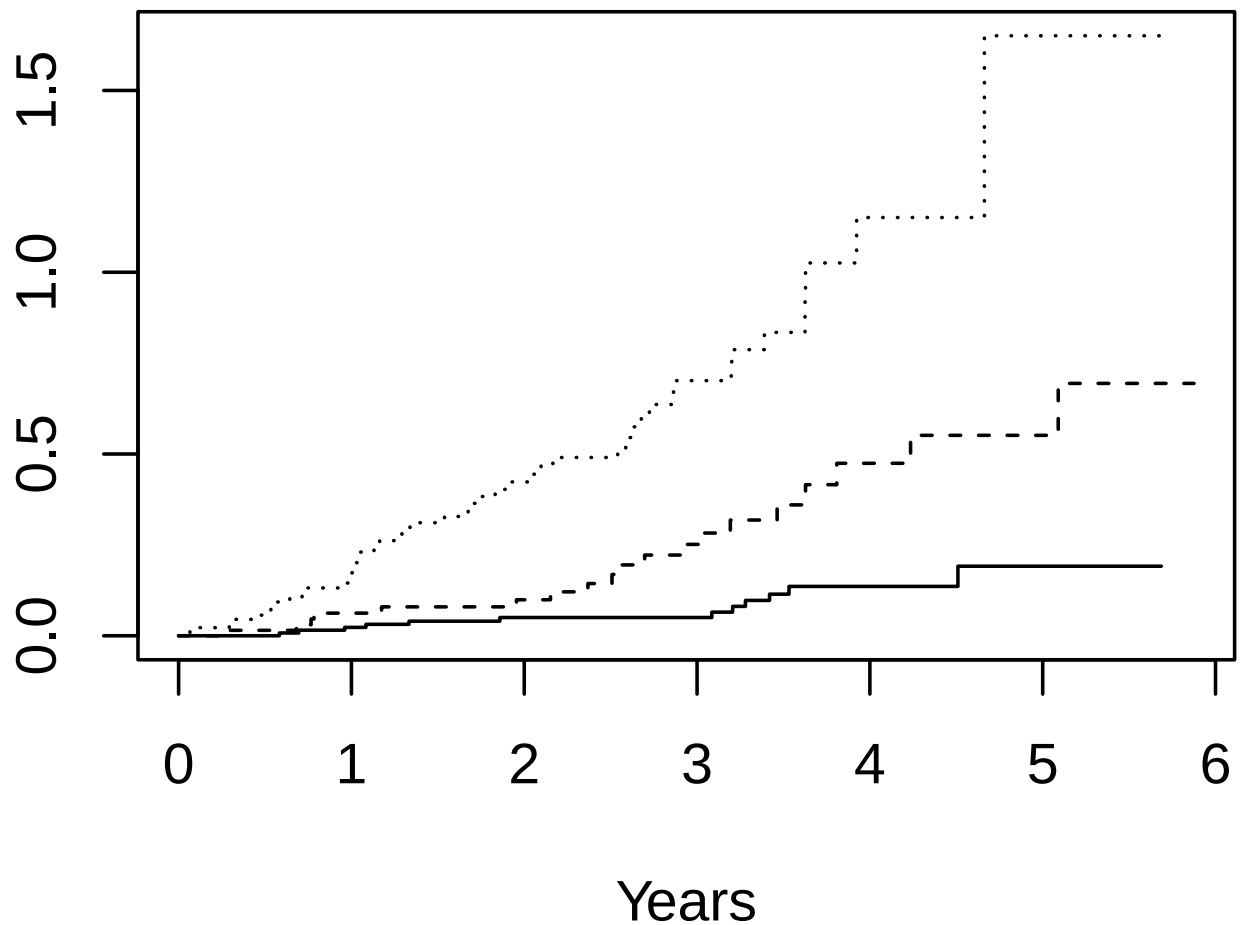
Cox in R

```
coxph(formula = Surv(days, status != 0) ~ factor(stage),
    data = pbc3, method = "breslow")

  n= 291, number of events= 77
   (58 observations deleted due to missingness)

                   coef exp(coef) se(coef)     z Pr(>|z|)
factor(stage)3 1.1644    3.2041   0.3698 3.149  0.00164 **
factor(stage)4 2.1469    8.5583   0.3265 6.575 4.86e-11 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

                exp(coef) exp(-coef) lower .95 upper .95
factor(stage)3     3.204     0.3121     1.552     6.614
factor(stage)4     8.558     0.1168     4.513    16.230

Likelihood ratio test= 56.61  on 2 df,   p=5e-13
Wald test            = 46.91  on 2 df,   p=7e-11
Score (logrank) test = 61.95  on 2 df,   p=4e-14
```

Breslow estimator `stage`



Nelson-Aalen estimator `stage`



Quantitative covariates

Using albumin measured in g/L (variable `alb`) from the PBC trial:

$$\alpha(t) \quad = \quad \alpha_0(t) \exp(\beta \cdot \text{alb})$$

or on the log-rate scale

$$\log(\alpha(t)) \quad = \quad \log(\alpha_0(t)) + \beta \cdot \text{alb}.$$

For all t this is a straight line with intercept $\log(\alpha_0(t))$ and slope β. The log-rate increases (or decreases) with β for each unit increase in albumin.

The null hypothesis $\beta = 0$, is a hypothesis of no effect of albumin (slope=0).

Having the model

$$\alpha(t) \quad = \quad \alpha_0(t)\exp(\beta \cdot \text{alb})$$

and comparing two patients with a albumin difference of 1 g/L. Let the reference have albumin $= Z$ g/L:

$$HR = \frac{\alpha_0(t)\exp(\beta \cdot (Z+1))}{\alpha_0(t)\exp(\beta \cdot Z)} = \exp(\beta).$$

Comparing two patients with a albumin difference of 10 g/L. Let the reference have albumin $= Z$ g/L:

$$HR = \frac{\alpha_0(t)\exp(\beta \cdot (Z+10))}{\alpha_0(t)\exp(\beta \cdot Z)} = \exp(\beta \cdot 10) = \exp(\beta)^{10}$$

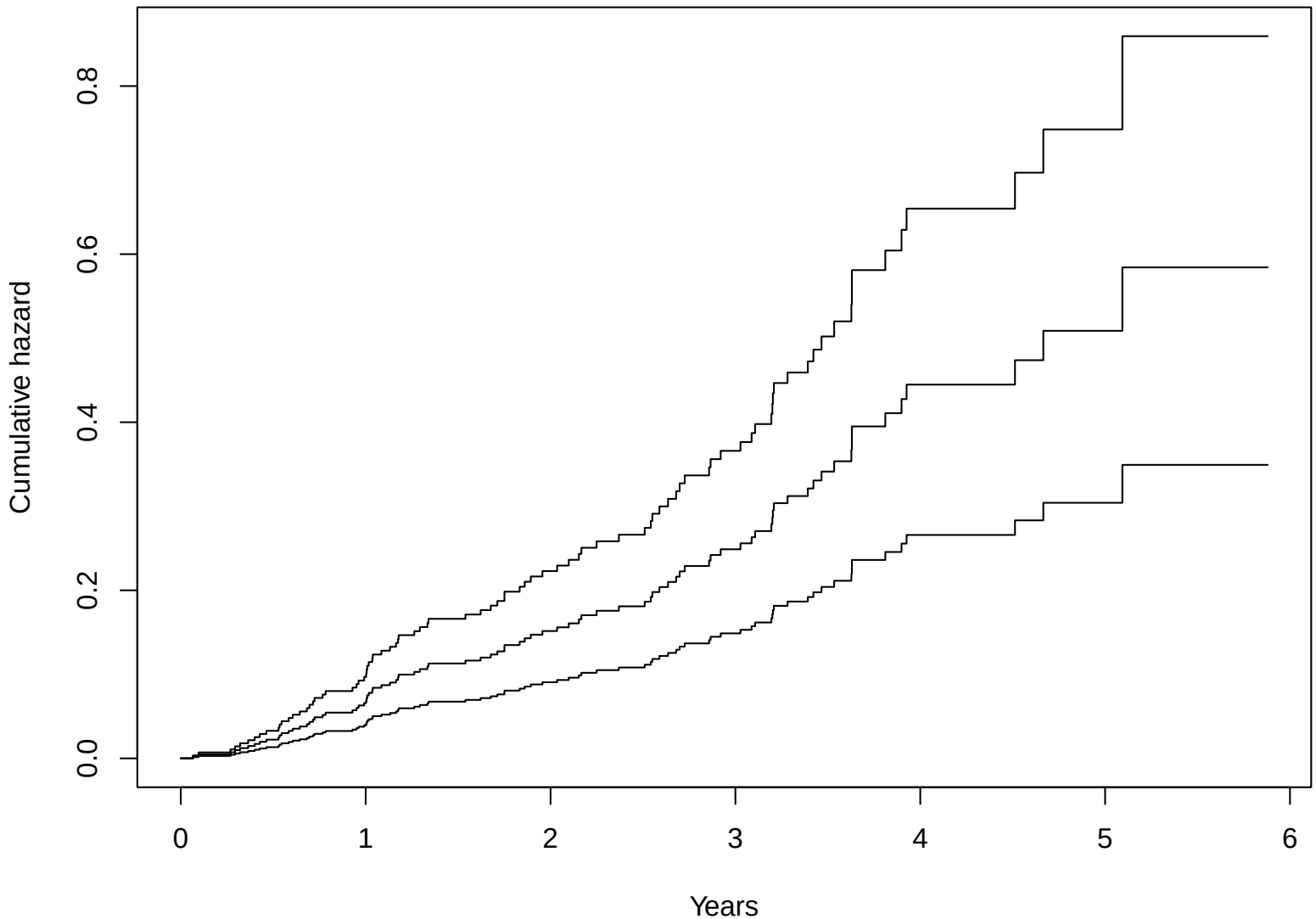
Cox in R

```
## quantitative
fitalb<-coxph(Surv(days,status!=0)~alb,
    data=pbc3, method="breslow")

## Breslow by quartile of alb
plot(survfit(fitalb,newdata = data.frame(alb=c(35,38,42))),
cumhaz = T,

## interaction
coxph(Surv(days,status!=0)~tment*factor(stage),
    data=pbc3, method="breslow")
```

Cox in R

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

  n= 343, number of events= 88
   (6 observations deleted due to missingness)

        coef exp(coef) se(coef)      z Pr(>|z|)
alb -0.12863   0.87930  0.02016 -6.382 1.75e-10 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

    exp(coef) exp(-coef) lower .95 upper .95
alb    0.8793      1.137    0.8452    0.9147

Concordance= 0.72  (se = 0.026 )
Likelihood ratio test= 40.13  on 1 df,   p=2e-10
Wald test            = 40.73  on 1 df,   p=2e-10
Score (logrank) test = 40.52  on 1 df,   p=2e-10
```

Breslow estimator `alb`



Multiple Cox regression

Table 2.4 *PBC3 trial in liver cirrhosis: Estimated coefficients (and SD) from a Cox model.*

Covariate		$\widehat{\beta}$	SD
Treatment	CyA vs. placebo	-0.496	0.226
Albumin	per 1 g/L	-0.116	0.021
Bilirubin	per 1 μmol/L	0.00895	0.00098

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)
```

Poisson regression (piece-wise exponential regression)

Alternatively to the Cox model, a piece-wise constant hazard model can be set up. For the PBC3 data, the model including only treatment is

$$\alpha_i(t) = \alpha_0(t)\exp(\beta Z_i),$$

but now the baseline hazard, instead of being completely unspecified as in the Cox model, is assumed to be constant in, e.g., 2-year intervals of follow-up time

$$\alpha_0(t) = \begin{cases} \alpha_1 & \text{if } t < 2, \\ \alpha_2 & \text{if } 2 \leq t < 4, \\ \alpha_3 & \text{if } 4 \leq t. \end{cases}$$

The resulting regression model is known as *Poisson* or *piece-wise exponential* regression.

Estimates of the parameters $\beta, \alpha_1, \alpha_2, \alpha_3$ are obtained by referring to the *maximum likelihood principle*.

