The Cox Proportional Hazards Regression Model

Henrik Ravn

Novo Nordisk

DSBS Course
Survival Analysis in Clinical Trials
January 2019

The Cox proportional hazards regression model

The Cox model assumes that the hazard for the ith individual ($i = 1, \ldots, n$) is

$$\begin{aligned} \lambda_i(t) &= \lambda_0(t) \exp(\beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{ip}) \\ &= \lambda_0(t) \exp(\beta' X_i) \end{aligned}$$

where $\beta_1, \beta_2, \ldots, \beta_p$ are regression parameters, X_{i1} is the covariate value for covariate 1 for individual i, *etc*.

$\lambda_0(t)$ is called the *baseline hazard*.

Time t is the time-scale of choice, e.g. age or time since randomization. As formulated here the only quantity depending on time t is the baseline hazard $\lambda_0(t)$.

$$\lambda_i(t) \quad = \quad \lambda_0(t) \exp(\beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{ip}).$$

If all covariates (X's) are zero we get

$$\lambda_i(t) \quad = \quad \lambda_0(t).$$

The interpretation of the baseline hazard is the hazard of an individual having all covariates equal to zero.

The Cox model does not make any assumptions about the shape of this baseline hazard, it is said to vary freely, and in the first place we are not interested in this baseline hazard. The focus is on the regression parameters.

The Cox model

$$\lambda_i(t) = \lambda_0(t)\exp(\beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{ip})$$

on the log-scale

$$\begin{aligned}\log(\lambda_i(t)) &= \log\left[(\lambda_0(t)\exp(\beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{ip})\right] \\ &= \log\left[\lambda_0(t)\right] + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{ip}.\end{aligned}$$

This means that the Cox model assumes that

- the effects of covariates are additive and linear on the log-rate scale.

Let's turn to the interpretation of the regression parameters $\beta_1, \beta_2, \ldots, \beta_p$.

One binary covariate

To make things simple we now study the effect of only one binary covariate, e.g. sex on the rate of dying

$$X_i = \begin{cases} 0 & \text{if individual } i \text{ is a female} \\ 1 & \text{if individual } i \text{ is a male} \end{cases}$$

The Cox model is

$$\lambda_i(t) \quad = \quad \lambda_0(t) \exp(\beta X_i).$$

With X_i defined as above we get

$$\lambda_i(t) = \begin{cases} \lambda_{female}(t) \quad = \lambda_0(t) & \text{if individual } i \text{ is a female} \\ \lambda_{male}(t) = \lambda_0(t) \exp(\beta) & \text{if individual } i \text{ is a male} \end{cases}$$

Mortality Rate Ratio – Hazard Ratio

The *hazard ratio* (HR) between males and females is

$$\text{HR} = \frac{\lambda_{male}(t)}{\lambda_{female}(t)} = \frac{\lambda_0(t)\exp(\beta)}{\lambda_0(t)} = \exp(\beta).$$

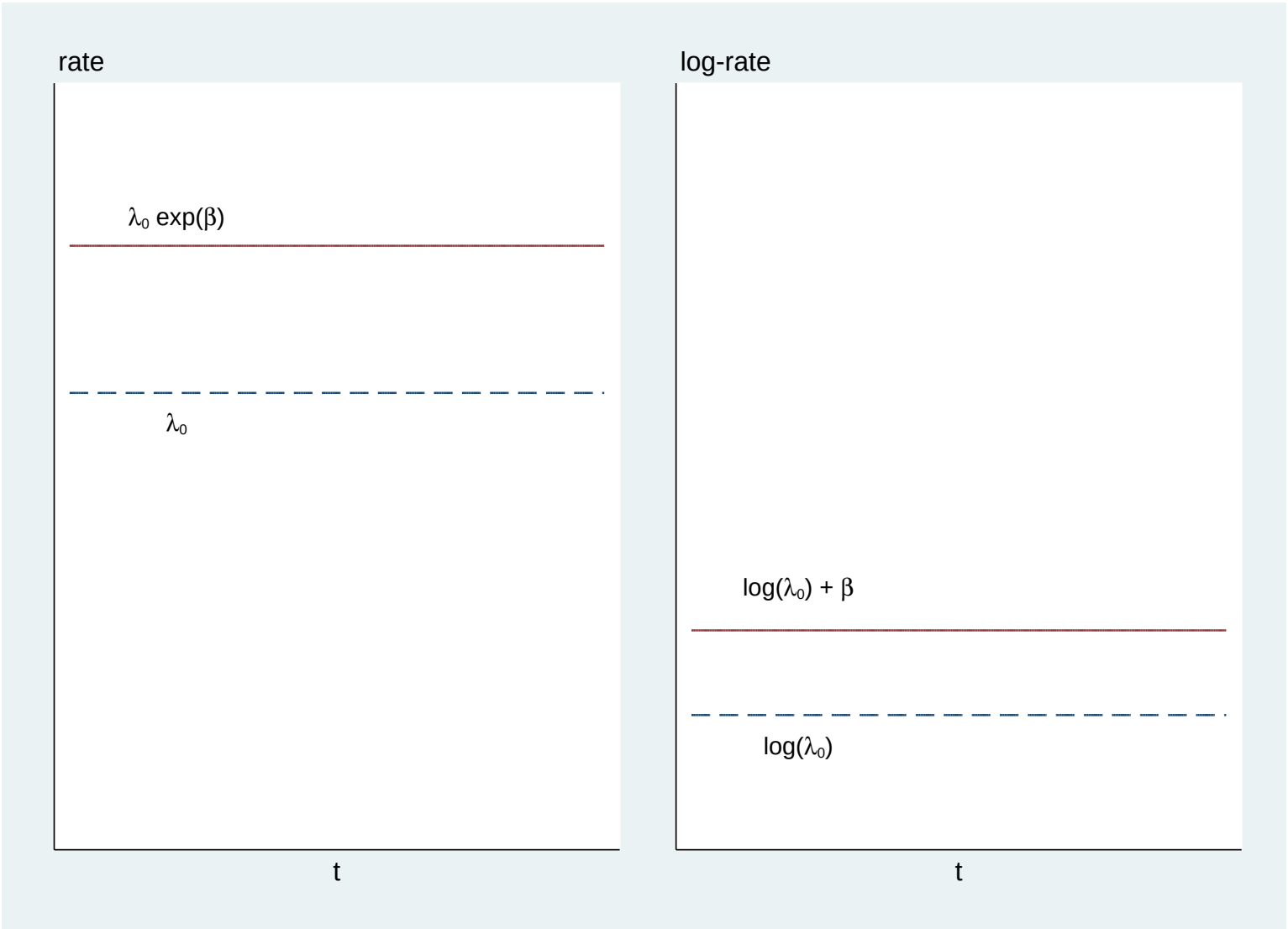
The ratio is independent of time, i.e. *proportional hazards* over time. Females are the *reference group*. On the log-scale

