

# Who needs the Cox model anyway

**Bendix Carstensen** Steno Diabetes Center Copenhagen  
Gentofte, Denmark  
<http://BendixCarstensen.com>

SDC Epi and Biostat Network, 11 March 2020

From /home/bendix/teach/AdvCoh/talks/Aarhus2020/slides.tex

Thursday 12<sup>th</sup> March, 2020, 10:38

1/ 47

## The dogma [1]

- ▶ do not condition on the future — **indisputable**
- ▶ do not count people after they are dead — **disputable**
- ▶ stick to this world — **expandable**

P. K. Andersen and N. Keiding:  
Interpretability and importance of functionals in competing risks and multistate models  
Stat Med, 31:1074–1088, 2012

2/ 47

## (further) dogma for “sticking to this world”

- ▶ rates are continuous in time (and “smooth”)
- ▶ rates may depend on more than one time scale
- ▶ ... **which** timescales is an **empirical** question
- ▶ But first we look at the machinery for modeling simple occurrence rates from follow-up studies (mortality, incidence, ...)

3/ 47

- ▶ In follow-up studies we estimate rates from:
  - ▶  $D$  — events, deaths
  - ▶  $Y$  — person-years
  - ▶  $\hat{\lambda} = D/Y$  rates
  - ▶ ... empirical counterpart of intensity — an **estimate**
- ▶ Rates differ between persons.
- ▶ Rates differ **within** persons:
  - ▶ by age
  - ▶ by calendar time
  - ▶ by disease duration
  - ▶ ...
- ▶ Multiple timescales — later

4/ 47

## Representation of follow-up data

A cohort or follow-up study records **events** and **risk time**

The outcome (response) is thus **bivariate**:  $(d, y)$

Follow-up **data** for each individual must therefore have (at least) three pieces of information recorded:

Date of entry	entry	date variable
Date of exit	exit	date variable
Status at exit	event	indicator (mostly 0/1)

5/ 47

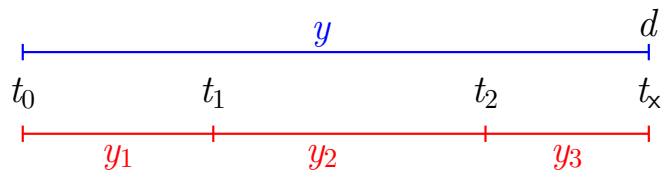
## From representation to likelihood

- ▶ Target is estimates of **occurrence rates** (mortality rates, incidence rates)
- ▶ ... and how these depend on covariates
- ▶ If we assume that mortality,  $\lambda$  is constant over time, then the log-likelihood from one person based on  $(d, y)$ :
  - ▶  $d$  — event, 0 or 1 (event)
  - ▶  $y$  — risk time (exit–entry)