Poisson regression (and Cox)

Table 2.4 *PBC3 trial in liver cirrhosis: Estimated coefficients (and SD) from a Cox model.*

Covariate		$\widehat{\beta}$	SD
Treatment	CyA vs. placebo	-0.496	0.226
Albumin	per 1 g/L	-0.116	0.021
Bilirubin	per 1 μmol/L	0.00895	0.00098

Table 2.5 *PBC3 trial in liver cirrhosis: Estimated coefficients (and SD) from a Poisson regression model.*

Covariate		$\widehat{\beta}$	SD
Treatment	CyA vs. placebo	-0.475	0.224
Albumin	per 1 g/L	-0.112	0.021
Bilirubin	per 1 μmol/L	0.00846	0.00094

Stratified Cox model

Stratified Cox model

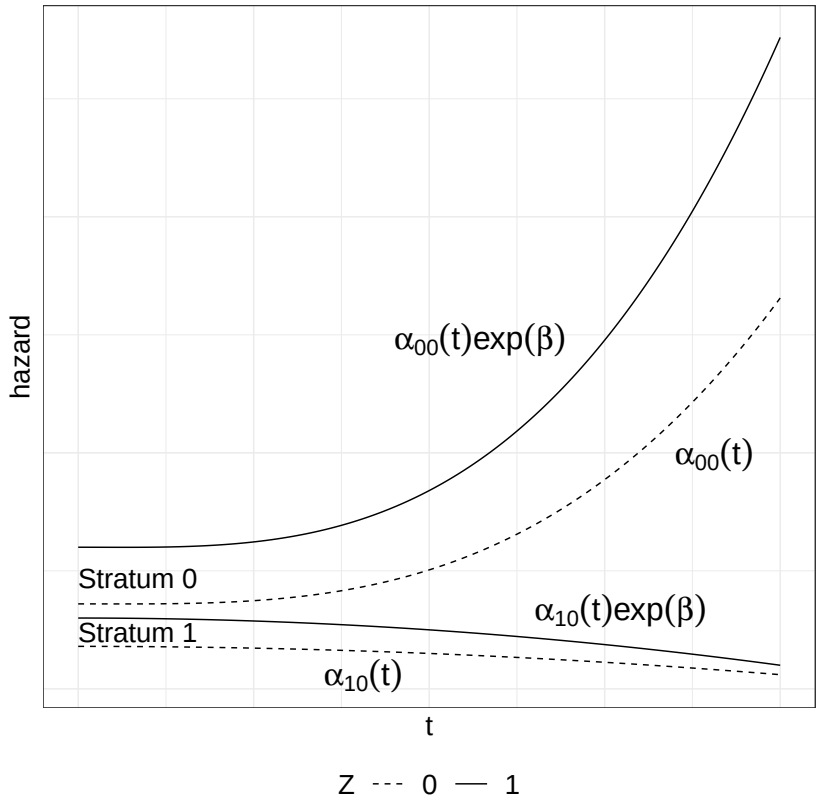
If the assumption of proportional hazards is questionable (more on control of this later) for a *categorical* covariate with, say, m categories it is possible to expand the Cox model to include different baseline hazards for each category

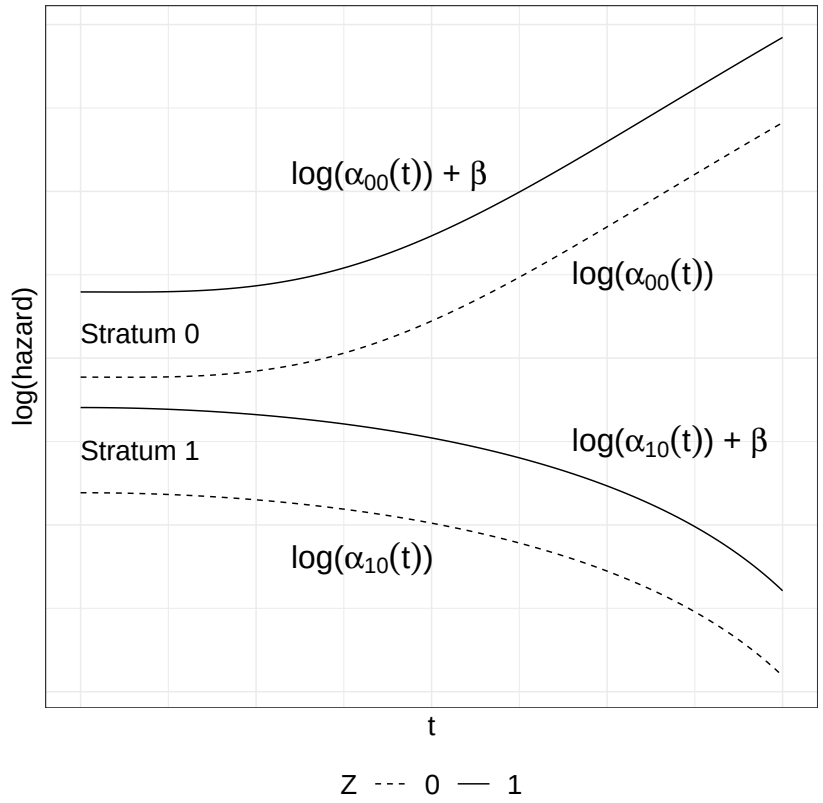
$$\alpha_i(t) \quad = \quad \alpha_{j0}(t)\exp(\beta^{\mathsf{T}} Z_i), \text{ when } i \text{ is in stratum } j$$

where $\alpha_{j0}(t)$ is the baseline hazard in stratum $j = 1, \ldots, m$. These baseline hazards are again allowed to depend on time, but no other assumptions are made.

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, i.e., no interaction between the stratifying variable and the covariates (can be relaxed).

Stratified Cox model





Likelihood Function for stratified model

For the stratified Cox model

$$\alpha(t) = \alpha_{j0}(t)\exp(\beta^{\mathsf{T}} Z), \quad j = 1, \ldots, m.$$

the partial likelihood is the product of the partial likelihood functions for the individual strata S_j

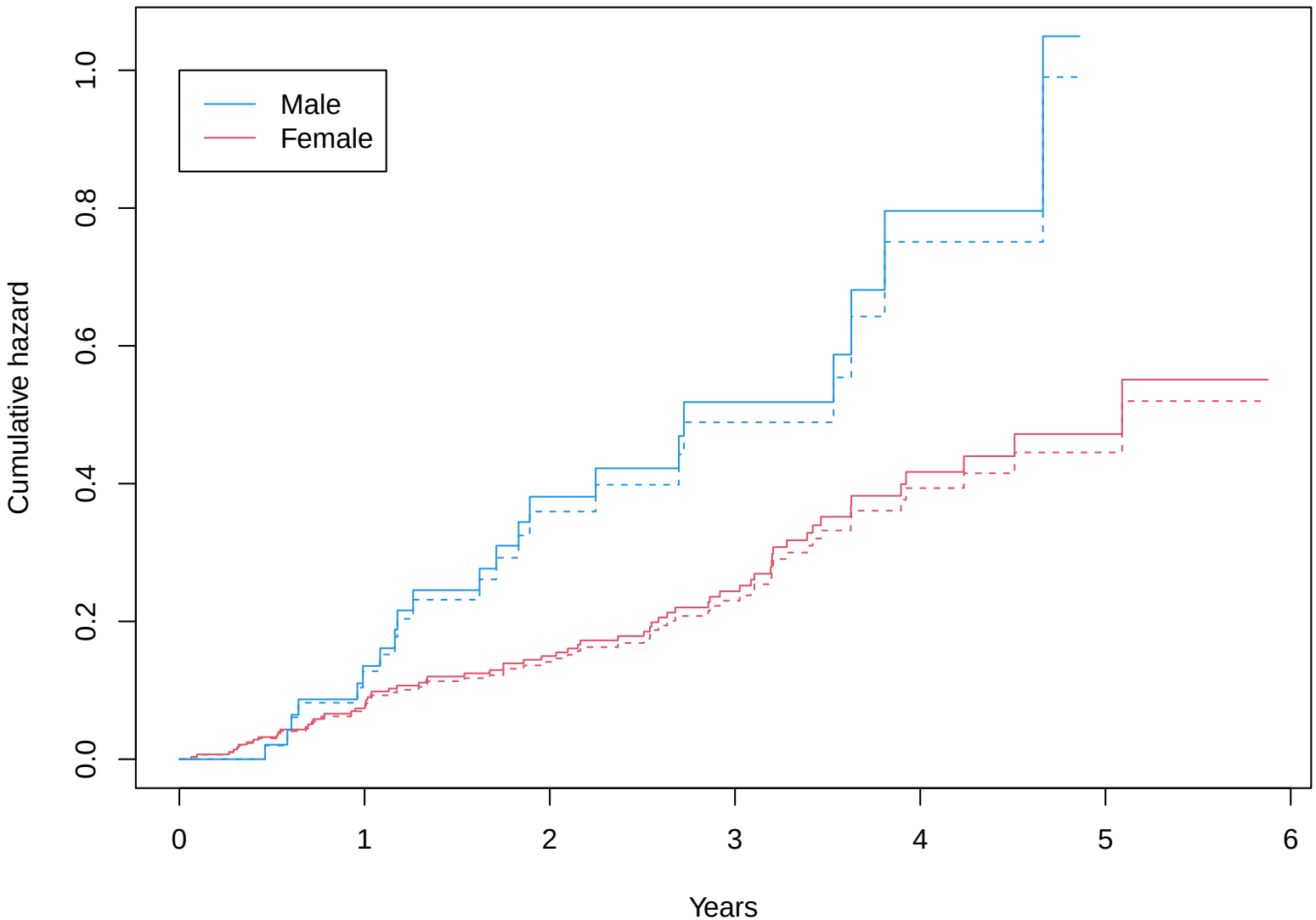
$$\begin{aligned} \mathrm{PL}_s(\boldsymbol{\beta}) &= \prod_j \prod_{i \in S_j} \prod_t \left(\frac{Y_i(t)\exp(\boldsymbol{\beta}^{\mathsf{T}} \boldsymbol{Z}_i)}{\sum_{k \in S_j} Y_k(t)\exp(\boldsymbol{\beta}^{\mathsf{T}} \boldsymbol{Z}_k)} \right)^{dN_i(t)} \\ &= \prod_j \prod_{i \in S_j} \left(\frac{\exp(\beta Z_i)}{\sum_{k \in R_j(X_i)} \exp(\beta Z_k)} \right)^{D_i}, \end{aligned}$$

where $R_j(t_i)$ is the risk set for each stratum S_j. Stratum-specific Breslow estimators for $A_{j0}(t) = \int_0^t \alpha_{j0}(u)du$ can be calculated.

Stratified Cox in R

```
## stratified model
fitstrata<-coxph(Surv(days,status!=0)~tment+strata(sex),
    data=pbc3, method="breslow")

## Breslow estimators
plot(survfit(fitstrata,newdata = data.frame(tment=0:1)),
     cumhaz = T)

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

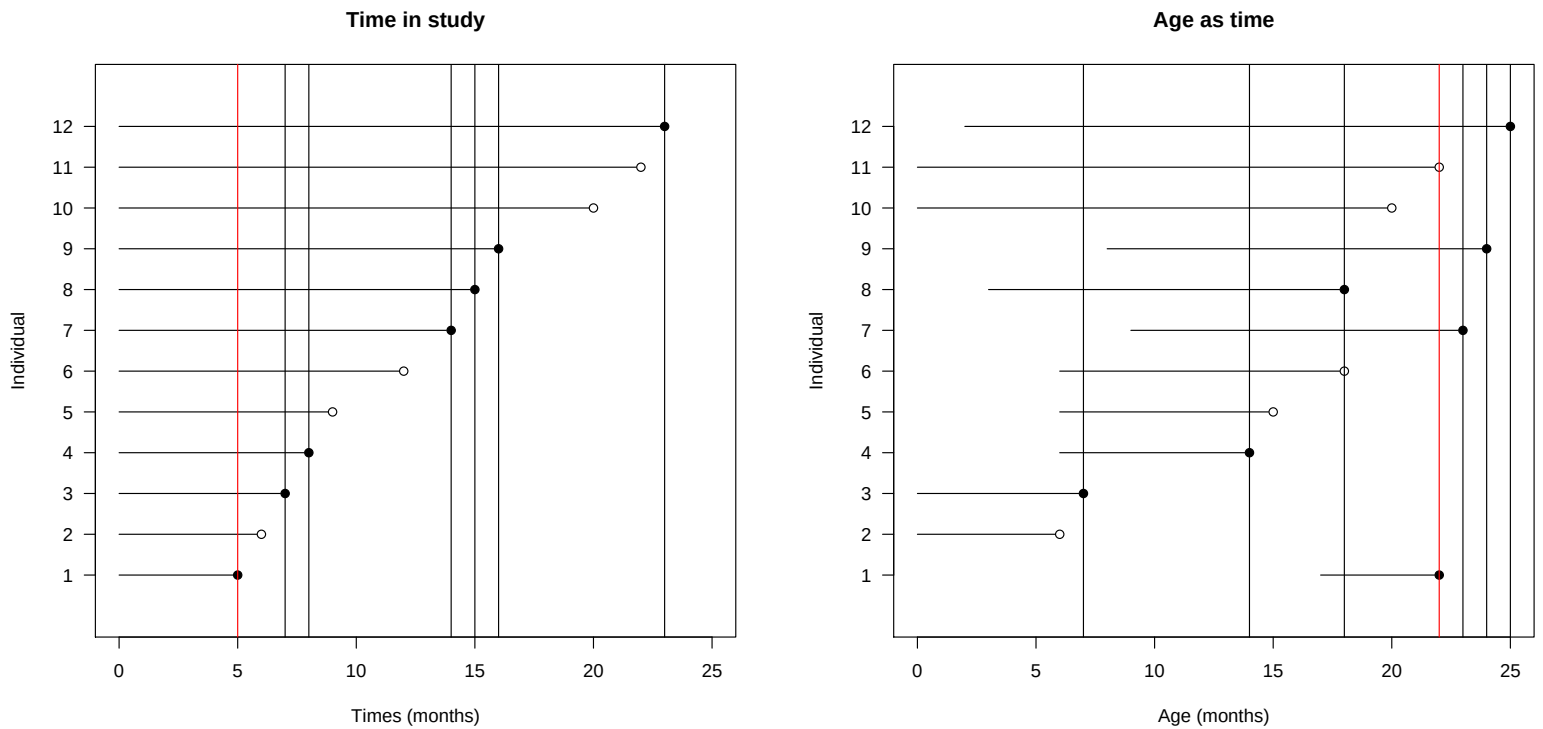
Cumulative hazards from stratified Cox model



Delayed entry

Delayed entry

Not often in randomized trial 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 is 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)
```