$$\log\left[\lambda_{male}(t)\right] - \log\left[\lambda_{female}(t)\right] = \log\left[\lambda_0(t)\exp(\beta)\right] - \log\left[\lambda_0(t)\right]$$

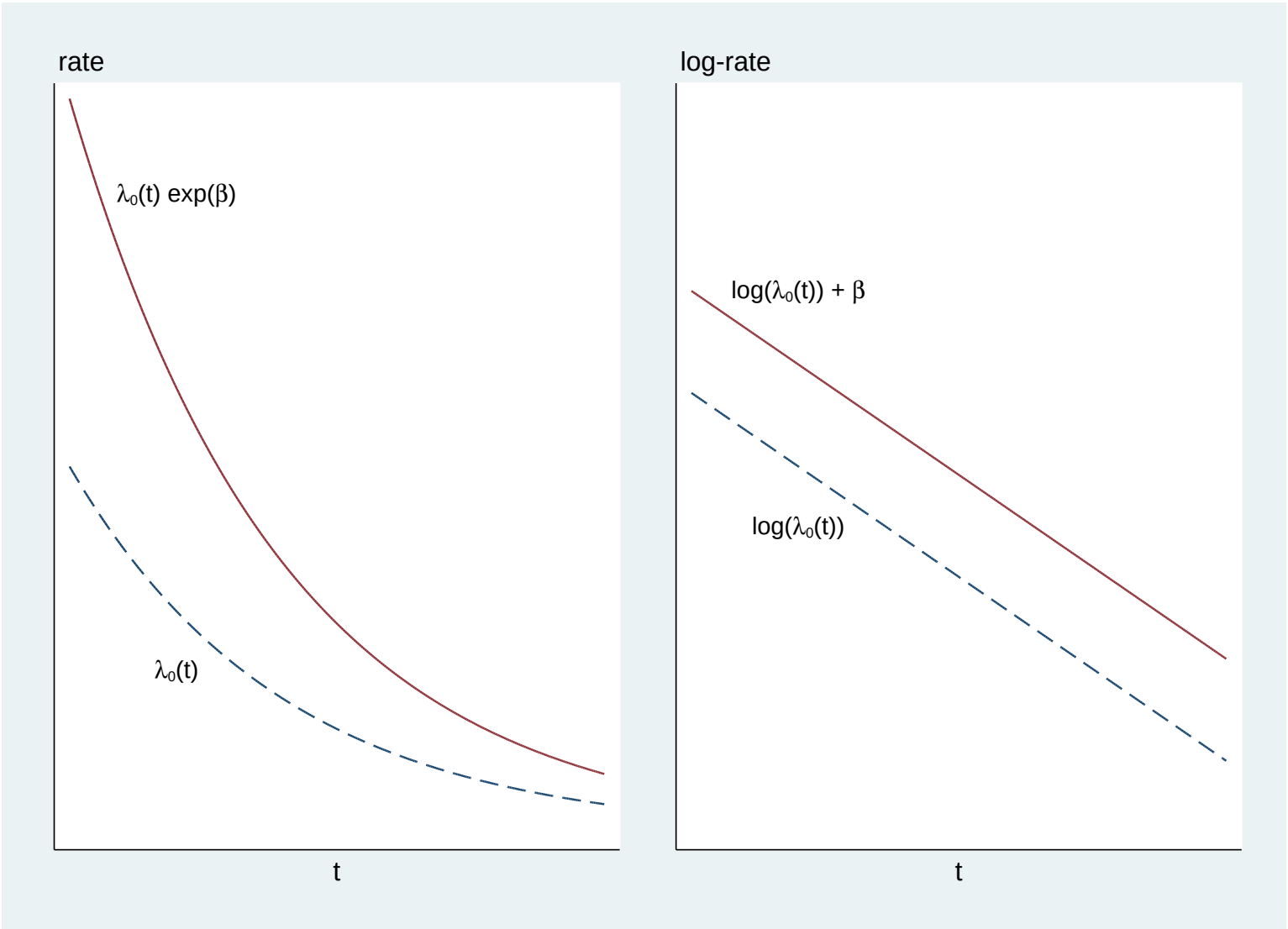
$$= \log\left[\lambda_0(t)\right] + \log\left[\exp(\beta)\right] - \log\left[\lambda_0(t)\right] = \beta.$$

Thus, the proportionality assumption is the same as a constant difference between the log-rates at any time t.

