

Survival models and Cox-regression

Bendix Carstensen Steno Diabetes Center Copenhagen,
Gentofte, Denmark
& Department of Biostatistics,
University of Copenhagen
b@bxc.dk
<http://BendixCarstensen.com>

IDEG 2017 training day, Abu Dhabi,
11 December 2017
<http://BendixCarstensen/Epi/Courses/IDEG2017>

From /home/bendix/edc/conf/IDEG2017/teach/slides/slides.tex



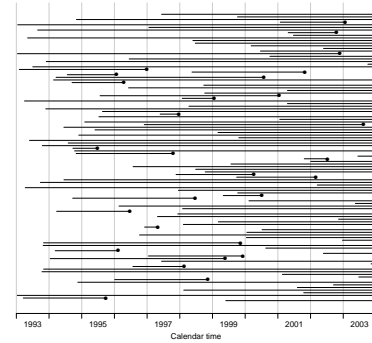
Steno Diabetes Center
Copenhagen

Rates and Survival
Lifetable estimators
Kaplan-Meier estimators
The Cox-model
Who needs the Cox-model anyway?
Multiple time scales and continuous rates

Each line a person

Each blob a death

Study ended at 31 Dec. 2003



Survival models and Cox-regression
Bendix Carstensen

Rates and Survival
Lifetable estimators
Kaplan-Meier estimators
The Cox-model
Who needs the Cox-model anyway?
Multiple time scales and continuous rates

Rates and Survival

Bendix Carstensen
Senior Statistician, Steno Diabetes Center

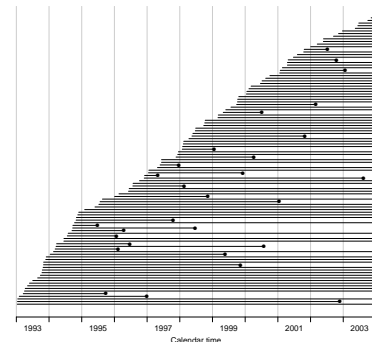
Survival models and Cox-regression
IDEG 2017 training day, Abu Dhabi,
11 December 2017

<http://BendixCarstensen/Epi/Courses/IDEG2017>

surv-rate

Ordered by date of entry

Most likely the order in your database.



Survival models and Cox-regression
Bendix Carstensen

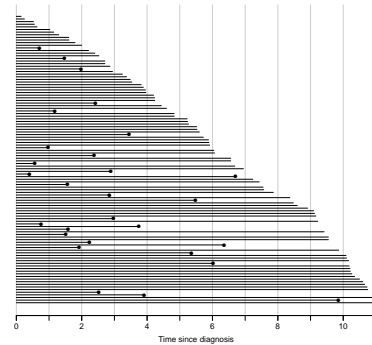
Rates and Survival
Lifetable estimators
Kaplan-Meier estimators
The Cox-model
Who needs the Cox-model anyway?
Multiple time scales and continuous rates

Survival data

Persons enter the study at some date.
Persons exit at a later date, either dead or alive.
Observation:
Actual time span to death ("event")
or
Some time alive ("at least this long")

Survival models and Cox-regression
Bendix Carstensen
Rates and Survival
Lifetable estimators
Kaplan-Meier estimators
The Cox-model
Who needs the Cox-model anyway?
Multiple time scales and continuous rates

Timescale changed to "Time since diagnosis".



Survival models and Cox-regression
Bendix Carstensen

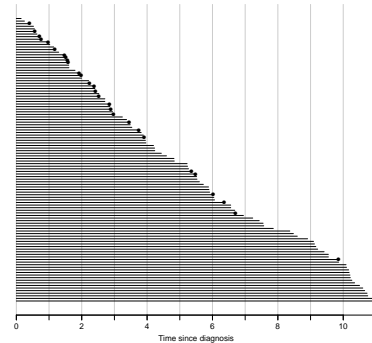
Rates and Survival
Lifetable estimators
Kaplan-Meier estimators
The Cox-model
Who needs the Cox-model anyway?
Multiple time scales and continuous rates

Examples of time-to-event measurements

- ▶ Time from diagnosis of cancer to death.
- ▶ Time from randomisation to death in a cancer clinical trial
- ▶ Time from HIV infection to AIDS.
- ▶ Time from marriage to 1st child birth.
- ▶ Time from marriage to divorce.
- ▶ Time to re-offending after being released from jail

Survival models and Cox-regression
Bendix Carstensen
Rates and Survival
Lifetable estimators
Kaplan-Meier estimators
The Cox-model
Who needs the Cox-model anyway?
Multiple time scales and continuous rates

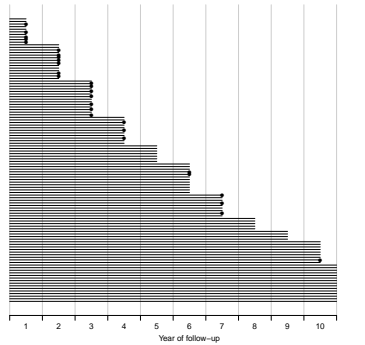
Patients ordered by survival time.