Time-dependent covariates

Time-dependent covariates

The Cox model may be expanded to include time-dependent covariates

$$\alpha_i(t) \quad = \quad \alpha_0(t) \exp(\beta^{\mathsf{T}} Z_i^*(t)).$$

Here $Z_i^*(t)$, is some summary of the covariate *history* $(Z(u); u \leq t)$, such as

- $Z_i^*(t) = Z_i(t)$, the value at time t
- $Z_i^*(t) = I(\text{vaccinated before } t)$
- $Z_i^*(t) = Z_i(0) \cdot f(t)$, baseline value times a known function

Estimation with time-dependent covariates

Cox's partial likelihood becomes

$$\mathrm{PL}(\beta) = \prod_i \left(\frac{\exp(\beta^\mathsf{T} Z_i^*(X_i))}{\sum_{j \in R(X_i)} \exp(\beta^\mathsf{T} Z_j^*(X_i))} \right)^{D_i}.$$

NB: $Z_j^*(t_i)$ should be known for all subjects at risk at *event times*.

Time-dependent covariates can be combined with stratified model and strata may also be time-dependent.

Interaction with time scale

Let Z be binary (treatment, placebo). An example of $Z_i^*(t) = Z_i(0) \cdot f(t)$ is the model

$$\alpha_i(t) \quad = \quad \alpha_0(t) \exp(\beta_1 Z_i + \beta_2 Z_i I(t \geq t_0)),$$

where

$$I(t \geq t_0) = \begin{cases} 0 & \text{if } t < t_0 \\ 1 & \text{if } t \geq t_0 \end{cases}$$

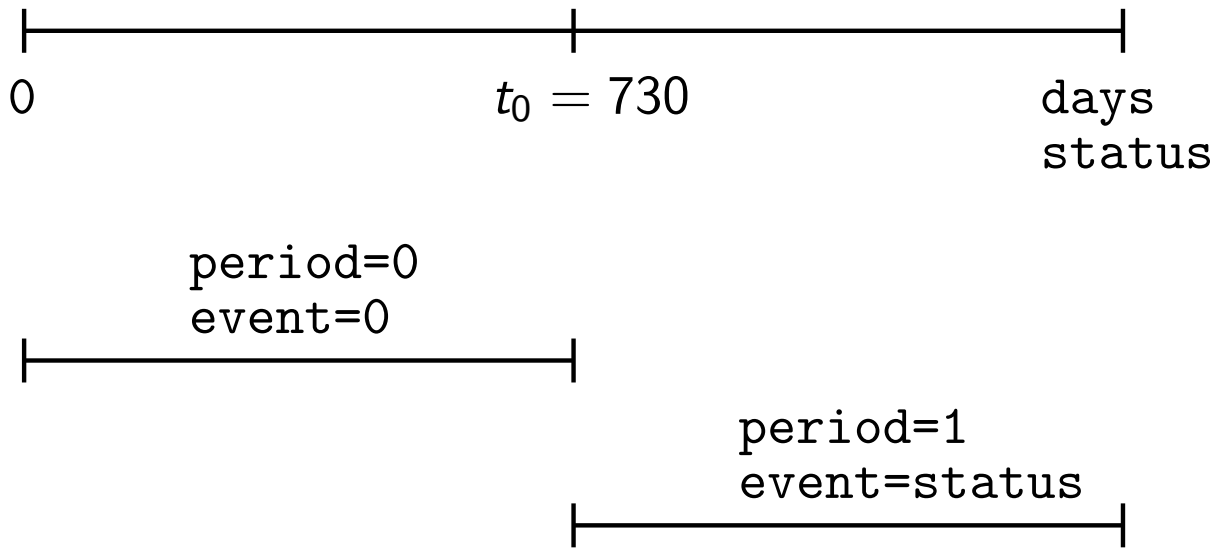
corresponding to an interaction between time and Z. The hazard ratio (treatment vs placebo) is then

$$HR = \begin{cases} \exp(\beta_1) & \text{if } t < t_0 \\ \exp(\beta_1 + \beta_2) & \text{if } t \geq t_0 \end{cases}$$

A (simple) test of PH-assumption is testing $\beta_2 = 0$.

Binary time-dependent covariate

Splitting up subjects in two records at followup $= t_0$ years (e.g., 2 years):



Create an expanded data set and use delayed entry (counting process style).

Time-dependent covariates in R

```
## Method I
coxph(Surv(days,status!= 0) ~ tment + tt(tment),
      data = pbc3,
      tt = function(x, t, ...) (x)*(t > 2*365))

## Method II; splitting
pbc3sp<-survSplit(Surv(days,status!=0) ~ ., data = pbc3,
                  cut=c(2*365), episode = "period")

coxph(Surv(tstart,days,event) ~ tment*factor(period),
      data = pbc3sp, method="breslow")
```

Goodness-of-fit I

Cox assumptions

- Linear predictor: The effects of covariates are additive and linear on the log-hazard scale
- Proportional hazards: The ratio of the hazard rates for two groups is constant over time (can be relaxed by stratified model)

Checking assumptions for the linear predictor

This is not different from any other model with a linear predictor (e.g., linear or logistic regression).

- No interaction between Z_{i1} and Z_{i2} can be tested by adding suitable interaction terms to the model.
- Linearity for a quantitative Z may be tested by adding, e.g., quadratic terms Z^2 or linear splines to the model. For chosen cut-points, say a_1, a_2, add

$$(Z - a_1)I(Z > a_1) \text{ and } (Z - a_2)I(Z > a_2)$$

  to a model that also includes Z. The dose-response relationship between Z and the linear predictor (the log(hazard) in the Cox model) is then a broken straight line and coefficients for the linear splines give the change in slope at each cut-point.
- Martingale residuals provide special techniques for the Cox model (later today)

Checking proportional hazards

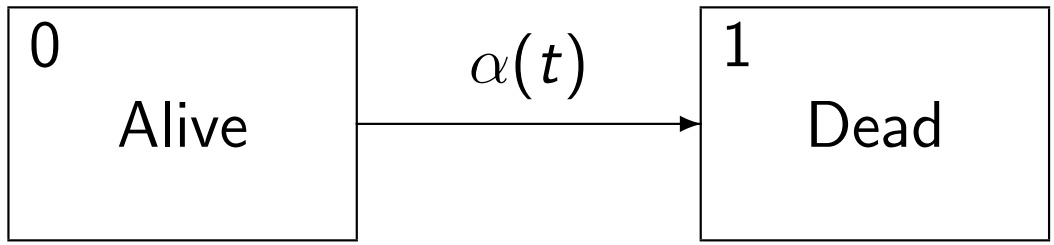
This is something special for the Cox model.

- 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 (later today)

Plug-in estimation of marginal parameters

Plug-in estimation of marginal parameters (predictions)

In the two-state model



we have the one-to-one relationship between survival and hazard

$$P(\text{State 0 at time } t) = S(t) = \exp(-\int_0^t \alpha(u)du) \quad \text{Eq. (1.2)}$$

A regression model for $\alpha(t)$ induces a regression model for $S(t)$.

For a Cox model

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

the survival function is then given by

$$S(t \mid Z) = \exp\Big(-\int_0^t \alpha_0(u) \exp(\beta^{\mathsf{T}} Z) du\Big) = \exp\big(-A_0(t) \exp(\beta^{\mathsf{T}} Z)\big)$$

The complementary log-log link transformation *cloglog* of a distribution function $F(t)$ is

$$\text{cloglog}(F(t)) = \log(-\log(1 - F(t))) = \log(-\log(S(t))), \text{ i.e.,}$$

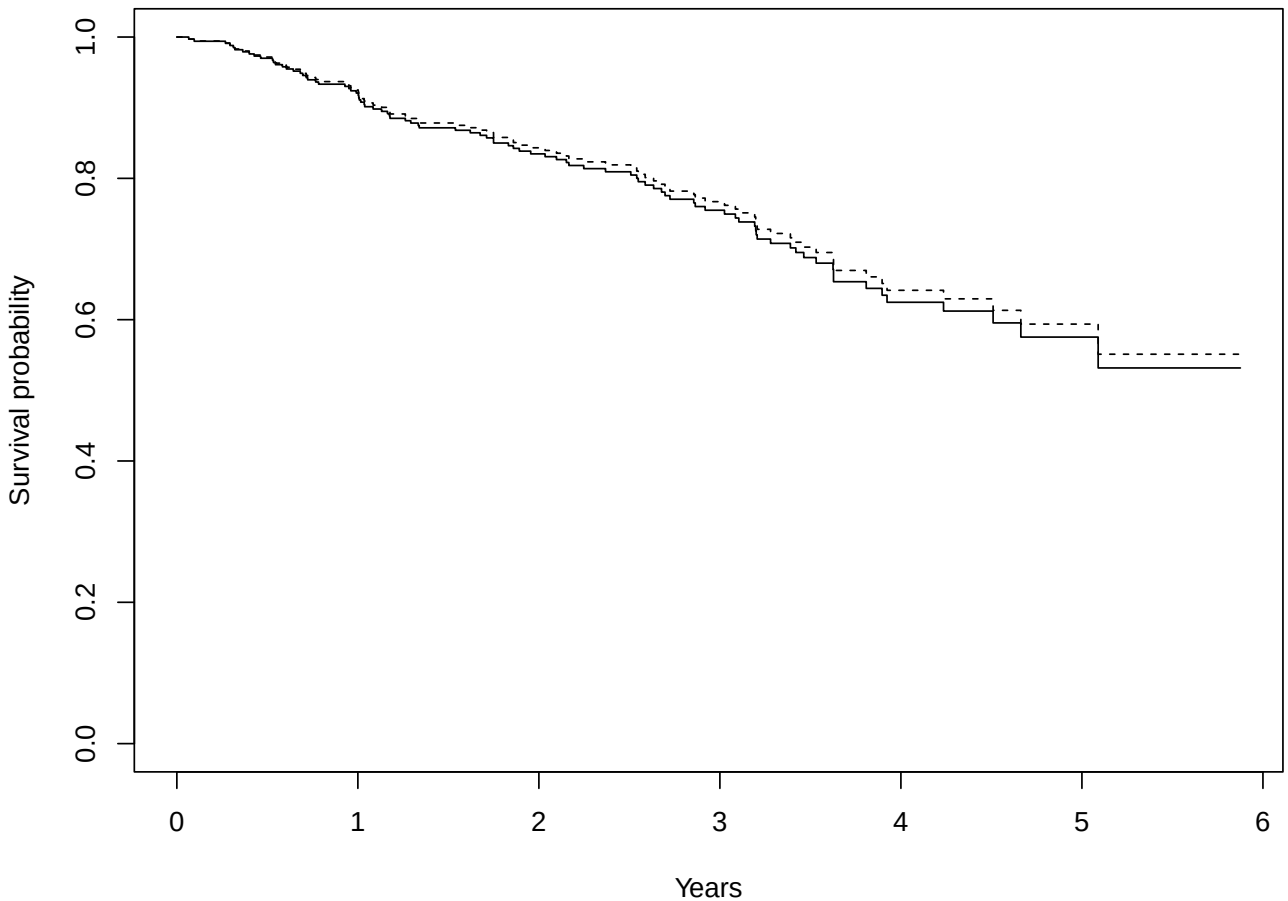
$$\log(-\log((S \mid Z) = \log(A_0(t)) + \beta^{\mathsf{T}} Z. \quad \text{Eq. (4.6)}$$

The cloglog link function takes us from the marginal parameter to the linear predictor. One typically uses Breslow estimator for $A_0(t)$.