Proportional hazards



Proportional hazards



$$\text{HR} = \frac{\lambda_0(t)\exp(\beta)}{\lambda_0(t)} = \exp(\beta)$$

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

Cox's partial likelihood function

Cox's partial likelihood function for $i = 1, \ldots, n$ is

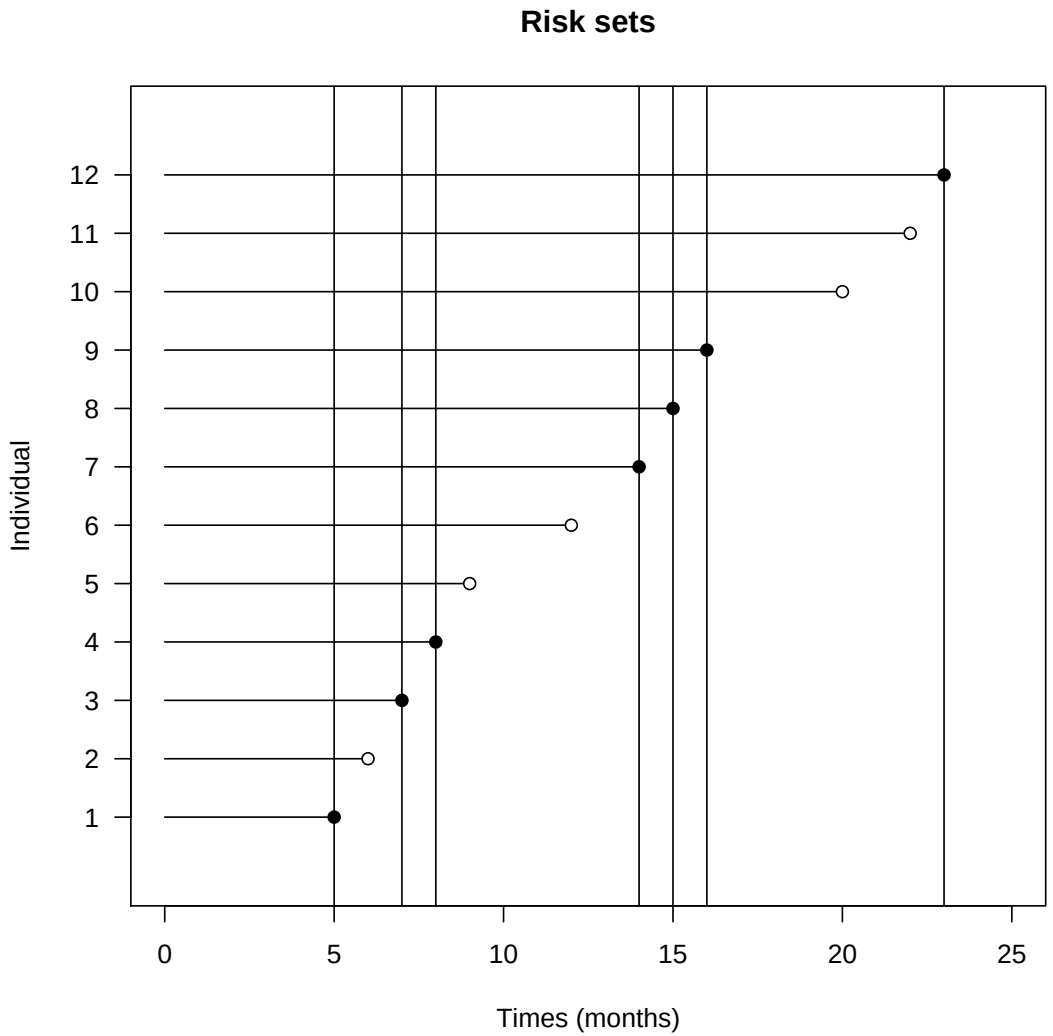
$$L(\beta) = \prod_{i=1}^{n} \left(\frac{\exp(\beta X_i)}{\sum_{j \in R(t_i)} \exp(\beta X_j)} \right)^{D_i}.$$

$R(t_i)$ is the **risk set** at time t_i, i.e. the set of individuals at risk of just before time t_i.

The partial likelihood function may be obtained from the general likelihood function presented earlier today by profiling out the baseline hazard function $\lambda_0(t)$.

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

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. If you e.g. have chosen age as the time-variable you have automatically adjusted for age. 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 (more later).

$$L(\beta) = \prod_{i=1}^{n} \left(\frac{\exp(\beta X_i)}{\sum_{j \in R(t_i)} \exp(\beta X_j)} \right)^{D_i}.$$

Math

The Cox partial likelihood is

$$L(\beta) = \prod_{i=1}^{n} \Big(\frac{\exp(\beta X_i)}{\sum_{j \in R(t_i)} \exp(\beta X_j)} \Big)^{D_i}.$$

The log-likelihood is then

$$\ell(\beta) = \log(L(\beta)) = \sum_{i=1}^{n} D_i \big(\beta X_i - \log \sum_{j \in R(t_i)} \exp(\beta X_j) \big)$$

and the *score* is

$$U(\beta) = \frac{d}{d\beta} \ell(\beta) = \sum_{i=1}^{n} \int_0^{\infty} \Big(X_i - \frac{\sum_j Y_j(t) X_j \exp(\beta X_j)}{\sum_j Y_j(t) \exp(\beta X_j)} \Big) dN_i(t).$$

More math

When evaluated at the true parameter value (β_0) and considered as a process in t (i.e., integrating to t instead of ∞), by the Doob-Meyer decomposition, $U_t(\beta_0)$ is a *martingale*.

This provides a way studying the large sample properties:
The martingale CLT gives asymptotic normality of the score and, by a standard Taylor expansion argument, we get asymptotic normality of $\widehat{\beta}$.
The variance of $\widehat{\beta}$ may be estimated from the observed information and the distribution of standard likelihood-based tests are also obtained.
One such test is the *score test* and, in fact, for a binary covariate this score test is simply the *logrank test*. Thus, the logrank test is closely related to the Cox model and, therefore, it has certain optimality properties against proportional hazards alternatives.

Cox in SAS – PROC PHREG

```
PROC PHREG DATA=pbc3;
  CLASS tment;
  MODEL followup*status(0)=tment / RISKLIMITS;
RUN;
```

```
PROC PHREG DATA=pbc3;
  CLASS tment(ref="0");
  MODEL followup*status(0)=tment / RISKLIMITS;
RUN;
```

Categorical covariates

PBC 3 trial: histological stage (1-2, 3, 4)

$$\text{stage} = \begin{cases} 2 & \text{histological stage} = 1 \text{ or } 2 \\ 3 & \text{histological stage} = 3 \\ 4 & \text{histological stage} = 4 \end{cases}$$

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

$$\text{stage3} = \begin{cases} 1 & \text{histological stage} = 3 \\ 0 & \text{otherwise} \end{cases}$$

and

$$\text{stage4} = \begin{cases} 1 & \text{histological stage} = 4 \\ 0 & \text{otherwise} \end{cases}$$