$$\ell(\lambda) = d \log(\lambda) - \lambda y$$

- ▶ This formula is not derived here — see note on website

6/ 47



Probability

$$P(d \text{ at } t_x | \text{entry } t_0)$$

$$= P(\text{surv } t_0 \rightarrow t_1 | \text{entry } t_0)$$

$$\times P(\text{surv } t_1 \rightarrow t_2 | \text{entry } t_1)$$

$$\times P(d \text{ at } t_x | \text{entry } t_2)$$

log-Likelihood

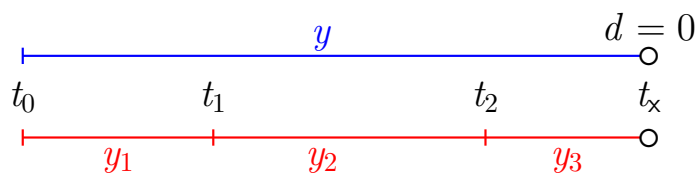
$$d \log(\lambda) - \lambda y$$

$$= 0 \log(\lambda) - \lambda y_1$$

$$+ 0 \log(\lambda) - \lambda y_2$$

$$+ d \log(\lambda) - \lambda y_3$$

7/ 47



Probability

$$P(\text{surv } t_0 \rightarrow t_x | \text{entry } t_0)$$

$$= P(\text{surv } t_0 \rightarrow t_1 | \text{entry } t_0)$$

$$\times P(\text{surv } t_1 \rightarrow t_2 | \text{entry } t_1)$$

$$\times P(\text{surv } t_2 \rightarrow t_x | \text{entry } t_2)$$

log-Likelihood

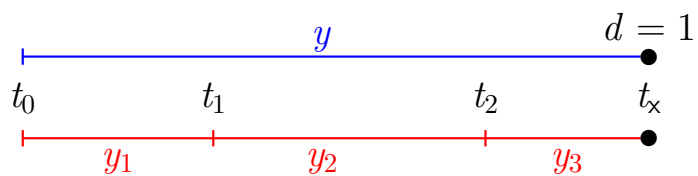
$$0 \log(\lambda) - \lambda y$$

$$= 0 \log(\lambda) - \lambda y_1$$

$$+ 0 \log(\lambda) - \lambda y_2$$

$$+ 0 \log(\lambda) - \lambda y_3$$

8/ 47



Probability

$$P(\text{event at } t_x | \text{entry } t_0)$$

$$= P(\text{surv } t_0 \rightarrow t_1 | \text{entry } t_0)$$

$$\times P(\text{surv } t_1 \rightarrow t_2 | \text{entry } t_1)$$

$$\times P(\text{event at } t_x | \text{entry } t_2)$$

log-Likelihood

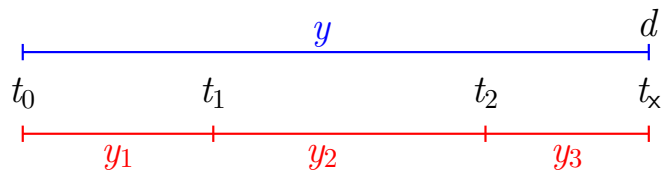
$$1 \log(\lambda) - \lambda y$$

$$= 0 \log(\lambda) - \lambda y_1$$

$$+ 0 \log(\lambda) - \lambda y_2$$

$$+ 1 \log(\lambda) - \lambda y_3$$

9/ 47



Probability

log-Likelihood

$$P(d \text{ at } t_x | \text{entry } t_0)$$

$$d \log(\lambda) - \lambda y$$

$$= P(\text{surv } t_0 \rightarrow t_1 | \text{entry } t_0)$$

$$= 0 \log(\lambda) - \lambda y_1$$

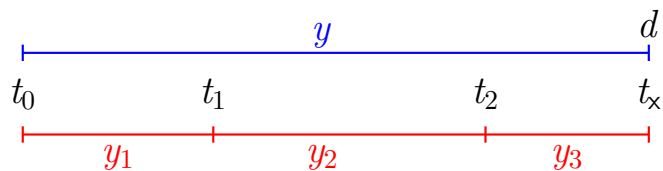
$$\times P(\text{surv } t_1 \rightarrow t_2 | \text{entry } t_1)$$

$$+ 0 \log(\lambda) - \lambda y_2$$

$$\times P(d \text{ at } t_x | \text{entry } t_2)$$

$$+ d \log(\lambda) - \lambda y_3$$

10/ 47



Probability

log-Likelihood

$$P(d \text{ at } t_x | \text{entry } t_0)$$

$$d \log(\lambda) - \lambda y$$

$$= P(\text{surv } t_0 \rightarrow t_1 | \text{entry } t_0)$$

$$= 0 \log(\lambda_1) - \lambda_1 y_1$$

$$\times P(\text{surv } t_1 \rightarrow t_2 | \text{entry } t_1)$$

$$+ 0 \log(\lambda_2) - \lambda_2 y_2$$

$$\times P(d \text{ at } t_x | \text{entry } t_2)$$

$$+ d \log(\lambda_3) - \lambda_3 y_3$$

— allows different rates ( $\lambda_i$ ) in each interval

11/ 47

## Likelihood for time-split data

- ▶ The setup is for a situation where it is assumed that rates are constant in each of the intervals
- ▶ Each **record** in the data set represents follow-up for one **person** in one (small) interval — many records for each person
- ▶ Each **record** in the data set contributes a **term** to the likelihood
- ▶ Each **term** looks like a contribution from a Poisson variate (albeit with values only 0 or 1), with mean  $\lambda y$
- ▶  $\Rightarrow$  Likelihood for **one** person's FU (rate likelihood) is the same as the likelihood for **several** independent Poisson variates:
- ▶ **Two** models, **one** likelihood.

12/ 47

## Analysis of time-split data

Observations classified by  $p$ —person and  $i$ —interval

- ▶  $d_{pi}$  — In the model as response
- ▶  $y_{pi}$  — risk time  
In the model as offset  $\log(y)$  ... or as part of the response
- ▶ Covariates are:
  - ▶ timescales (age, period, time in study)
  - ▶ other variables for this person (constant in each interval).
- ▶ Model rates using the covariates in `glm`:  
— no difference in how time-scales and other covariates are modeled

13/ 47

## A look at the Cox model

$$\lambda(t, x) = \lambda_0(t) \times \exp(x'\beta)$$

A model for the rate as a function of  $t$  and  $x$ .

Covariates:

- ▶  $x$
- ▶  $t$
- ▶ ... often the effect of  $t$  is ignored (forgotten?)
- ▶ *i.e.* left unreported

14/ 47

## Cox-likelihood

The (partial) log-likelihood for the regression parameters:

$$\ell(\beta) = \sum_{\text{death times}} \log \left( \frac{e^{\eta_{\text{death}}}}{\sum_{i \in \mathcal{R}_t} e^{\eta_i}} \right)$$

is also a **profile likelihood** in the model where observation time has been subdivided in small pieces (empirical rates) and each small piece provided with its own parameter:

$$\log(\lambda(t, x)) = \log(\lambda_0(t)) + x'\beta = \alpha_t + \eta$$

15/ 47

## The Cox-likelihood as profile likelihood

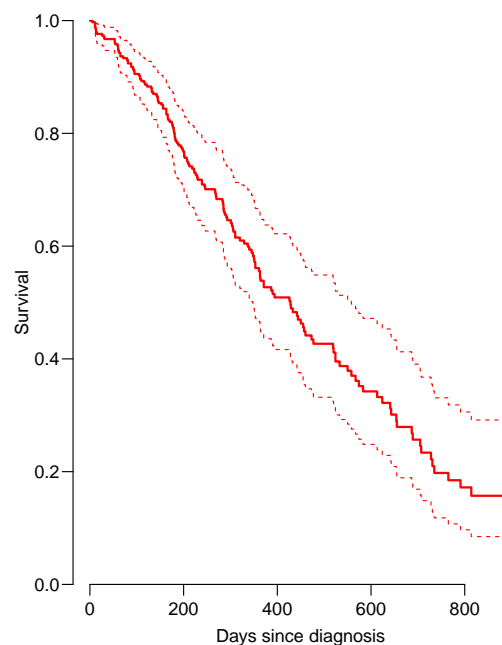
- ▶ One parameter per death time to describe the effect of time (i.e. the chosen timescale).