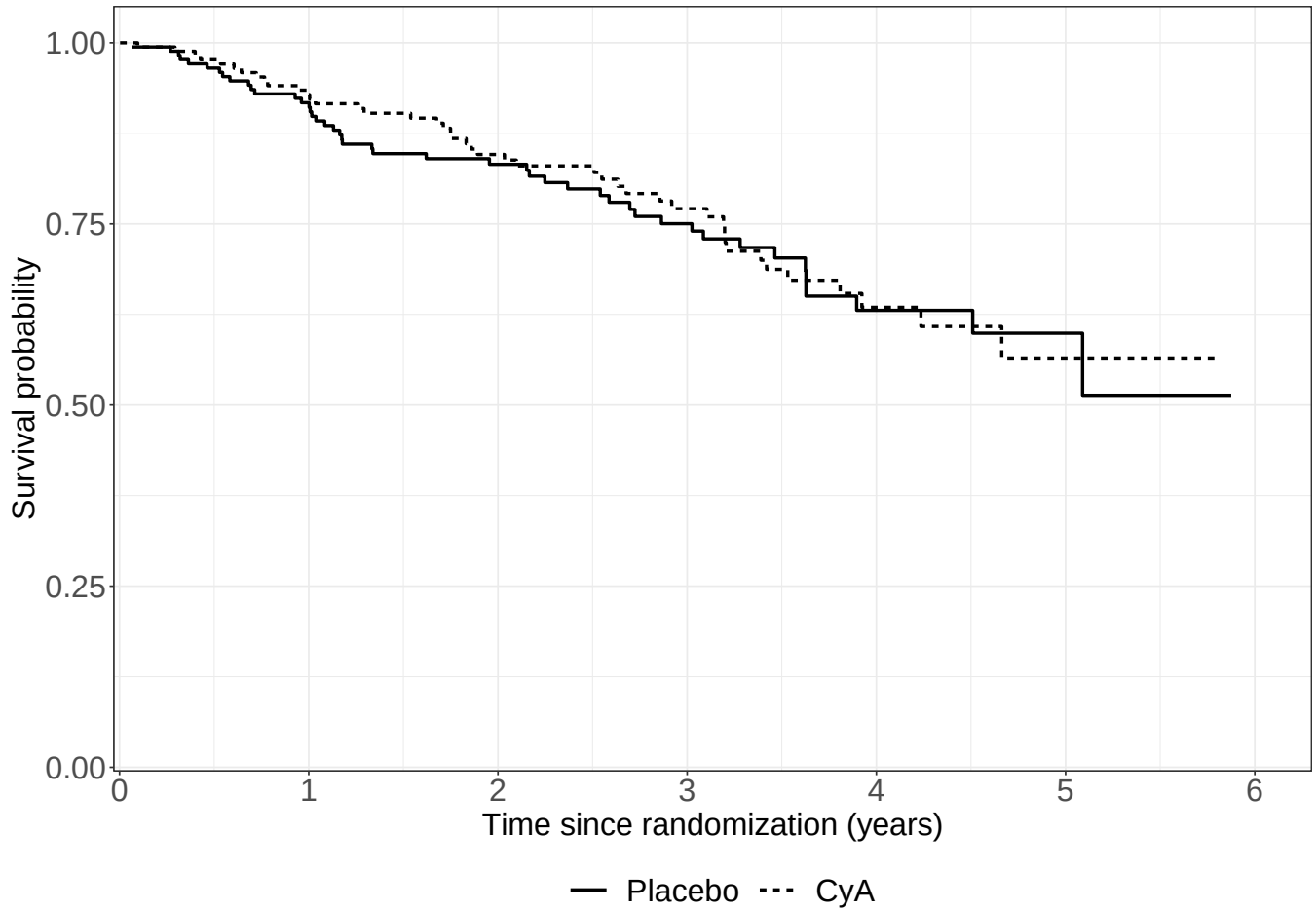
Predict after Cox in R

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

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

Prediction from Cox model



Kaplan-Meier



Multiple Cox regression

```
pbc3$log2bili <- with(pbc3, log2(bili))
coxfit<-coxph(Surv(days,status!=0)~ tment + alb + log2bili,
              data=pbc3, method="breslow")
```

```
              coef exp(coef) se(coef)      z Pr(>|z|)
tment     -0.57406   0.56323  0.22447 -2.557   0.0105
alb       -0.09093   0.91308  0.02164 -4.201 2.65e-05
log2bili   0.66500   1.94449  0.07443  8.935  < 2e-16
---

          exp(coef) exp(-coef) lower .95 upper .95
tment        0.5632     1.7755    0.3628    0.8745
alb          0.9131     1.0952    0.8752    0.9526
log2bili     1.9445     0.5143    1.6805    2.2499
```

Predict after Cox in R

```
pbc3$log2bili <- with(pbc3, log2(bili))

coxfit<-coxph(Surv(days,status!=0)~tment+alb+log2bili,
              data=pbc3, method="breslow")

preddata<-data.frame(tment=0:1,alb=38,log2bili=log2(45))

plot(survfit(coxfit, newdata = preddata),lty=1:2)
```

Prediction from Cox model

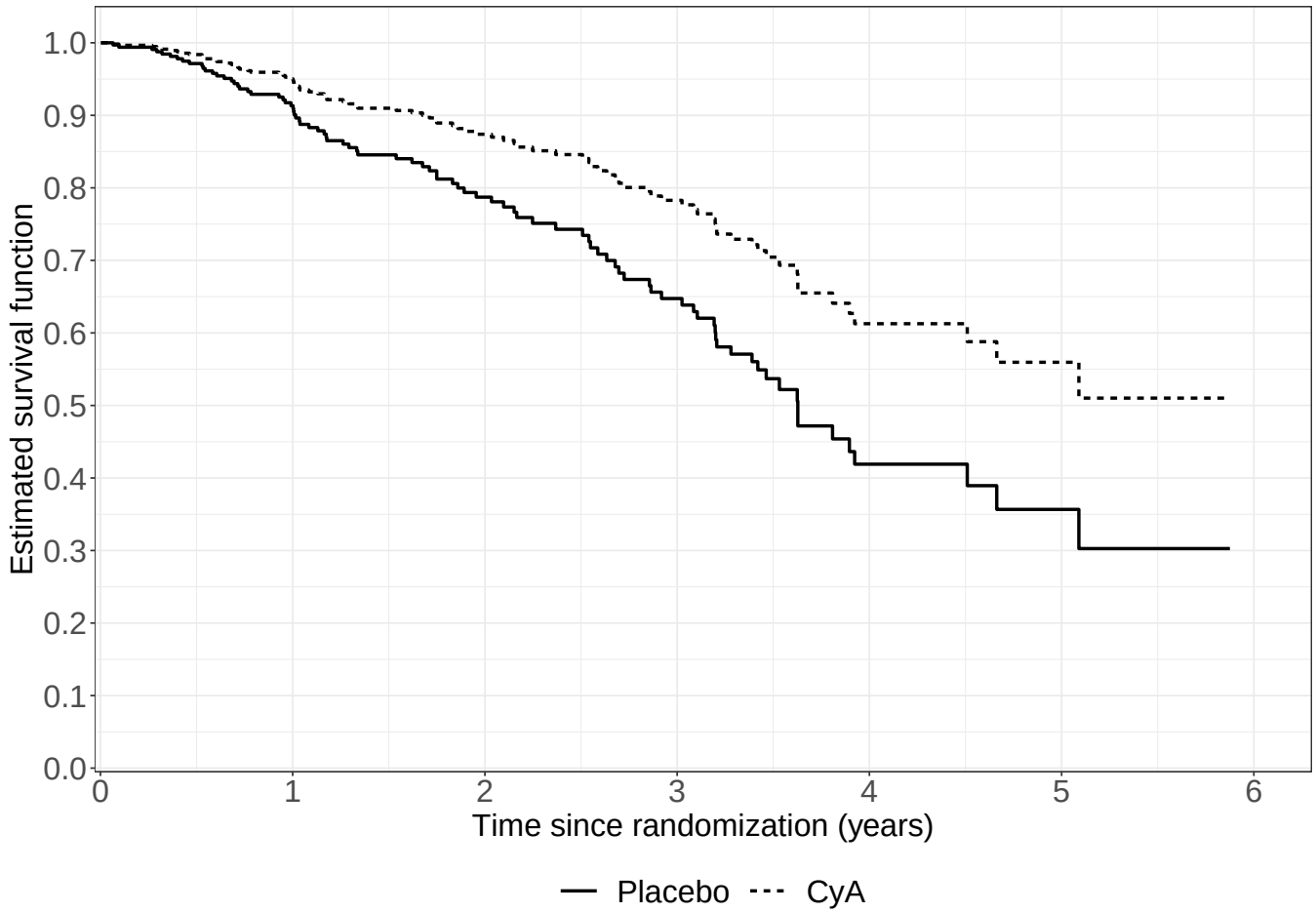


Figure 4.5: Estimated survival curves for a patient with albumin = 38 g/L and bilirubin = 45 μmol/L based on a Cox model.

Predicted risk difference and RMST

Based on a Cox model, it is thus possible to obtain a 'regression model' for survival (or risk) at a favorite time τ

$$\widehat{S}(\tau \mid Z),$$

and one could then, e.g., for treatment, estimate the risk difference

$$\widehat{RD}(\tau \mid Z) = \widehat{S}_0(\tau \mid Z) - \widehat{S}_1(\tau \mid Z).$$

Similarly, for the τ-*restricted mean survival time* (RMST)

$$\widehat{\varepsilon}_0(\tau \mid Z) = \int_0^\tau \widehat{S}(t \mid Z)dt.$$

One does, however, not get a simple direct relationship between the variables and the marginal parameters.

Direct adjusted survival curves (g-formula)

From a Cox model with treatment variable Z^0 and other covariates Z, the risk difference at τ between treatment groups could be estimated by direct adjustment/standardization:

$$\frac{1}{n}\Big(\sum_i \widehat{S}(\tau \mid Z_i^0 = 1, Z_i) - \widehat{S}(\tau \mid Z_i^0 = 0, Z_i)\Big)$$

This is also known as the g-formula in (modern) causal inference. It is summarizing the survival experience of an average patient for a *given population*.

Direct adjusted survival curves (g-formula)

Algorithm:

1. The value of the variable `tment` is set to 0 for all observations in the PBC data set.
2. The survival curve for each observation in the modified data set is computed.
3. All the survival curves computed in step 2 are averaged.

And similar for `tment` set to 1.