We only need two indicators, because if both are zero the patient will be in histological stage $= 1$ or 2, which is then the reference. The Cox model becomes

$$\lambda(t) \quad = \quad \lambda_0(t)\exp(\beta_1 \cdot \text{stage3} + \beta_2 \cdot \text{stage4}),$$

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

$$\lambda(t) = \begin{cases} \lambda_0(t)\exp(\beta_1) & \text{if histological stage} = 3 \\ \lambda_0(t)\exp(\beta_2) & \text{if histological stage} = 4 \\ \lambda_0(t) & \text{histological stage} = 1 \text{ or } 2 \end{cases}$$

Categorical covariates in SAS

```
PROC PHREG DATA=pbc3;
  CLASS stage(REF="2");
  MODEL followup*status(0)=stage / RL;
RUN;
```

Likelihood ratio test and confidence intervals

```
PROC PHREG DATA=pbc3;
  CLASS stage(REF="2");
  MODEL followup*status(0)=stage / RL=PL
     TYPE3(LR);
RUN;
```

Categorical covariates in SAS

```
PROC PHREG DATA=pbc3;
  CLASS tment(ref="0") stage(REF="2");
  MODEL followup*status(0)=tment stage / RL;
RUN;
```

Quantitative covariates

Using bilirubin measured in micromoles/L (variable `bili`) from the PBC trial:

$$\lambda(t) \quad = \quad \lambda_0(t)\exp(\beta \cdot \text{bilirubin})$$

or on the log-rate scale

$$\log(\lambda(t)) \quad = \quad \log(\lambda_0(t)) + \beta \cdot \text{bilirubin}.$$

For all t this is a straight line with intercept $\log(\lambda_0(t))$ and slope β. The log-rate increases (or decreases) with β for each unit increase in bilirubin.

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

Having the model

$$\lambda(t) \quad = \quad \lambda_0(t)\exp(\beta \cdot \text{bili})$$

and comparing two patients with a bilirubin difference of 1 micromole/L. Let the reference have bilirubin $= X$ micromoles/L:

$$HR = \frac{\lambda_0(t)\exp(\beta \cdot (X+1))}{\lambda_0(t)\exp(\beta \cdot X)} = \exp(\beta).$$

Comparing two patients with a bilirubin difference of 10 micromoles/L. Let the reference have bilirubin $= X$ micromoles/L:

$$HR = \frac{\lambda_0(t)\exp(\beta \cdot (X+10))}{\lambda_0(t)\exp(\beta \cdot X)} = \exp(\beta \cdot 10) = \exp(\beta)^{10}$$

Quantitative covariates in SAS

```
PROC PHREG DATA=pbc3;
  MODEL followup*status(0)=bili / RL;
RUN;

PROC PHREG DATA=pbc3;
  bili10=bili/10;
  MODEL followup*status(0)=bili10 / RL;
RUN;
```

Interactions

Interactions between covariates follow those of other regression models (logistic, linear) – here illustrated in SAS

```
PROC PHREG DATA=pbc3;
  CLASS tment(REF="0") stage(REF="2");
  MODEL followup*status(0)=tment stage
      tment*stage / RL;
RUN;

PROC PHREG DATA=pbc3;
  CLASS tment(REF="0") stage(REF="2");
  MODEL followup*status(0)=stage tment(stage) /
      RL;
RUN;
```

Stratified Cox model

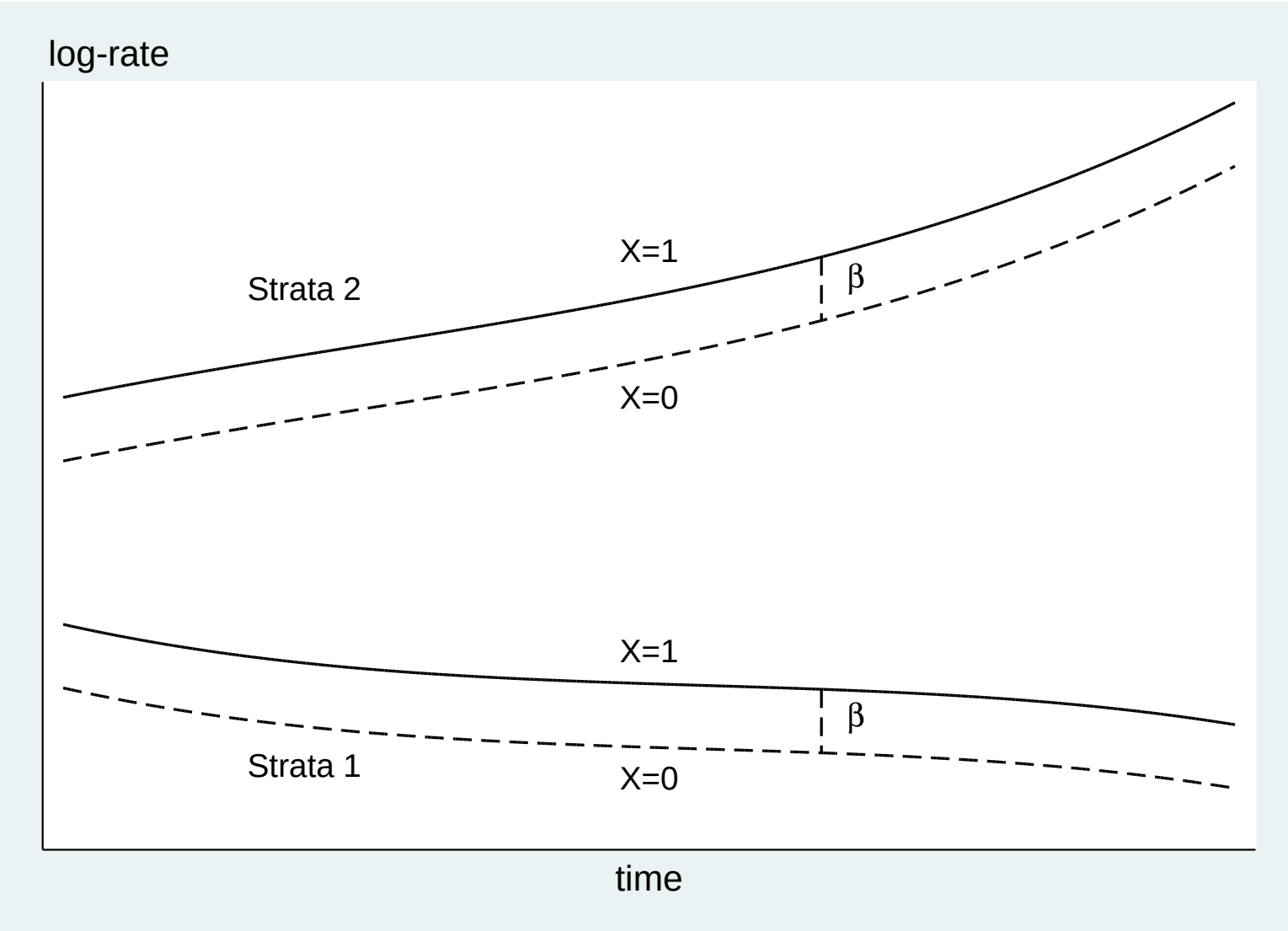
If the assumption of proportional hazards is violated (more on control of this later) for a categorical covariate with K categories it is possible to expand the Cox model to include different baseline hazards for each category