$$\log(\lambda(t, x_i)) = \log(\lambda_0(t)) + \underbrace{\beta_1 x_{1i} + \dots + \beta_p x_{pi}}_{\eta_i} = \alpha_t + \eta_i$$

- ▶ Profile likelihood:
  - ▶ Derive estimates of  $\alpha_t$  as function of data and  $\beta$ s
    - assuming constant rate between death/censoring times
  - ▶ Insert in likelihood, now only a function of data and  $\beta$ s
  - ▶ This turns out to be Cox's partial likelihood
- ▶ Cumulative intensity ( $\Lambda_0(t)$ ) obtained via the Breslow-estimator

16/ 47

### Mayo Clinic lung cancer data: 60 year old woman



17/ 47

## The Cox-likelihood: mechanics of computing

- ▶ The likelihood is computed by suming over risk-sets:

$$\ell(\eta) = \sum_t \log \left( \frac{e^{\eta_{\text{death}}}}{\sum_{i \in \mathcal{R}_t} e^{\eta_i}} \right)$$

- ▶ this is essentially splitting follow-up time at event- (and censoring) times
- ▶ ... repeatedly in every cycle of the iteration
- ▶ ... simplified by not keeping track of risk time
- ▶ ... but only works along **one** time scale

18/ 47

$$\log(\lambda(t, x_i)) = \log(\lambda_0(t)) + \underbrace{\beta_1 x_{1i} + \dots + \beta_p x_{pi}}_{\eta_i} = \alpha_t + \eta_i$$

- ▶ Suppose the time scale has been divided into small intervals with at most one death in each:
- ▶ Empirical rates:  $(d_{it}, y_{it})$  — each  $t$  has at most one  $d_{it} = 1$ .
- ▶ Assume w.l.o.g. the  $y$ s in the empirical rates all are 1.
- ▶ Log-likelihood contributions that contain information on a specific time-scale parameter  $\alpha_t$  will be from:
  - ▶ the (only) empirical rate  $(1, 1)$  with the death at time  $t$ .
  - ▶ all other empirical rates  $(0, 1)$  from those who were at risk at time  $t$ .

19/ 47

Note: There is one contribution from each person at risk to the part of the log-likelihood at  $t$ :

$$\begin{aligned} \ell_t(\alpha_t, \beta) &= \sum_{i \in \mathcal{R}_t} d_i \log(\lambda_i(t)) - \lambda_i(t) y_i \\ &= \sum_{i \in \mathcal{R}_t} \{d_i(\alpha_t + \eta_i) - e^{\alpha_t + \eta_i}\} \\ &= \alpha_t + \eta_{\text{death}} - e^{\alpha_t} \sum_{i \in \mathcal{R}_t} e^{\eta_i} \end{aligned}$$

where  $\eta_{\text{death}}$  is the linear predictor for the person that died at  $t$ .

20/ 47

The derivative w.r.t.  $\alpha_t$  is:

$$D_{\alpha_t} \ell_t(\alpha_t, \beta) = 1 - e^{\alpha_t} \sum_{i \in \mathcal{R}_t} e^{\eta_i} = 0 \quad \Leftrightarrow \quad e^{\alpha_t} = \frac{1}{\sum_{i \in \mathcal{R}_t} e^{\eta_i}}$$

If this estimate is fed back into the log-likelihood for  $\alpha_t$ , we get the **profile likelihood** (with  $\alpha_t$  “profiled out”):

$$\log\left(\frac{1}{\sum_{i \in \mathcal{R}_t} e^{\eta_i}}\right) + \eta_{\text{death}} - 1 = \log\left(\frac{e^{\eta_{\text{death}}}}{\sum_{i \in \mathcal{R}_t} e^{\eta_i}}\right) - 1$$

which is the same as the contribution from time  $t$  to Cox’s partial likelihood.

21/ 47

## Splitting the dataset a priori

- ▶ The Poisson approach needs a dataset of empirical rates  $(d, y)$  with suitably small values of  $y$ .
- ▶ — each individual contributes many empirical rates
- ▶ (one per risk-set contribution in Cox-modelling)
- ▶ From each empirical rate we get:
  - ▶ Poisson-response  $d$
  - ▶ Risk time  $y \rightarrow \log(y)$  as offset
  - ▶ time scale covariates: current age, current date, ...
  - ▶ other covariates
- ▶ Contributions not independent, but likelihood is a product
- ▶ Same likelihood as for independent Poisson variates
- ▶ Poisson `glm` with spline/factor effect of time

22/ 47

## History

This is not new, the profile likelihood was pointed out by Holford [2] in 1976, and the practical implementation was demonstrated by Whitehead in 1980 [3], using GLIM. ... so I am telling an old story here.