Direct adjusted survival curves (g-formula)

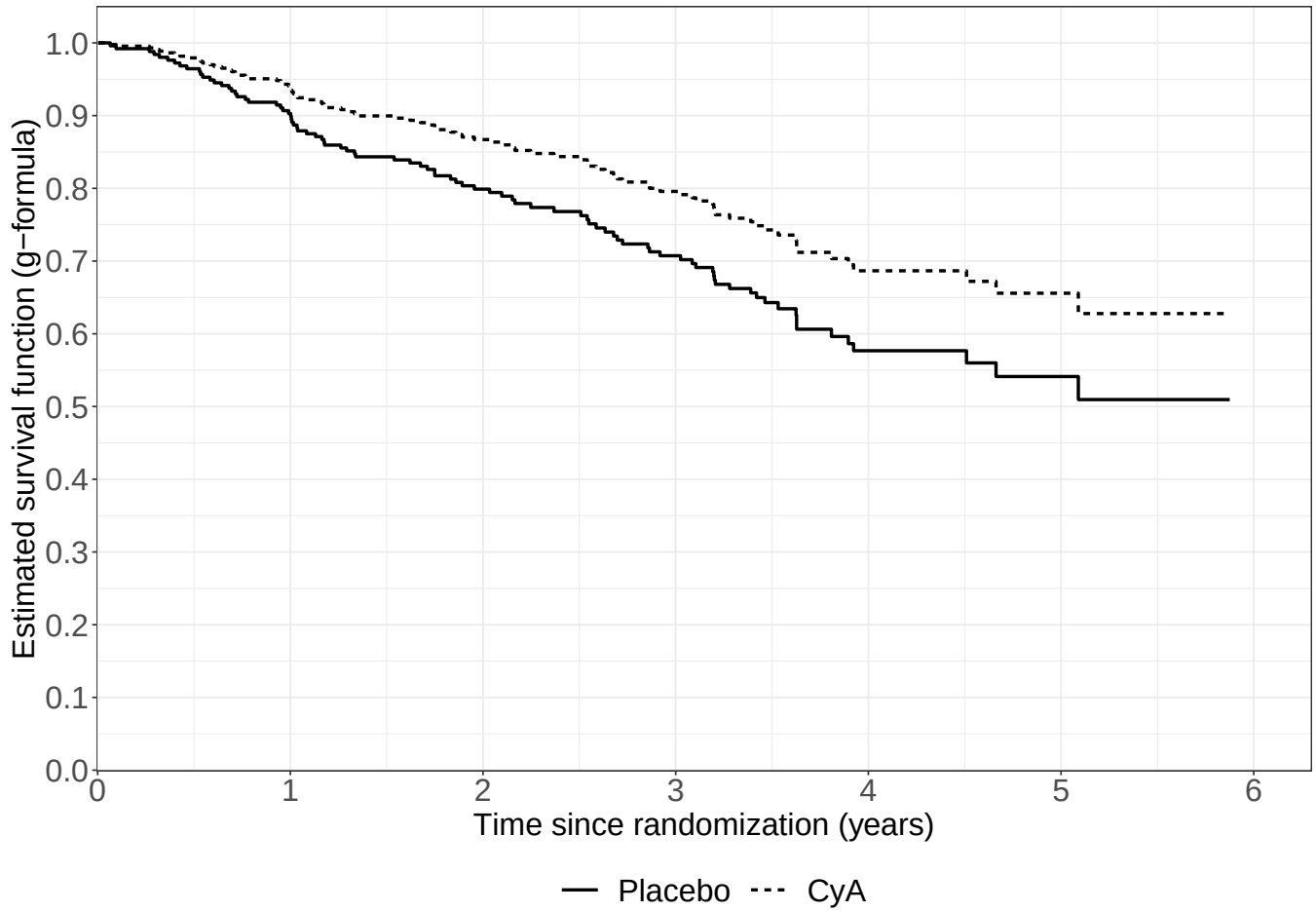


Figure 4.7

Prediction and time-dependent covariates

Without time-dependent covariates (and without competing risks) the survival function at time t is:

$$S(t \mid Z) = [\exp(-A_0(t))]^{\exp(\beta^\mathsf{T} Z)}$$

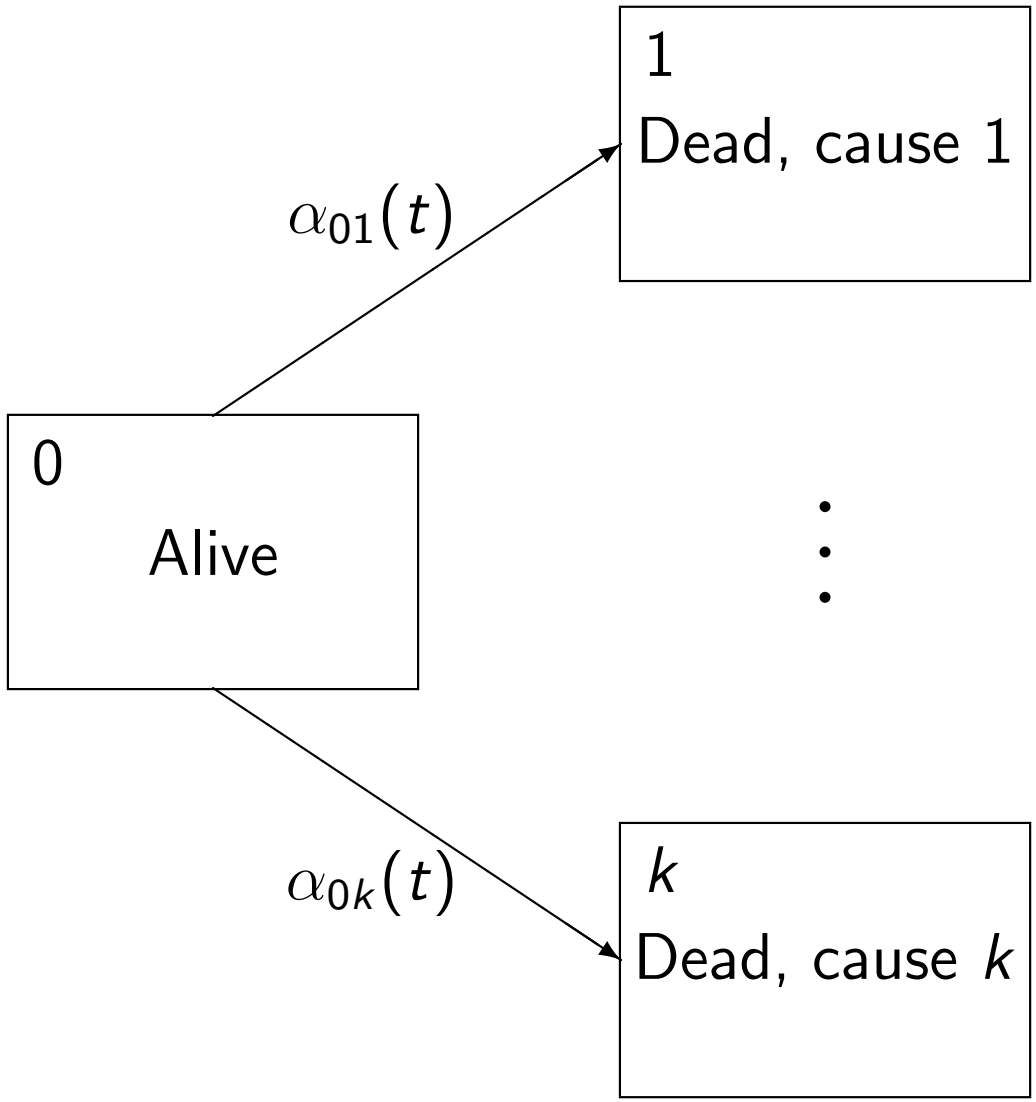
When there are time-dependent covariates in the hazard model then this need not work since the probability of surviving till time t depends on the (stochastic) behavior of $(Z(u); 0 \leq u < t)$.

Covariates must be *exogenous* (external) for the prediction to work.

With *endogenous* (internal) covariates, there is extra randomness not described by the hazard function and there is a need for joint models.

Cox regression for competing risks

Competing risks



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):

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

Competing risks – Likelihood (Sec 3.1)

Data: (X_i, D_i), $\quad i = 1, ..., n$ where $X_1, ..., X_n$ are observation times and $D_i = h$, $h = 1, ..., k$ if observed failure from cause h, $D_i = 0$ if censored.

Likelihood:

$$
\begin{aligned}
L &= \prod_{i=1}^{n} S(X_i) \prod_{h=1}^{k} (\alpha_h(X_i))^{I(D_i=h)} \\
&= \prod_{i=1}^{n} \left(\exp(-\sum_{h=1}^{k} A_h(X_i))\right) \prod_{h=1}^{k} (\alpha_h(X_i))^{I(D_i=h)} \\
&= \prod_{h=1}^{k} \left(\prod_{i=1}^{n} \exp(-A_h(X_i))(\alpha_h(X_i))^{I(D_i=h)}\right).
\end{aligned}
$$

Inference for cause-specific hazards

- Product over causes, h,
- The hth factor is what we would get if only that cause were studied *and all other causes were right-censorings*
- This has nothing to do with 'independence' of causes - it is solely a consequence of the definition of cause-specific hazards as hazards of exclusive events.
- It means that all standard hazard-based models for survival data apply when analyzing cause-specific hazards
  - non-parametric: estimate $A_h(t) = \int_0^t \alpha_h(u)\mathrm{d}u$, $h = 1, \ldots, k$ by Nelson-Aalen estimator, compare using, e.g., logrank tests
  - Cox regression (Poisson regression)
  - Breslow estimator for cumulative cause-specific hazards

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.

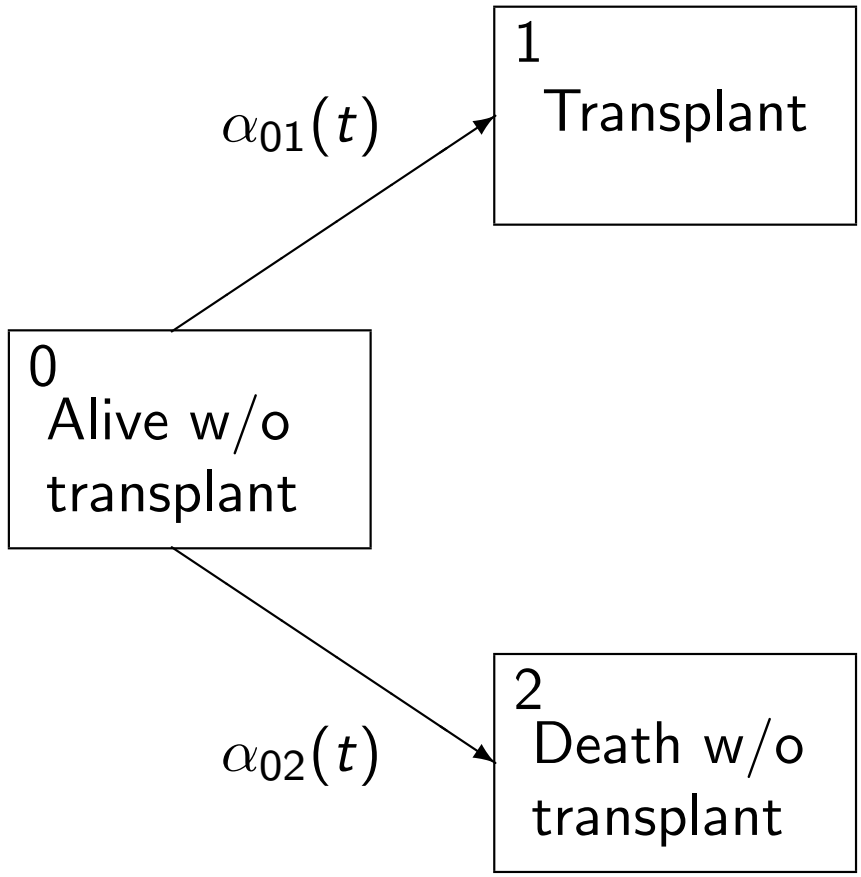
It is technically possible to fit Cox models for cause-specific hazards with

- identical or proportional baselines for some causes
- regression coefficients that are shared between several causes

However, that is rarely relevant.

These features may be more relevant for other multi-state models than the competing risks model.