$$\lambda(t) \quad = \quad \lambda_{0k}(t) \exp(\beta X),$$

where $\lambda_{0k}(t)$ for $k = 1, \ldots, K$ is the baseline hazard in each of the K groups. These baseline hazards are 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.

Stratified Cox model



Likelihood Function for stratified model

In a stratified model

$$\lambda(t) \quad = \quad \lambda_{0k}(t)\exp(\beta X),$$

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

$$L(\beta) = \prod_{k=1}^{K} \prod_{i \in A_k} \left(\frac{\exp(\beta X_i)}{\sum_{j \in R_k(t_i)} \exp(\beta X_j)} \right)^{D_i},$$

$R_k(t_i)$ is the risk set for stratum k (A_k).

Stratified Cox in SAS

```
PROC PHREG DATA=pbc3;
   CLASS tment;
   MODEL followup*status(0)=tment / RISKLIMITS;
   STRATA sex;
RUN;
```

No need to declare the STRATA variable as a CLASS variable.

Cox assumptions

- The baseline hazard $\lambda_0(t)$ is non-parametric.
- The effects of covariates are additive and linear on the log-rate scale.
- Proportional hazards: The ratio of the hazard rates for two groups is constant over time.
- Time t is "automatically" adjusted for.

The Breslow estimator

The cumulative baseline hazard $\Lambda_0(t) = \int_0^t \lambda_0(s)ds$ from the Cox model $\lambda_i(t) = \lambda_0(t)\exp(\beta \cdot X_i)$ can be estimated by the Breslow estimator

$$\widehat{\Lambda}_0(t) = \sum_{t_i \leq t} \frac{d(t_i)}{\sum_{j \in R(t_i)} \exp(\widehat{\beta} \cdot X_j)},$$

where $\widehat{\beta}$ is the maximum likelihood estimate of β, $d(t_i)$ number of deaths at t_i, and $R(t_i)$ is the risk set of individuals at risk and under observation at time t_i. Having no covariates, the Breslow estimator is the Nelson-Aalen estimator

$$\widehat{\Lambda}_0(t) = \sum_{t_i \leq t} \frac{d(t_i)}{\sum_{j \in R(t_i)} \exp(\widehat{\beta} \cdot X_j)} = \sum_{t_i \leq t} \frac{d(t_i)}{Y(t_i)}$$

Predicted probabilities from Cox model

We have the relationship between survival and hazards functions

$$S(t, X) = [\exp(-\Lambda_0(t))]^{\exp(\beta' X)} = [S_0(t)]^{\exp(\beta' X)}$$

The predicted survival probabilities from a Cox model for a set of covariates X may be estimated by

$$\widehat{S}(t, X) \quad = \quad \widehat{S}_0(t)^{\exp(\hat{\beta}' X)},$$

where

$$\widehat{S}_0(t) = \exp(-\widehat{\Lambda}_0(t))$$

using the Breslow estimator for $\Lambda_0(t)$ or alternative estimators for $S_0(t)$.

Predicted probabilities from Cox model in SAS

```
PROC PHREG DATA=pbc3 PLOTS=SURVIVAL;
  CLASS tment;
  MODEL followup*status(0)=tment / RISKLIMITS;
  BASELINE / METHOD=BRESLOW;
RUN;
```

Remember, the estimated curve (probabilities) is based on the PH-assumption and given for the reference group for `tment`, i.e. on treatment.

Predicted probabilities from Cox model in SAS

```
PROC PHREG DATA=pbc3 PLOTS=SURVIVAL;
  CLASS tment;
  MODEL followup*status(0)=tment bili /
     RISKLIMITS;
  BASELINE / METHOD=PL;
RUN;

PROC PHREG DATA=pbc3 PLOTS=SURVIVAL;
  CLASS tment;
  MODEL followup*status(0)=tment bili /
     RISKLIMITS;
  BASELINE / METHOD=BRESLOW;
RUN;
```

Again remember, the estimated curve (probabilities) is based on the PH-assumptions and for a subject with an average bilirubin and on treatment.

Kaplan-Meier estimator from PROC PHREG

```
PROC PHREG DATA=pbc3 PLOTS=SURVIVAL;
  MODEL followup*status(0)=;
  STRATA tment;
  BASELINE / METHOD=PL;
  OUTPUT OUT=kmdata SURVIVAL=s / METHOD=PL;
RUN;
```

If not specifying METHOD=PL, the "exp(-N-Aa)" estimator is obtained.

Alternative estimator for $S(t)$ from PROC PHREG

```
PROC PHREG DATA=pbc3 PLOTS=SURVIVAL;
  MODEL followup*status(0)=;
  STRATA tment;
  BASELINE / METHOD=BRESLOW;
  OUTPUT OUT=survdata SURVIVAL=s /
     METHOD=BRESLOW;
RUN;
```

The survival function is estimated using the Breslow estimator for the integrated baseline hazard, which in the case of no covariates is the Nelson-Aalen estimator

$$\widehat{S}_0(t) = \exp(-\widehat{\Lambda}_0(t)).$$

The METHOD=BRESLOW is the default!