23/ 47

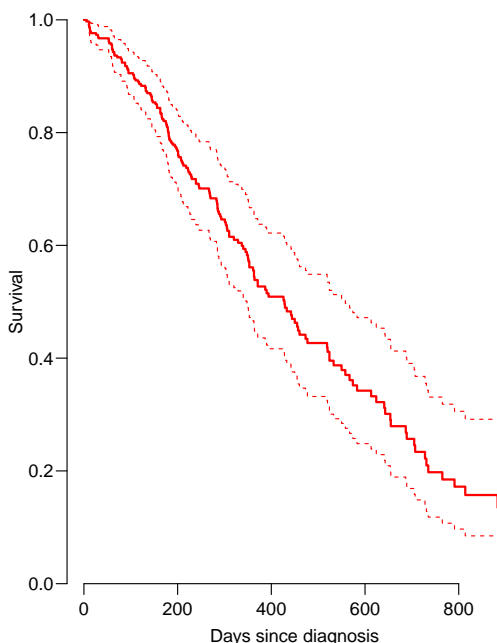
## Example: Mayo Clinic lung cancer

- ▶ Survival after lung cancer
- ▶ Covariates:
  - ▶ Age at diagnosis
  - ▶ Sex
  - ▶ Time since diagnosis
- ▶ Cox model
- ▶ Split data:
  - ▶ Poisson model, time as factor
  - ▶ Poisson model, time as spline

24/ 47



# Mayo Clinic lung cancer 60 year old woman



25/ 47

## Example: Mayo Clinic lung cancer I

```
> library( survival )
> library( Epi )
> library( popEpi )
> Lung <- Lexis( exit = list( tfe=time ),
+               exit.status = factor(status,labels=c("Alive","Dead")),
+               data = lung )
```

NOTE: entry.status has been set to "Alive" for all.  
NOTE: entry is assumed to be 0 on the tfe timescale.

```
> summary( Lung )
```

Transitions:

	To						
From	Alive	Dead	Records:	Events:	Risk time:	Persons:	
	Alive	63	165	228	165	69593	228

26/ 47

## Example: Mayo Clinic lung cancer II

```
> system.time(
+ mL.cox <- coxph( Surv( tfe, tfe+lex.dur, lex.Xst=="Dead" ) ~
+                   age + factor( sex ),
+                   method="breslow", data=Lung ) )
```

```
user system elapsed
0.027 0.021 0.020
```

```
> Lung.s <- splitMulti( Lung, tfe=c(0,sort(unique(Lung$time))) )
> summary( Lung.s )
```

Transitions:

	To						
From	Alive	Dead	Records:	Events:	Risk time:	Persons:	
	Alive	19857	165	20022	165	69593	228

```
> nlevels( factor( Lung.s$tfe ) )
```

```
[1] 186
```

27/ 47

## Example: Mayo Clinic lung cancer III

```
> subset( Lung.s, lex.id==96 )[,1:11]

  lex.id tfe lex.dur lex.Cst lex.Xst inst time status age sex ph.ecog
1:    96  0      5   Alive   Alive  12  30      2  72  1      2
2:    96  5      6   Alive   Alive  12  30      2  72  1      2
3:    96 11      1   Alive   Alive  12  30      2  72  1      2
4:    96 12      1   Alive   Alive  12  30      2  72  1      2
5:    96 13      2   Alive   Alive  12  30      2  72  1      2
6:    96 15     11   Alive   Alive  12  30      2  72  1      2
7:    96 26      4   Alive    Dead  12  30      2  72  1      2

> system.time(
+ mLs.pois.fc <- glm( cbind(lex.Xst=="Dead",lex.dur) ~ - 1 + factor( tfe ) +
+                   age + factor( sex ),
+                   family=poisreg, data=Lung.s, eps=10^-8, maxit=25 )
+ )

  user  system elapsed
12.789  19.108   9.286
```

28/ 47

## Example: Mayo Clinic lung cancer IV

```
> length( coef(mLs.pois.fc) )

[1] 188

> t.kn <- c(0,25,100,500,1000)
> dim( Ns(Lung.s$tfe,knots=t.kn) )

[1] 20022    4

> system.time(
+ mLs.pois.sp <- glm( cbind(lex.Xst=="Dead",lex.dur) ~ Ns( tfe, knots=t.kn ) +
+                   age + factor( sex ),
+                   family=poisreg, data=Lung.s ) )

  user  system elapsed
0.252   0.454   0.221
```