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)
```

Breslow estimator for cumulative cause-specific hazard in R

```
fitcr<-coxph(Surv(days,factor(status))~tment+alb+log2bili+sex+age,
      method = "breslow", data = pbc3, id=id)

predcr<-data.frame(tment=0:1,alb=38,log2bili=log2(45),sex=0,age=55)

plot(survfit(fitcr,newdata=predcr), cumhaz = T)
```

Plug-in estimation of marginal parameters

Estimate cumulative incidence $F_h(t \mid Z)$ by plug-in:

$$\widehat{F}_h(t \mid Z) = \int_0^t \widehat{S}(u- \mid Z) d\widehat{A}_h(u \mid Z).$$

Here,

$$\widehat{A}_h(u \mid Z) = \widehat{A}_{h0}(u) \exp(\widehat{\beta}_{h1} Z_1 + ... + \widehat{\beta}_{hp} Z_p)$$

is the cumulative cause-h-hazard estimate from the Cox model and $\widehat{S}(u \mid Z)$ the Cox model based estimator for the overall survival function, e.g.,

$$\widehat{S}(u \mid Z) = \exp\left(- \sum_h \widehat{A}_h(u \mid Z) \right).$$

Additionally, plug-in estimation for risk difference and years lost can be obtained.

Cumulative incidences from cause-specific hazards

Important to notice:

- The Cox models impose a simple structure between covariates and *rates*.
- Due to the non-linear relationship between rates and risks, this simple relationship does not carry over to the cumulative incidences.
- In particular, the way in which a covariate affects a rate can be different from the way in which it affects the corresponding risk: this will depend on how it affects the rates for the competing causes.
- ... more in Part II

Cumulative incidences from cause-specific hazards

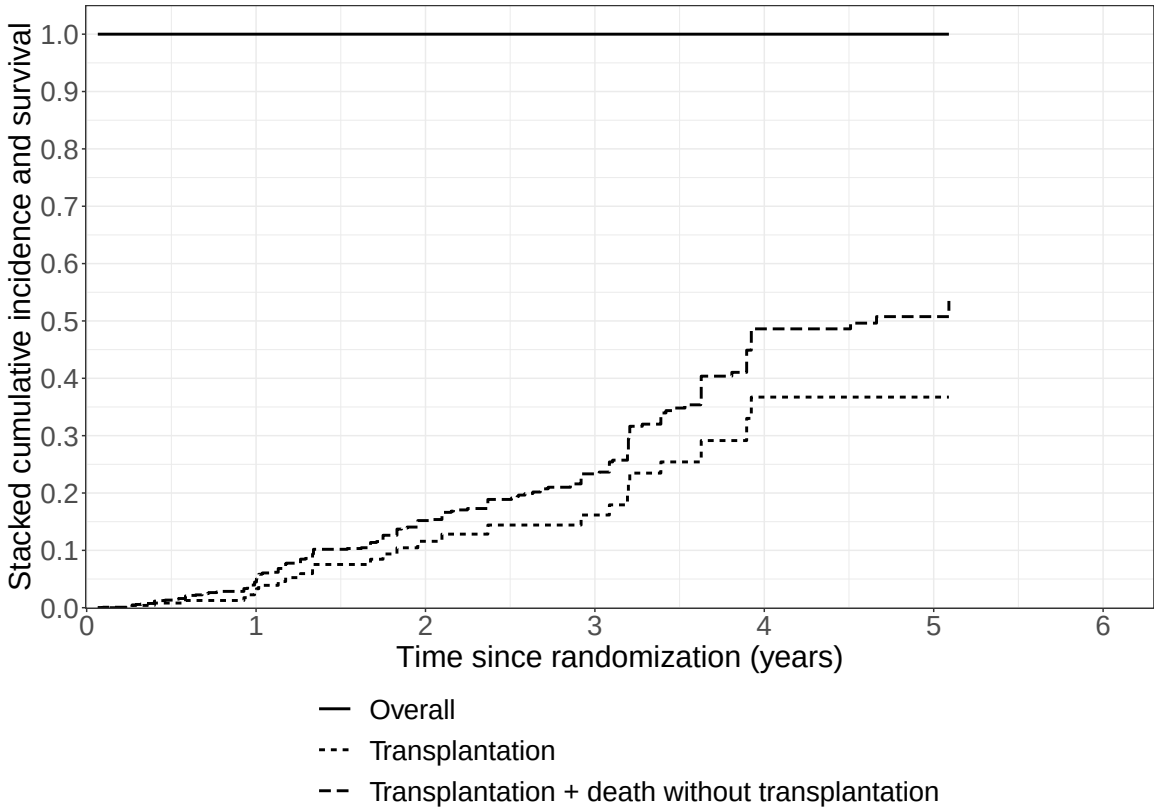


Figure 4.12a Predicted, stacked cumulative incidence and survival curves for a woman in the placebo group, 40 years old, alb=38, bili=45, based on cause-specific Cox models.

Cumulative incidences from cause-specific hazards

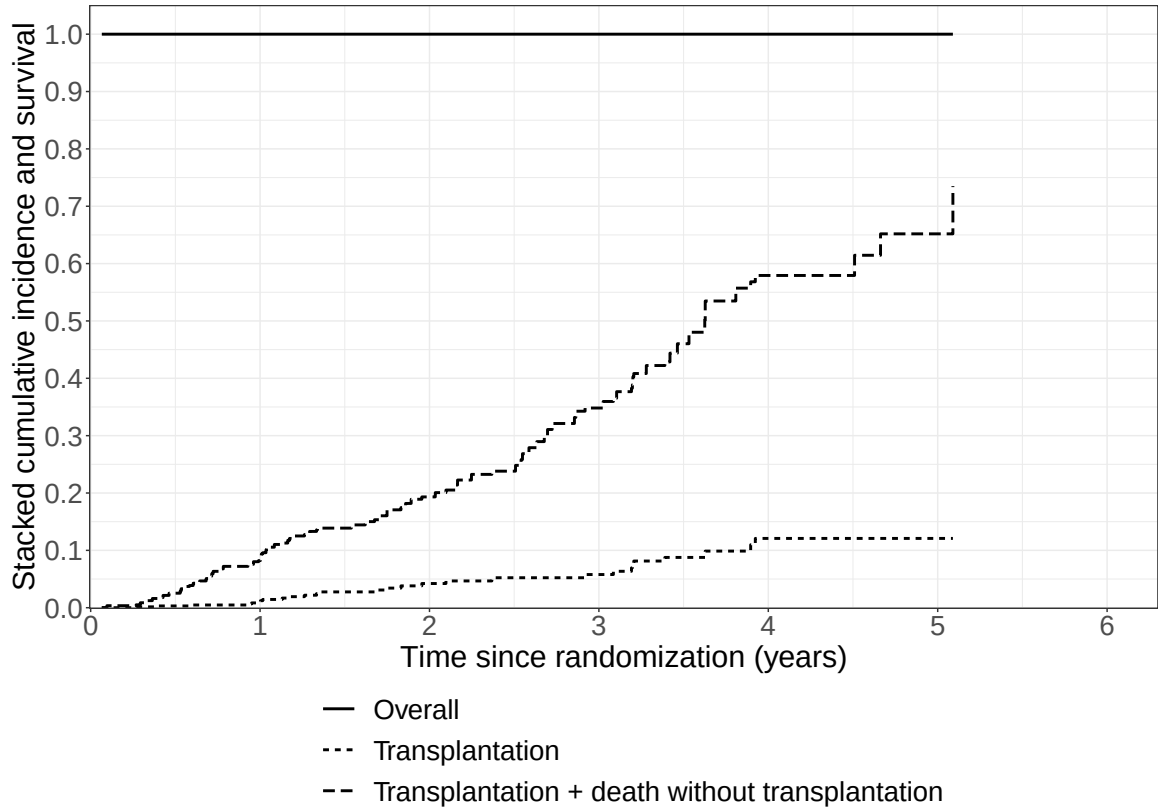


Figure 4.12c Predicted, stacked cumulative incidence and survival curves for a woman in the placebo group, 60 years old, alb=38, bili=45, based on cause-specific Cox models.

Cumulative incidences from cause-specific hazards in R