Predicted risk difference

Without covariates, this can be estimated from the KM-estimator

$$\widehat{S}_2(\tau) - \widehat{S}_1(\tau).$$

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

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

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 – SAS

```
PROC PHREG DATA=pbc3 PLOTS(OVERLAY)=SURVIVAL;
  CLASS tment(REF="0") biligroup sex;
  MODEL followup*status(0)=tment biligroup sex;
  BASELINE / DIRADJ GROUP=tment;
run;
```

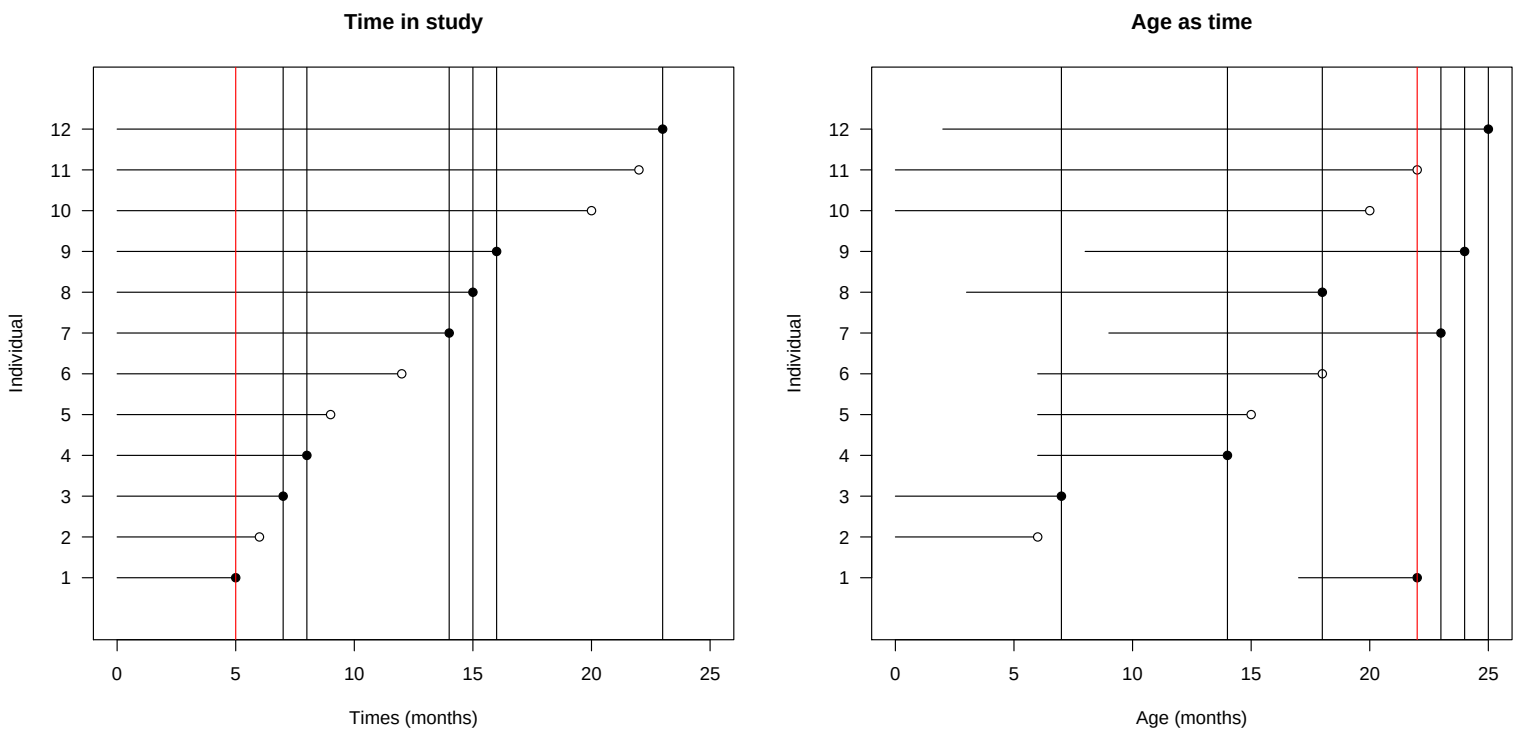
Algorithm:

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.

Delayed entry aka Left-truncation

Not often in randomised 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 handling of left-truncation is "easily" done by careful control of the risk sets $R(t_i)$ in the likelihood function:

$$L(\beta) = \prod_{i=1}^{d} \frac{\exp(\beta X_i)}{\sum_{j \in R(t_i)} \exp(\beta X_j)}$$

Only individuals at risk and under observation is included in the risk set $R(t_i)$ at time t_i. Remember, that the time-variable is automatically adjusted for, which means that if the time-variable in the Cox model is changed, so is the adjustment.

Additionally, a time of entry (e.g. age at entry) into the study will be needed in the data.

Left-truncation in SAS

The *Counting process style* of input for left-truncation

```
PROC PHREG DATA=epidata;
  CLASS vaccine;
  MODEL (inage,outage)*dead(0)= vaccine;
RUN;
```

Alternatively,

```
PROC PHREG DATA=epidata;
  CLASS vaccine;
  MODEL outage*dead(0)= vaccine / ENTRY=inage;
RUN;
```

Delayed entry, KM, logrank – SAS

As already mentioned, delayed entry has not been implemented in PROC LIFETEST, which means that neither KM nor logrank test is avaiable with delayed entries. However, delayed entry may be handled using PROC PHREG, e.g. Kaplan-Meier:

```
PROC PHREG DATA=epidata PLOT=S;
  MODEL outage*dead(0)= / ENTRY=inage;
  STRATA vaccine;
  BASELINE / METHOD=PL;
RUN;
```

and use the *score test* from the Cox model

```
PROC PHREG DATA=epidata;
  CLASS vaccine;
  MODEL outage*dead(0)=vaccine / ENTRY=inage;
RUN;
```