29/ 47

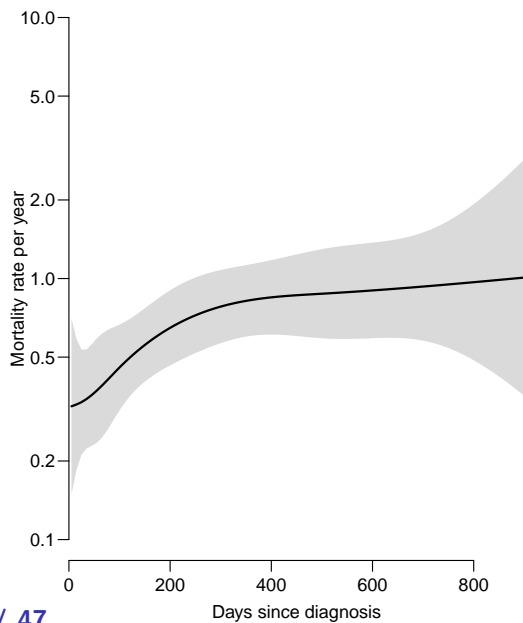
## Example: Mayo Clinic lung cancer V

```
> ests <-
+ rbind( ci.exp(mL.cox),
+       ci.exp(mLs.pois.fc,subset=c("age","sex")),
+       ci.exp(mLs.pois.sp,subset=c("age","sex")) )
> cmp <- cbind( ests[c(1,3,5) ,],
+             ests[c(1,3,5)+1,] )
> rownames( cmp ) <- c("Cox","Poisson-factor","Poisson-spline")
> colnames( cmp )[c(1,4)] <- c("age","sex")

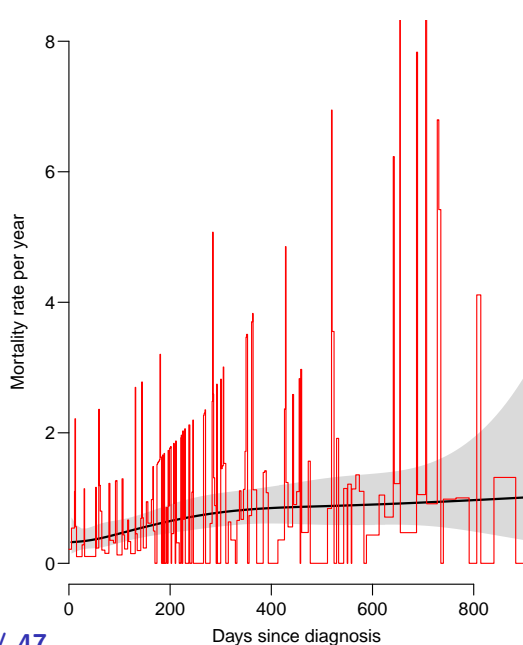
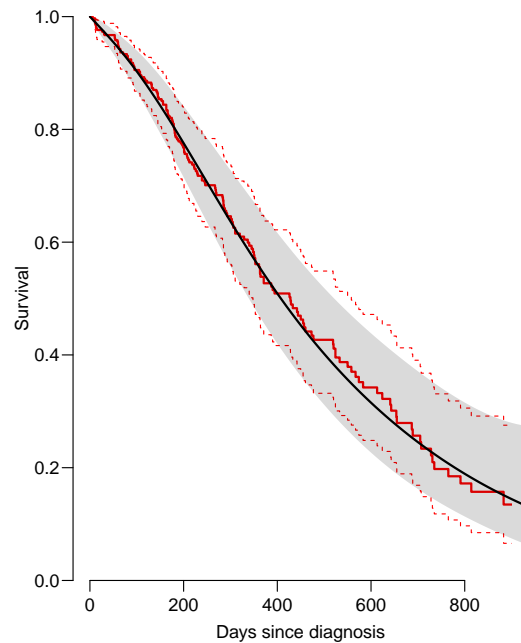
> round( cmp, 7 )

           age      2.5%      97.5%           sex      2.5%      97.5%
Cox      1.017158 0.9989388 1.035710 0.5989574 0.4313720 0.8316487
Poisson-factor 1.017158 0.9989388 1.035710 0.5989574 0.4313720 0.8316487
Poisson-spline 1.016189 0.9980321 1.034677 0.5998287 0.4319854 0.8328858
```

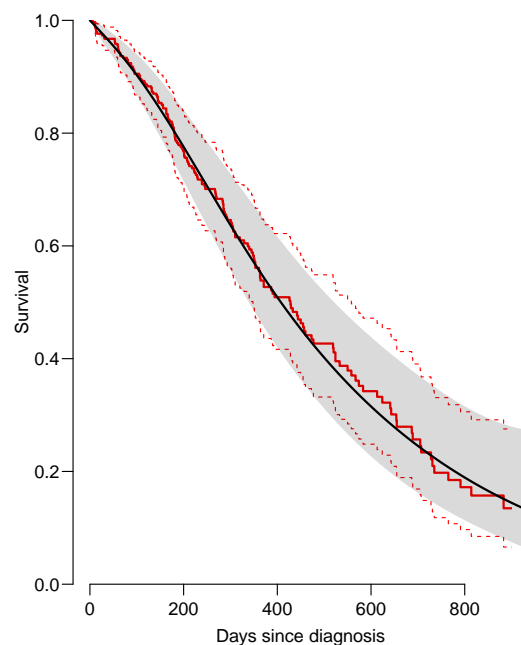
30/ 47



31/ 47



31/ 47



## Deriving the survival function

```
> mLS.pois.sp <- glm( lex.Xst=="Dead" ~ Ns( tfe, knots=t.kn ) +
+                   age + factor( sex ),
+                   offset = log(lex.dur),
+                   family=poisson, data=Lung.s, eps=10^-8, maxit=25 )
```

```
> nd <- data.frame( tfe=seq(10,1000,10)-5, age=60, sex=1 )
> lambda <- ci.pred( mLS.pois.sp, nd )
> survP <- ci.surv( mLS.pois.sp, nd, int=10 )
```