```
plot(survfit(fitcr,newdata=predcr))
```

Fine-Gray regression model?

Fine-Gray regression model?

It is not an intensity-based model!

... more in Part II

Goodness-of-fit

Sections 2.2.2, 5.7, 5.8.4

Cox assumptions

- Linear predictor: The effects of covariates are additive and linear on the log-hazard scale
- Proportional hazards: The ratio of the hazard rates for two groups is constant over time (can be relaxed by stratified model)

Checking assumptions for the linear predictor

This is not different from any other model with a linear predictor (e.g., linear or logistic regression).

- No interaction between Z_{i1} and Z_{i2} can be tested by adding suitable interaction terms to the model.
- Linearity for a quantitative Z may be tested by adding, e.g., quadratic terms Z^2 or linear splines to the model. For chosen cut-points, say a_1, a_2, add

$$(Z - a_1)I(Z > a_1) \text{ and } (Z - a_2)I(Z > a_2)$$

  to a model that also includes Z. The dose-response relationship between Z and the linear predictor (the log(hazard) in the Cox model) is then a broken straight line and coefficients for the linear splines give the change in slope at each cut-point.
- Martingale residuals provide special techniques for the Cox model.

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

Martingale residuals

Consider a given event (no cause j in notation though the event could be a given cause of failure). Recall that the counting process for subject i is

$$N_i(t) = I(X_i \leq t, D_i = 1)$$

and counts $+1$ at the observed time of failure. Note that $N_i(\infty) = D_i$. Let $Y_i(t) = I(X_i \geq t)$ be the *at-risk indicator* for subject i. Then the *martingale residual* (process) is

$$M_i(t) = N_i(t) - \int_0^t Y_i(u) \exp(\beta^{\mathsf{T}} Z_i) \alpha_0(u) du$$

which is 'estimated' as:

$$\widehat{M}_i(t) = N_i(t) - \int_0^t Y_i(u) \exp(\widehat{\beta}^{\mathsf{T}} Z_i) d\widehat{A}_0(u).$$

Often, the martingale residual is simply defined as $\widehat{M}_i(\infty)$.

Martingale residuals

Martingale residuals may be used directly to check the functional form of a quantitative covariate (e.g., log-linearity) by plotting *cumulative martingale residuals*

$$\sum_i I(Z \leq z)\widehat{M}_i(\infty)$$

against z, together with a large no of paths generated from the approximate asymptotic distribution (under linearity) and it is also possible to get a significance test.

We illustrate the method on the PBC3 data and the composite endpoint. The method is available in the `R`-package `timereg` (and in SAS `PROC PHREG`).

PBC3 data – composite endpoint

Table 2.4 *PBC3 trial in liver cirrhosis: Estimated coefficients (and SD) from a Cox model.*

Covariate		$\widehat{\beta}$	SD
Treatment	CyA vs. placebo	-0.496	0.226
Albumin	per 1 g/L	-0.116	0.021
Bilirubin	per 1 μmol/L	0.00895	0.00098

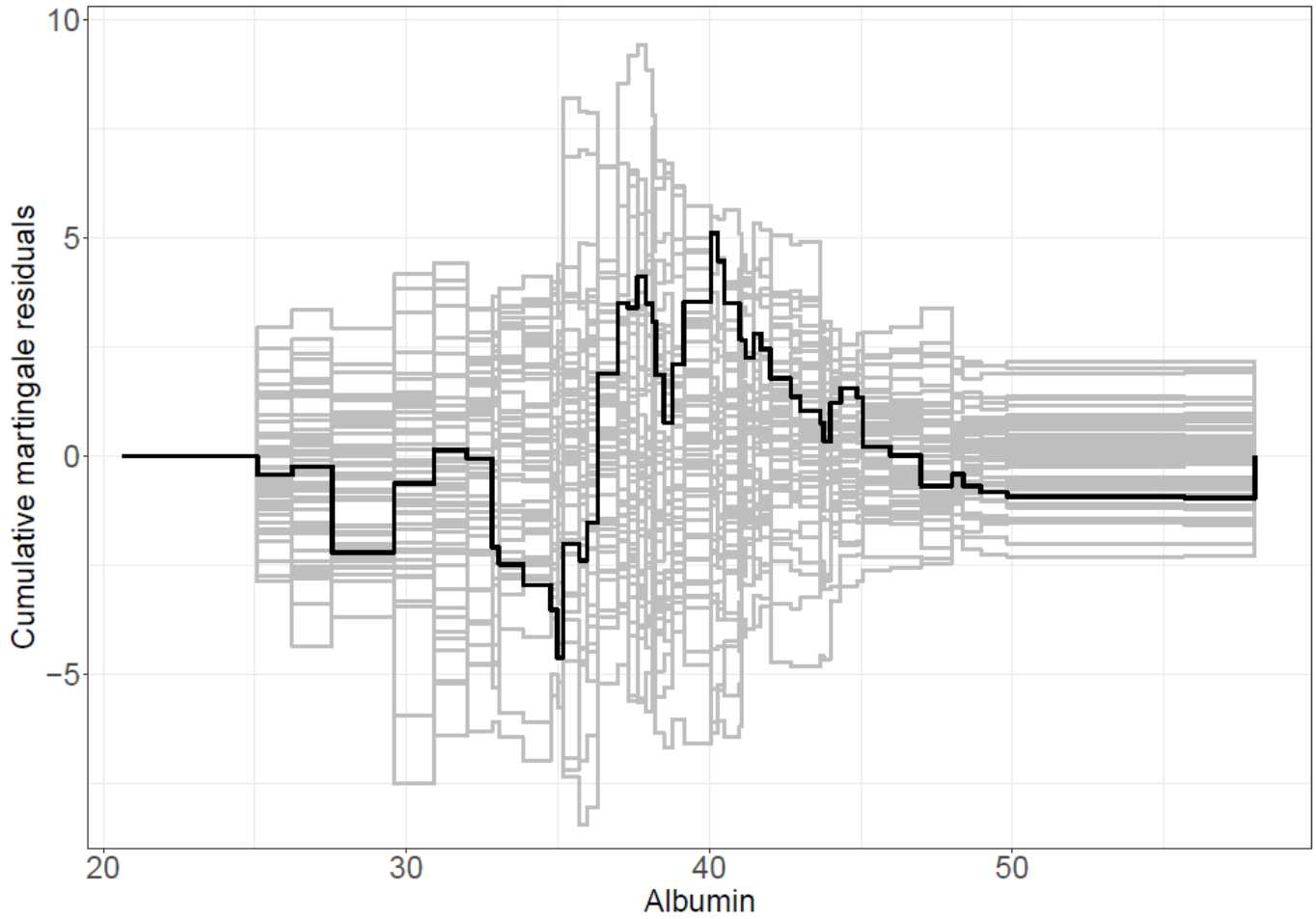


Figure 5.12 *PBC3 trial in liver cirrhosis: Checking linearity using cumulative martingale residuals plotted against albumin.*

P-value 0.459

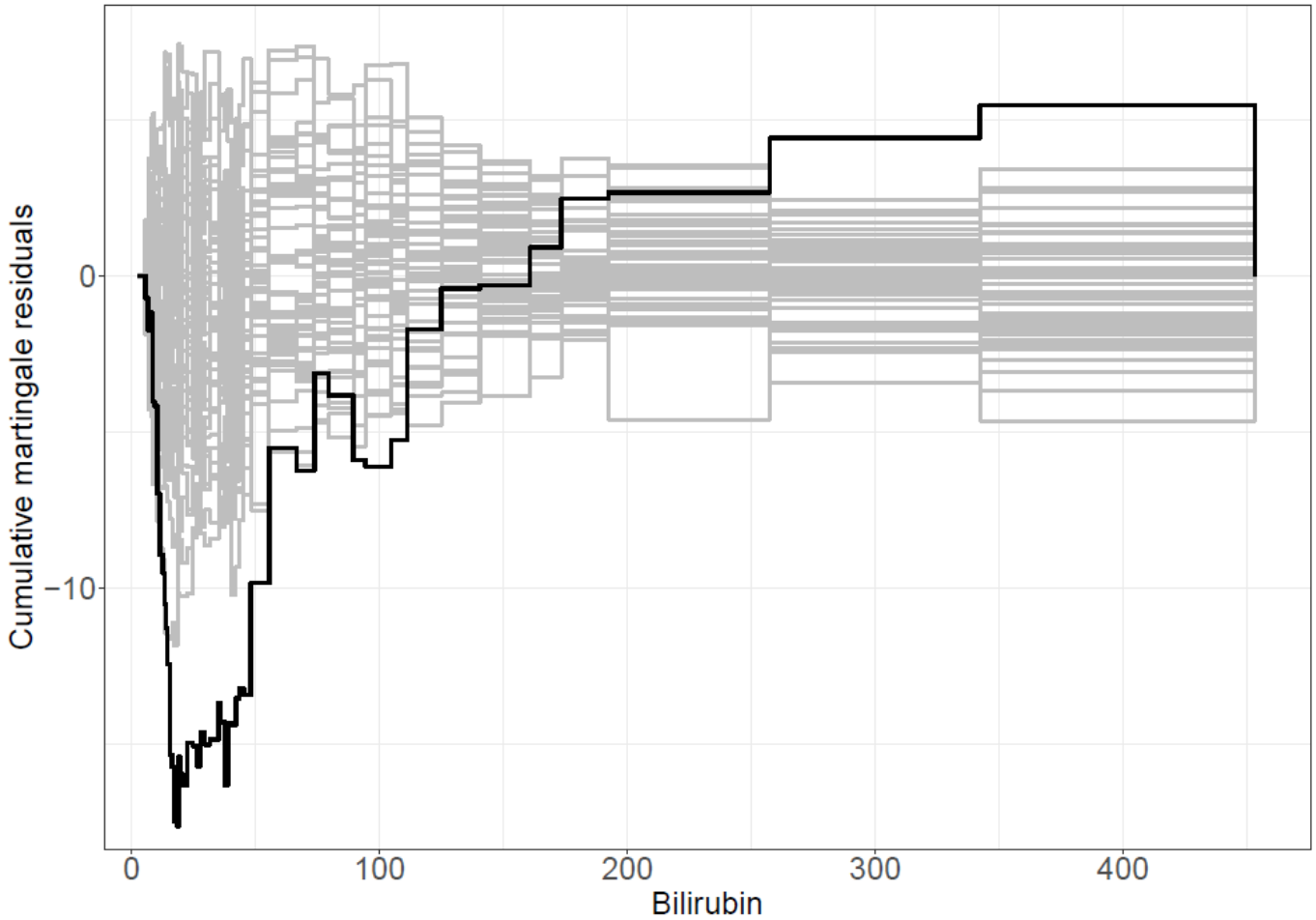


Figure 5.13 *PBC3 trial in liver cirrhosis: Checking linearity using cumulative martingale residuals plotted against bilirubin.*

P-value <0.001

Table 2.7 *PBC3 trial in liver cirrhosis: Estimated coefficients (and SD) from Cox and Poisson models with linear effects of albumin and* $\log_2$*(bilirubin).*

		Cox model		Poisson model	
Covariate		$\widehat{\beta}$	SD	$\widehat{\beta}$	SD
Treatment	CyA vs. placebo	-0.574	0.224	-0.546	0.223
Albumin	per 1 g/L	-0.091	0.022	-0.087	0.022
$\log_2$(bilirubin)	per doubling	0.665	0.074	0.647	0.073

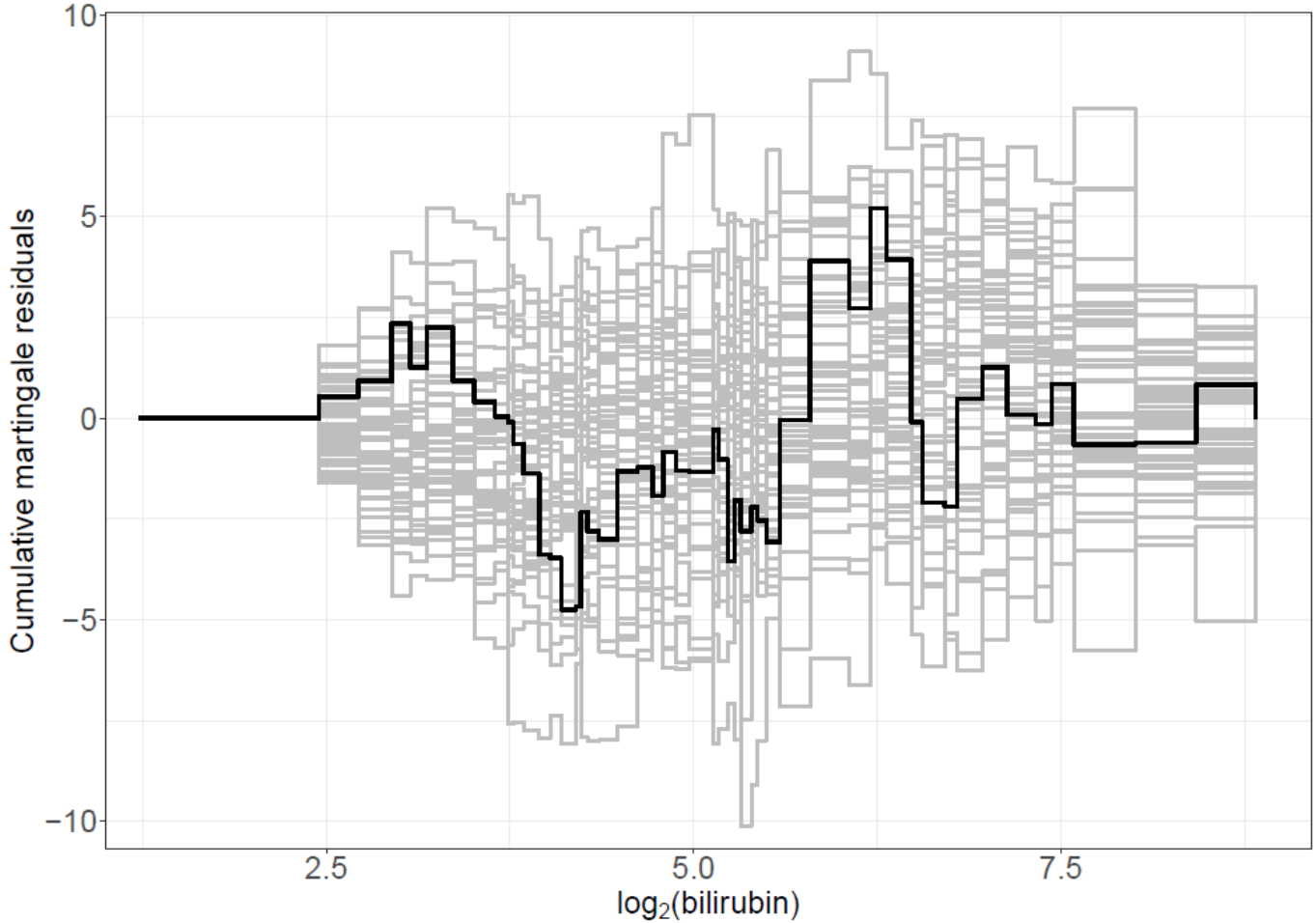


Figure 5.14 *PBC3 trial in liver cirrhosis: Checking linearity using cumulative martingale residuals plotted against* $\log_2$*(bilirubin).*

P-value 0.481

Score (Schoenfeld) residuals

The *score* for covariate j is:

$$U_j(\beta, \infty) = \sum_i D_i \big(Z_{ij} - \frac{\sum_{\ell \in R(X_i)} Z_{\ell j} \exp(\beta^{\mathsf{T}} Z_\ell)}{\sum_{\ell \in R(X_i)} \exp(\beta^{\mathsf{T}} Z_\ell)} \big), \quad j = 1, ..., p.$$

The term $U_{ij}(\beta, \infty)$ for subject i (only failures) is the *score-* (or *Schoenfeld-*) *residual*: $U_{ij}(\widehat{\beta}, \infty) = Z_{ij} - E_j(\widehat{\beta}, X_i)$.

The *score process* $U_j(\beta, t)$ only adds terms for subjects with $X_i \leq t$.

A *scaled* (or *weighted*) version divides by the estimated variance (say, V_j) of U_j (or by $\sqrt{V_j}$).

Schoenfeld (score) residuals

Scaled Schoenfeld residuals may be used directly to check for proportional hazards for a covariate: plot *cumulative scaled score residuals* (i.e., the scaled score process) against time, and it is possible to get a significance test.

This significance test is based on *re-sampling* from the distribution of the process under the model and evaluating where, in the re-sampled distribution, the observed process is.

We illustrate the method on the PBC3 data and the composite endpoint. The method is available in the `R`-package `timereg` (and in SAS `PROC PHREG`).

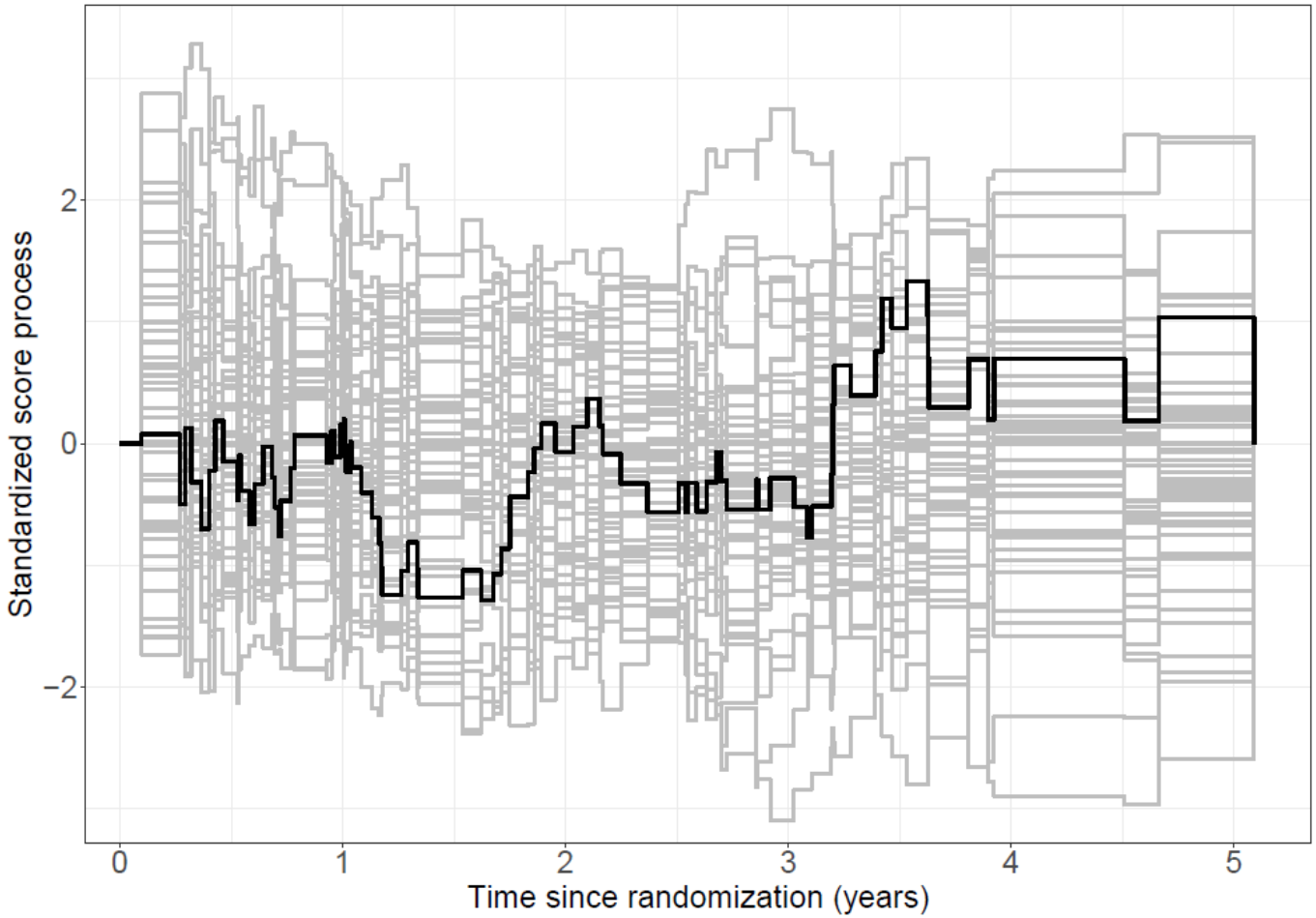


Figure 5.15 *PBC3 trial in liver cirrhosis: Checking proportional hazards using cumulative Schoenfeld residuals for treatment (standardized) plotted against the time-variable.*

P-value 0.919

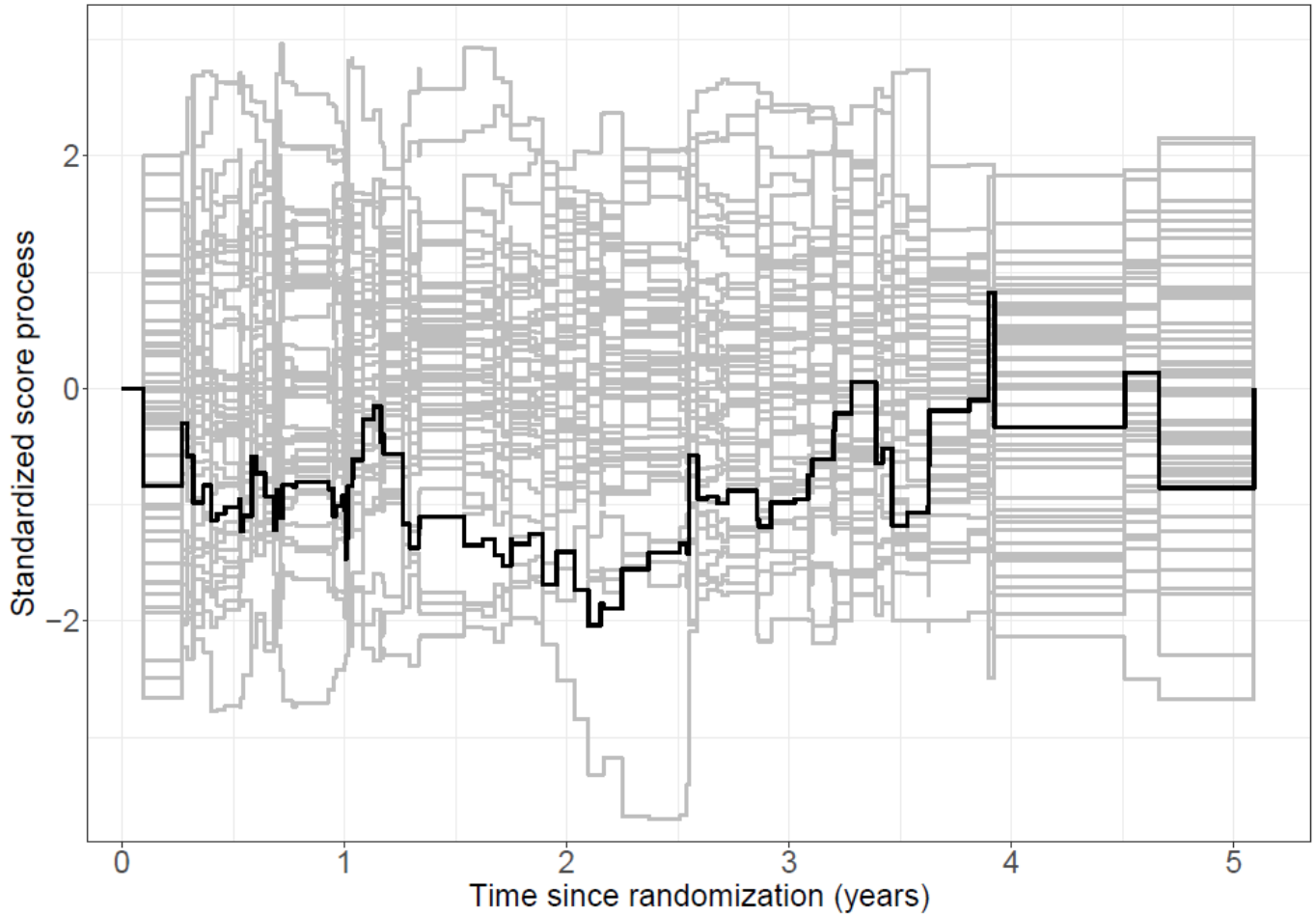


Figure 5.16 *PBC3 trial in liver cirrhosis: Checking proportional hazards using cumulative Schoenfeld residuals for albumin (standardized) plotted against the time-variable.*

P-value 0.418

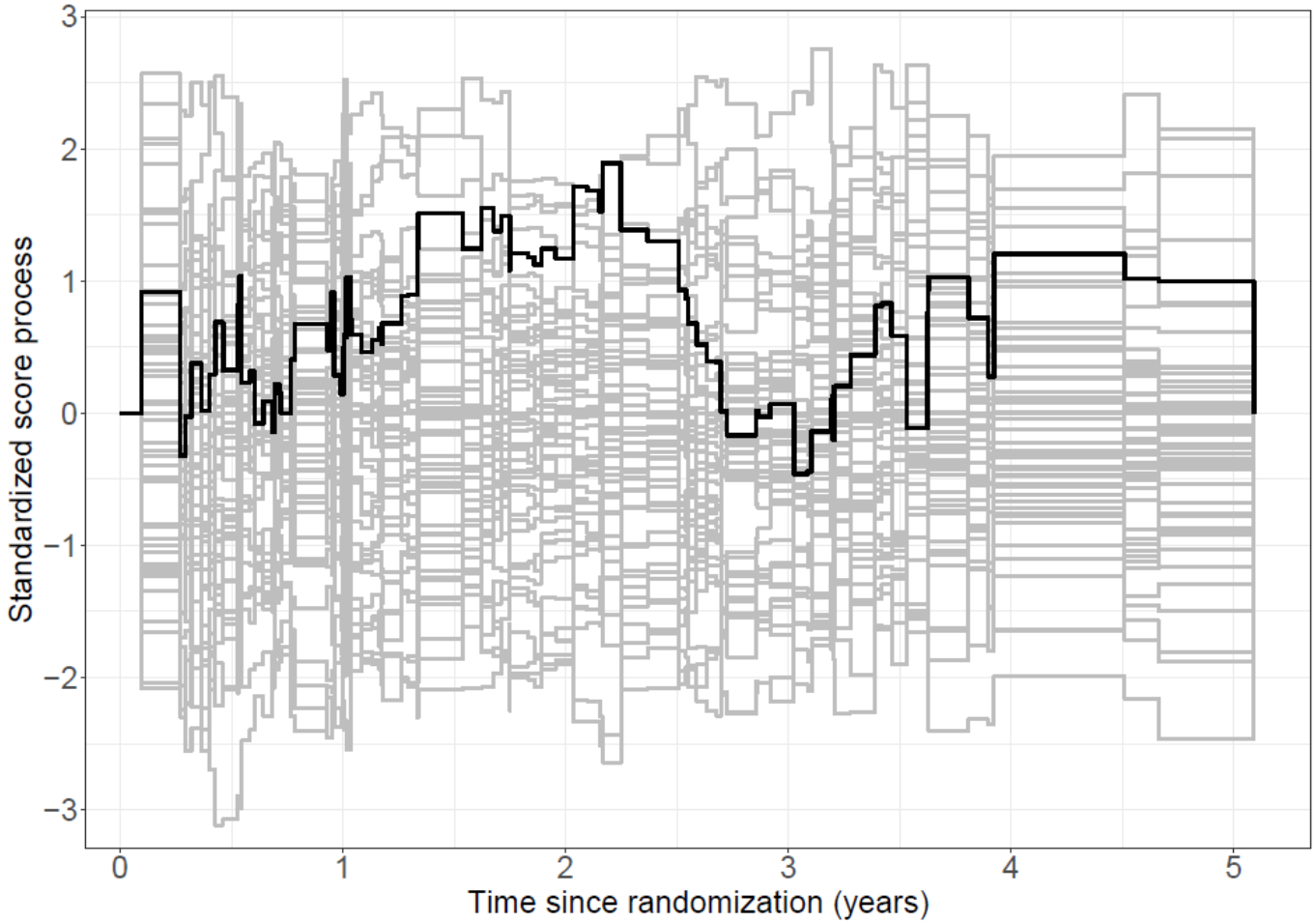


Figure 5.17 *PBC3 trial in liver cirrhosis: Checking proportional hazards using cumulative Schoenfeld residuals for* $\log_2$*(bilirubin) (standardized) plotted against the time-variable.*

P-value 0.568

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

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 – Cox models

1. Investigate if treatment (`tment`) impacts the hazard using a Cox model. Compare the score and logrank test.
2. Estimate effect of bilirubin using a Cox model by categorizing bilirubin into a binary variable (splitting at upper quartile as yesterday). Use the Breslow estimator to estimate the integrated hazards based on the Cox model and compare to the non-parametric Nelson-Aalen estimates.
3. Estimate treatment effect adjusted for bilirubin using a Cox model by categorizing bilirubin into a binary variable (splitting at upper quartile as yesterday). Estimate the effect of treatment within each level of bilirubin and test the assumption of 'no interaction' between the two variables.
4. Estimate the treatment effect adjusted for bilirubin as binary, now in a stratified Cox model. Compare the score test to the stratified logrank test. Estimate the effect of treatment for each stratum and test the assumption of 'no interaction' between treatment and bilirubin in the stratified Cox model.

R exercises – two state-model

Intensities – Cox models

5. Investigate (again) if bilirubin impacts the hazard using a Cox model, but now using bilirubin as a quantitative variable. Evaluate linearity of bilirubin in the Cox model by adding a quadratic term of bilirubin. Is a transformation relevant?
6. Investigate in a multiple Cox model treatment effect adjusted for both albumin and `log2(bili)` (as quantitative variables), and test for 'no interaction' between treatment and `log2(bili)`.
7. Replicate the Cox models in row three ($f(t) = I(t > 2)$) in Table 3.11, p. 96.

Marginal parameters

8. Reproduce Figure 4.5, p. 123.

R exercises – competing risks

Intensities

1. Reproduce Table 2.13, p. 62.
2. Change above Cox models to Cox models stratified by sex.
3. Remove sex from the model for transplantation.

Marginal parameters

4. Produce unstacked versions of Figure 4.12 (a) and (c), p. 129.