Time-dependent covariates

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

$$\lambda_i(t) \quad = \quad \lambda_0(t) \exp(\beta' X_i^*(t)).$$

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

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

Estimation with time-dependent covariates

Coxs partial likelihood becomes

$$L(\beta) = \prod_i \left(\frac{\exp(\beta' X_i^*(t_i))}{\sum_{j \in R(t_i)} \exp(\beta' X_j^*(t_i))} \right)^{D_i}.$$

Breslow estimator for cumulative baseline hazard:

$$\widehat{\Lambda}_0(t) = \sum_{t_i \leq t} \frac{D_i}{\sum_{j \in R(t_i)} \exp(\widehat{\beta}' X_j^*(t_i)},$$

NB: $X_j^*(t_i)$ 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 X be binary (treatment,placebo). An example of $X_i^*(t) = X_i(0) \cdot f(t)$ is the model

$$\lambda_i(t) \quad = \quad \lambda_0(t) \exp(\beta_1 X_i + \beta_2 X_i I(t \geq t_0)),$$

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 X. 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 $\beta_2 = 0$.

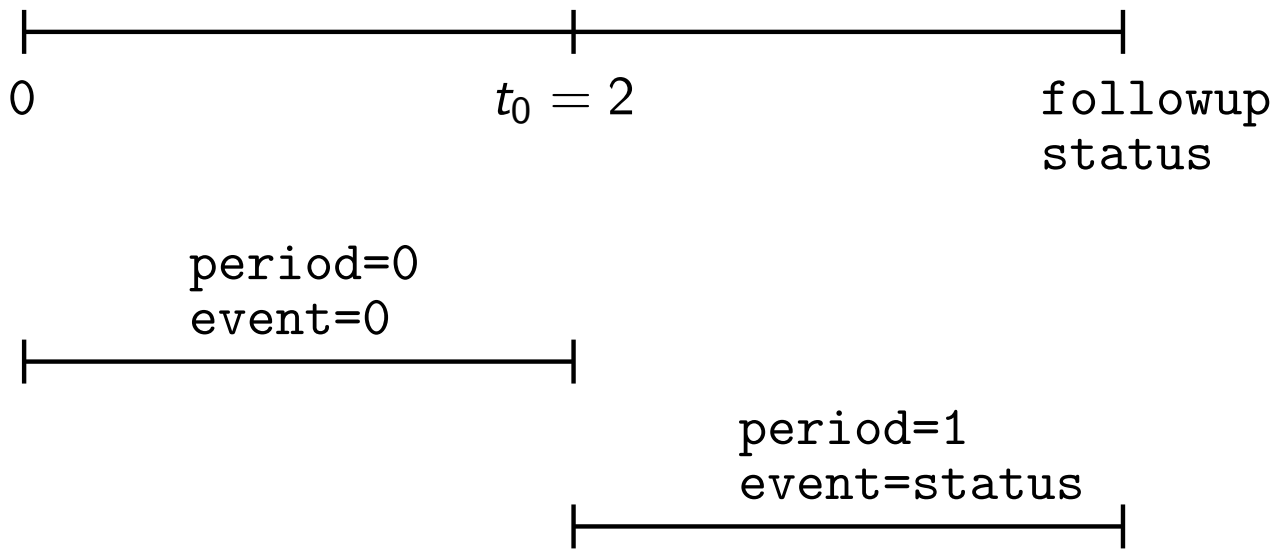
Alternative parametrisation

$$\lambda(t) \quad = \quad \lambda_0(t) \exp(\beta_1 \cdot X \cdot I(t \leq t_0) + \beta_2 \cdot X \cdot I(t > t_0)),$$

and we get the effect of X before and after time t_0, ($\exp(\beta_1)$) and ($\exp(\beta_2)$) respectively.

Binary time-dependent covariate in SAS (I)

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



Create an expanded data set and use counting process style of input

Time-dependent covariate in SAS (I)

```
DATA split;SET pbc3;
  IF followup<=2 THEN DO;
    period=0;in=0;out=followup;event=status;
    OUTPUT;
  END;
  IF followup>2 THEN DO;
    period=0;in=0;out=2;event=0;OUTPUT;
        period=1;in=2;out=followup;event=status;
    OUTPUT;
  END;
RUN;
PROC PHREG DATA=split;
  CLASS tment(REF="0") period;
  MODEL (in,out)*event(0)=tment|period;
  HAZARDRATIO tment / AT(period=ALL) DIFF=REF;
RUN;
```

Time-dependent covariate in SAS (II)

The programming statements re-calculate the covariates `tment1` and `tment2` at each event time represented by the variable `followup`.

```
PROC PHREG DATA=pbc3;
  tment1=tment*(followup<=2);
  tment2=tment*(followup>2);
  CLASS tment(ref="0");
  MODEL followup*status(0)=tment1 tment2 / RL;
  Equality: TEST tment1=tment2;
RUN;
```

For the first event time `followup`=0.06571 (24 days) and the programming statements are evaluated as

```
  tment1=tment*(0.06571<=2);
  tment2=tment*(0.06571>2);
```

The other parametrisation

```
PROC PHREG DATA=pbc3;
  tment2=tment*(followup>2);
  CLASS tment(ref="0");
  MODEL followup*status(0)=tment tment2 / RL;
  CONTRAST ">2 yrs" tment 1 tment2 1 /
     ESTIMATE=EXP;
RUN;
```

Stanford Heart Transplant Data

In the Stanford Heart Transplantation Study, patients identified as been eligible (N=103) for a heart transplant were put on a waiting list (time 0) and followed until transplantation, death or censorship. In total 69 received transplant during follow-up, whereas 34 did not. On the next frame you will find the variables in the transplant data set. Here we will discuss how to analyse and at the exercises we will do some of the analyses.

Stanford Heart Transplant Data variables

`age`	age (in years) at entry into the study.
`cens`	0 = Censoring 1 = Dead
`days`	number of days from entry to dead/censoring.
`trans`	1 = if the person had a heart transplantation 0 = otherwise.
`wait`	number of days from entry to transplantation NB: if trans=0 then wait=.
`mismatch`	1 = mismatch between HLA type in donor and patient 0 = no mismatch NB: if trans=0 then mismatch=.