Code and output for the entire example available in  
<http://bendixcarstensen.com/AdvCoh/WNtCMA/>

32/ 47

## What the Cox-model really is

Taking the life-table approach *ad absurdum* by:

- ▶ dividing time very finely and
- ▶ modeling one covariate, the time-scale, with one parameter per distinct value.
- ▶ the **model** for the time scale is really with exchangeable time-intervals.
- ▶ ⇒ difficult to access the baseline hazard (which looks terrible)
- ▶ ⇒ uninitiated tempted to show survival curves where irrelevant

33/ 47

## Models of this world

- ▶ Replace the  $\alpha_t$ s by a parametric function  $f(t)$  with a limited number of parameters, for example:
  - ▶ Piecewise constant
  - ▶ Splines (linear, quadratic or cubic)
  - ▶ Fractional polynomials
- ▶ the two latter brings model into “this world”:
  - ▶ smoothly varying rates
  - ▶ parametric closed form representation of baseline hazard
  - ▶ finite no. of parameters
- ▶ Makes it really easy to use rates directly in calculations of
  - ▶ expected residual life time
  - ▶ state occupancy probabilities in multistate models
  - ▶ ...

34/ 47

## The baseline hazard and survival functions

Using a parametric function to model the baseline hazard gives the possibility to plot this with confidence intervals for a given set of covariate values,  $x_0$

The survival function in a multiplicative Poisson model has the form:

$$S(t) = \exp\left(-\sum_{\tau < t} \exp(g(\tau) + x_0' \gamma)\right)$$

This is just a non-linear function of the parameters in the model,  $g$  and  $\gamma$ . So the variance can be computed using the  $\delta$ -method.

35/ 47

## $\delta$ -method for survival function

1. Select timepoints  $t_i$  (fairly close).
2. Get estimates of log-rates  $f(t_i) = g(t_i) + x_0' \gamma$  for these points:

$$\hat{f}(t_i) = \mathbf{B} \hat{\beta}$$

where  $\beta$  is the total parameter vector in the model.

3. Variance-covariance matrix of  $\hat{\beta}$ :  $\hat{\Sigma}$ .
4. Variance-covariance of  $\hat{f}(t_i)$ :  $\mathbf{B} \hat{\Sigma} \mathbf{B}'$ .
5. Transformation to the rates is the coordinate-wise exponential function, with derivative  $\text{diag}[\exp(\hat{f}(t_i))]$

36/ 47

6. Variance-covariance matrix of the rates at the points  $t_i$ :

$$\text{diag}(e^{\hat{f}(t_i)}) \mathbf{B} \hat{\Sigma} \mathbf{B}' \text{diag}(e^{\hat{f}(t_i)})'$$

7. Transformation to cumulative hazard ( $\ell$  is interval length):

$$\ell \times \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} e^{\hat{f}(t_1)} \\ e^{\hat{f}(t_2)} \\ e^{\hat{f}(t_3)} \\ e^{\hat{f}(t_4)} \end{bmatrix} = \mathbf{L} \begin{bmatrix} e^{\hat{f}(t_1)} \\ e^{\hat{f}(t_2)} \\ e^{\hat{f}(t_3)} \\ e^{\hat{f}(t_4)} \end{bmatrix}$$

37/ 47

8. Variance-covariance matrix for the cumulative hazard is:

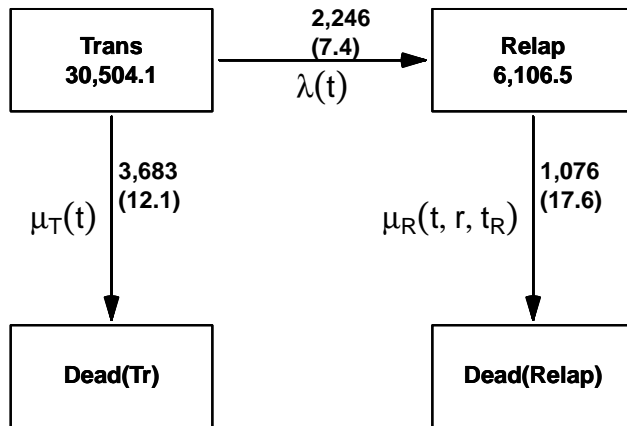
$$\mathbf{L} \text{diag}(e^{\hat{f}(t_i)}) \mathbf{B} \hat{\Sigma} \mathbf{B}' \text{diag}(e^{\hat{f}(t_i)})' \mathbf{L}'$$

This is all implemented in the `ci.cum()` function in Epi.