Stanford Heart Transplant Data

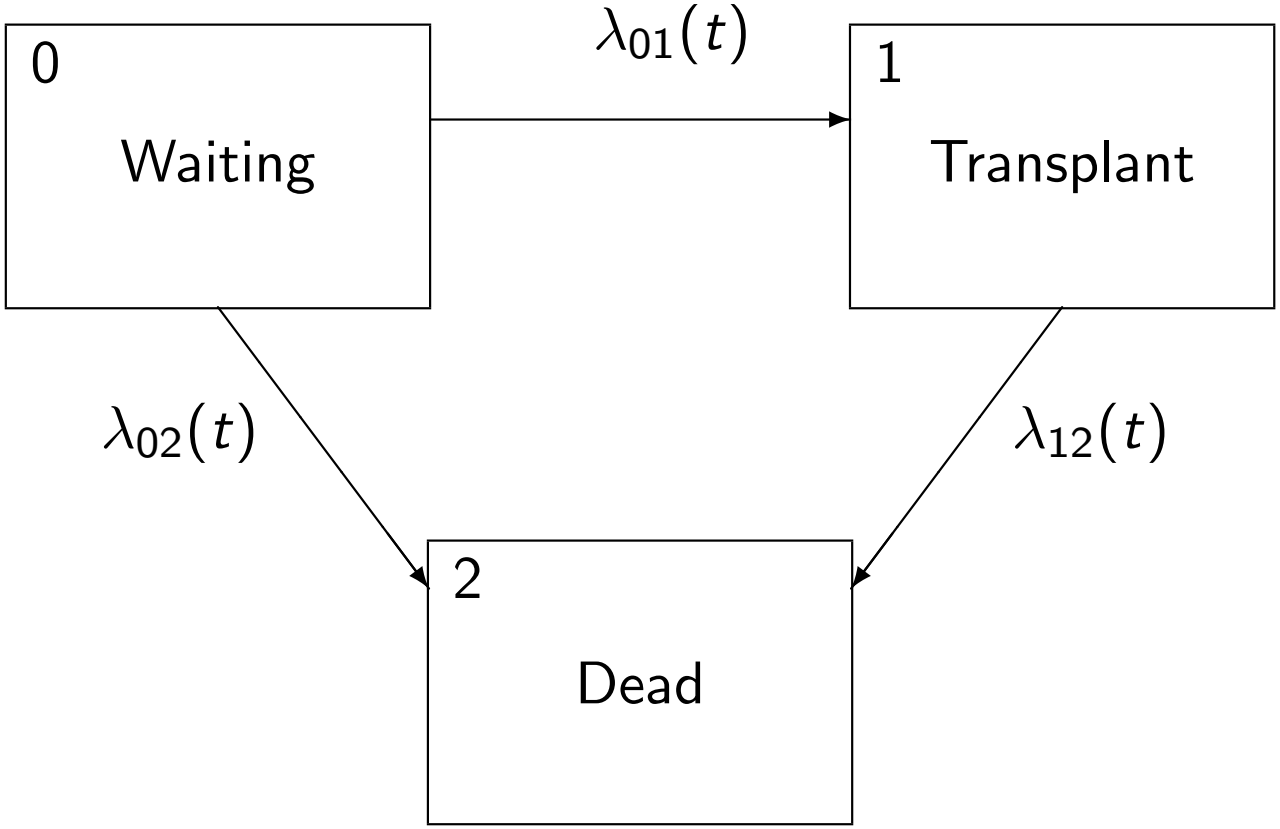
Obs	age	cens	days	trans	wait
54	48	1	100	1	46
55	41	1	102	0	.
56	28	0	109	1	96
57	46	1	110	1	60
58	23	0	131	1	21
59	41	1	149	0	.
60	47	1	153	1	26
61	43	1	165	1	4

Stanford Heart Transplant Data – SAS

```
PROC PHREG DATA=stanford;
  MODEL days*cens(0)=trans / RL;
RUN;
```

What is wrong?

Multi-state model



Stanford Heart Transplant Data – SAS

We need a time-dependent covariate:

```
PROC PHREG DATA=stanford;
  xtrans=(wait<days)*(wait>.);
  MODEL days*cens(0)=xtrans / RL;
RUN;
```

Stanford Heart Transplant Data – SAS

Alternatively

```
DATA updated;SET stanford;
IF wait = . THEN DO
  entry=0;exit=days;death=cens;xtrans=0;OUTPUT;
END;
IF wait ne . THEN DO;
  entry=0;exit=wait;death=0;xtrans=0;OUTPUT;
  entry=wait;exit=days;death=cens;xtrans=1;OUTPUT;
END;
RUN;

PROC PHREG DATA=updated;
  MODEL (entry,exit)*death(0) = xtrans / RL;
RUN;
```

Prediction and time-dependent covariates

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

$$S(t, X) = [\exp(-\Lambda_0(t))]^{\exp(\beta' X)}$$

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) behaviour of $(X(u); 0 \leq u < t)$.

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.

Checking assumptions for the linear predictor

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

- As noted earlier today, no interaction between X_{i1} and X_{i2} can be tested by adding suitable interaction terms to the model
- Linearity for quantitative X may be tested by adding quadratic terms X^2 or linear splines to the model. For chosen cut-points, say a_1, a_2 and add

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

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

Checking proportional hazards

This has developed into a whole "industry" within survival analysis and many approaches have been put forward.

- Graphical methods based on the stratified model, e.g. by ploting $\log(\widehat{\Lambda})_{0j}(t)$ against t (or $\log(t)$) for each stratum j and see if curves have constant vertical distance
- Modeling time-dependent effects
- Methods based on residuals
  - Martingale residuals
  - Score (Schoenfeld) residuals
  - Pseudo residuals

`ASSESS` statement in `PROC PHREG` uses several of these residuals. Martingale residuals may also be used directly to check the *functional form* of a quantitative covariate.