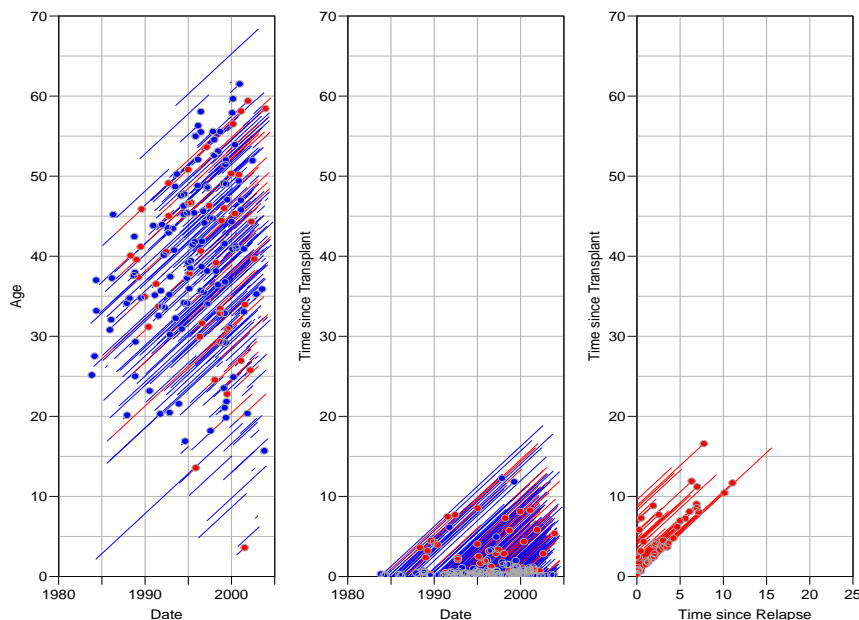
38/ 47

# EBMT transplant data

Iacobelli & Carstensen: Multistate Models with Multiple Timescales, Stat Med 2013, [4]



39/47 other covariates: Age and date at Tx, sex, donor type, CML type



40/ 47

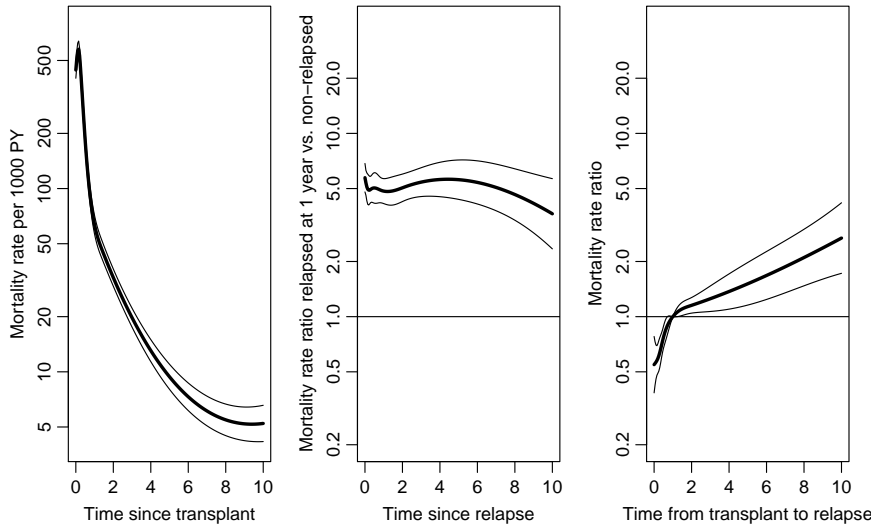
## Markov property: Empirical question

Model for mortality rates (with and without relapse):

- ▶  $t$  time since transplant
- ▶  $r$  time since relapse (if relapsed)
- ▶  $t_r$  time from transplant to relapse
- ▶ Fit the model for all transitions:
  - ▶ split follow-up time
  - ▶ fit Poisson model with covariates
  - ▶ and spline terms for each time scale.
- ▶ Lexis machinery [5, 6] from the **Epi** package for **R**
- ▶ ... for representation and manipulation of follow-up data.

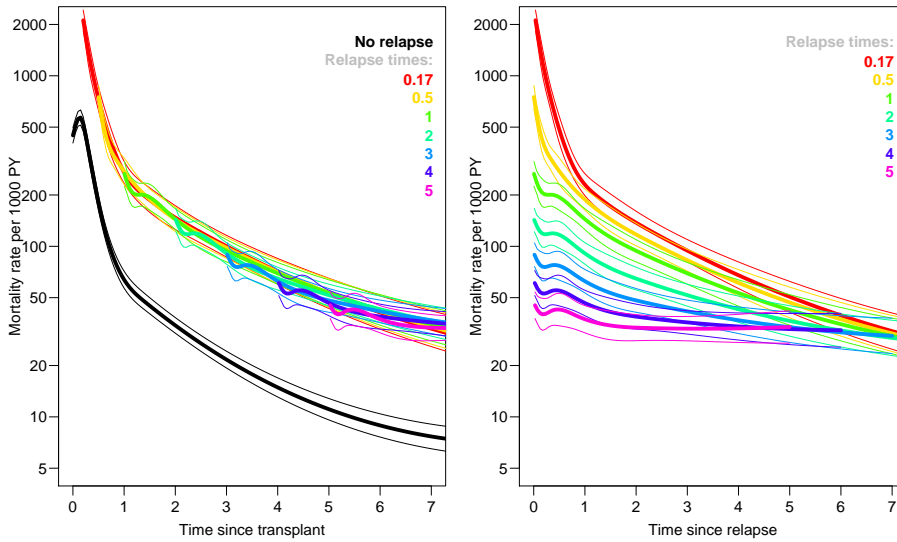
41/ 47

$$\log(\mu) = h(t) + k(r) + g(t - r) + X\beta$$



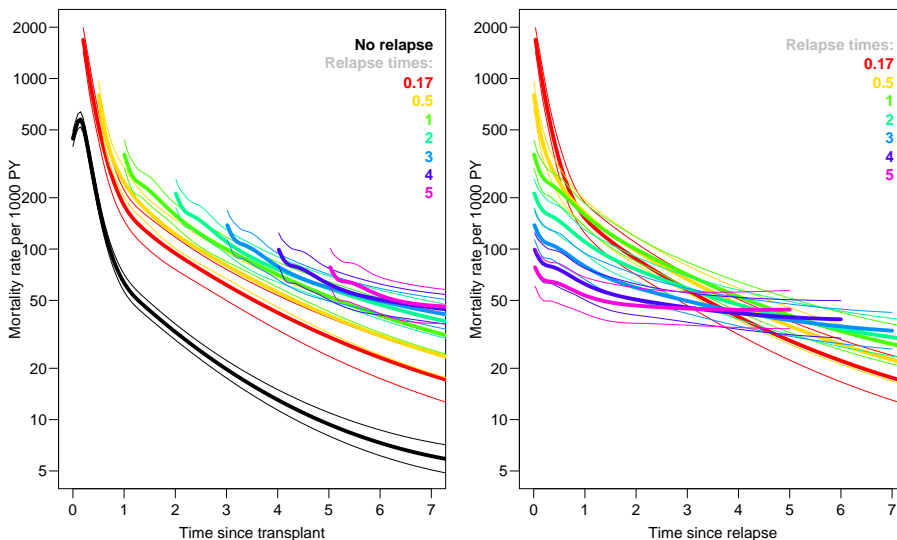
42/ 47  $t$ : time since transplant  $r$ : time since relapse

$$\log(\mu) = h(t) + k(r) + X\beta$$



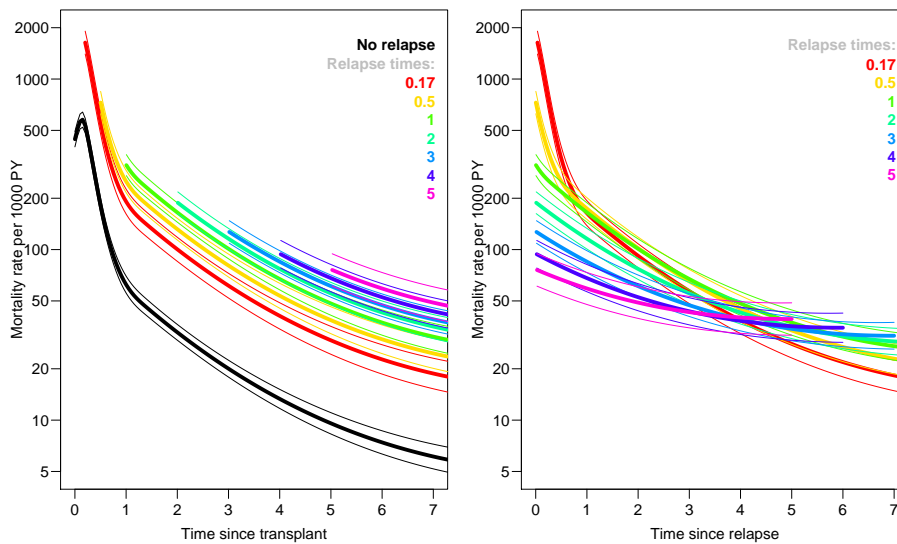
43/ 47  $t$ : time since transplant  $r$ : time since relapse

$$\log(\mu) = h(t) + k(r) + g(t - r) + X\beta$$



44/ 47  $t$ : time since transplant  $r$ : time since relapse






$$\log(\mu) = h(t) + g(t - r) + X\beta$$




45/ 47

$t$ : time since transplant     $r$ : time since relapse

## References I

-  P. K. Andersen and N. Keiding.  
Interpretability and importance of functionals in competing risks and multistate models.  
*Stat Med*, 31:1074–1088, 2012.
-  T R Holford.  
Life table with concomitant information.  
*Biometrics*, 32:587–597, 1976.
-  Whitehead J.  
Fitting Cox's regression model to survival data using GLIM.  
*Applied Statistics*, 29(3):268–275, 1980.
-  S. Iacobelli and B. Carstensen.  
Multiple time scales in multi-state models.  
*Stat Med*, 32(30):5315–5327, Dec 2013.
-  Martyn Plummer and Bendix Carstensen.  
Lexis: An R class for epidemiological studies with long-term follow-up.  
*Journal of Statistical Software*, 38(5):1–12, 1 2011.

## References II

-  Bendix Carstensen and Martyn Plummer.  
Using Lexis objects for multi-state models in R.  
*Journal of Statistical Software*, 38(6):1–18, 1 2011.

Direct link to these slides and to a document with details is at:  
[bendixcarstensen.com](http://bendixcarstensen.com)

Examples of this type of modeling at:  
[bendixcarstensen.com/AdvCoh/Lexis-ex](http://bendixcarstensen.com/AdvCoh/Lexis-ex)

**Thanks for your attention**