Survival models and Cox-regression
Bendix Carstensen

Rates and Survival
Lifetable estimators
Kaplan-Meier estimators
The Cox-model
Who needs the Cox-model anyway?
Multiple time scales and continuous rates

Survival times grouped into bands of survival.



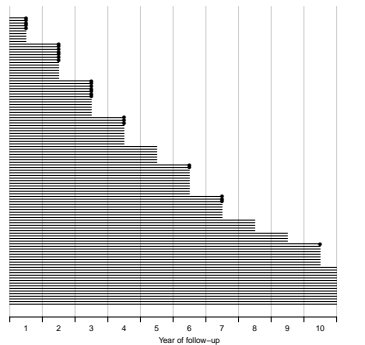
Intensity / rate / hazard — same same

- ▶ The **intensity** or **hazard function**
- ▶ Probability of event in interval, relative to interval length:

$$\lambda(t) = P \{ \text{event in } (t, t + h] \mid \text{alive at } t \} / h$$

- ▶ Characterizes the distribution of survival times as does f (density) or F (cumulative distribution).
- ▶ Theoretical counterpart of a(n empirical) **rate**.

Patients ordered by survival status within each band.



Rate and survival

$$S(t) = \exp \left(- \int_0^t \lambda(s) ds \right) \quad \lambda(t) = \frac{S'(t)}{S(t)}$$

Survival is a *cumulative* measure, the rate is an *instantaneous* measure.

Note: A cumulative measure requires an origin!
... it is always survival **since** some timepoint.

Survival after Cervix cancer

Year	Stage I			Stage II		
	<i>N</i>	<i>D</i>	<i>L</i>	<i>N</i>	<i>D</i>	<i>L</i>
1	110	5	5	234	24	3
2	100	7	7	207	27	11
3	86	7	7	169	31	9
4	72	3	8	129	17	7
5	61	0	7	105	7	13
6	54	2	10	85	6	6
7	42	3	6	73	5	6
8	33	0	5	62	3	10
9	28	0	4	49	2	13
10	24	1	8	34	4	6

Estimated risk in year 1 for Stage I women is $5/107.5 = 0.0465$

Estimated 1 year survival is $1 - 0.0465 = 0.9535$

Lifetable estimator.

Observed survival and rate

- ▶ **Survival studies:**

Observation of (right censored) survival time:

$$X = \min(T, Z), \quad \delta = 1\{X = T\}$$

— sometimes conditional on $T > t_0$ (left truncation, delayed entry).

- ▶ **Epidemiological studies:**

Observation of (components of) a rate:

$$D/Y$$

D: no. events, *Y* no of person-years, in a prespecified time-frame.

Survival function

Persons enter at time 0:

Date of birth, date of randomization, date of diagnosis.

How long do they survive?

Survival time T — a stochastic variable.

Distribution is characterized by the survival function:

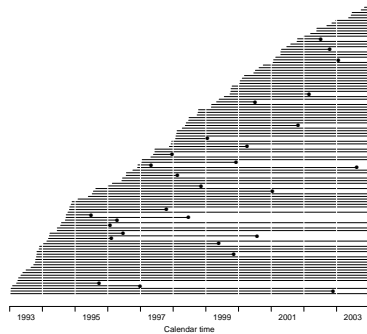
$$\begin{aligned} S(t) &= P \{ \text{survival at least till } t \} \\ &= P \{ T > t \} = 1 - P \{ T \leq t \} = 1 - F(t) \end{aligned}$$

$F(t)$ is the cumulative risk of death before time t .

Empirical rates for individuals

- ▶ At the *individual* level we introduce the **empirical rate**: (d, y) ,
— number of events ($d \in \{0, 1\}$) during y risk time.
- ▶ A person contributes several observations of (d, y) , with associated covariate values.
- ▶ Empirical rates are **responses** in survival analysis.
- ▶ The timescale t is a **covariate** — varies within each individual:
 t : age, time since diagnosis, calendar time.
- ▶ Don't confuse with y — difference between two points on **any** timescale we may choose.

Empirical rates by calendar time.



Rates and Survival (surv-rate)

Survival models and Cox-regression
Bendix Carstensen

Rates and Survival

Lifetable estimators

Kaplan-Meier estimators

The Cox-model

Who needs the Cox-model anyway?

Multiple time scales and continuous rates

16 / 94

Poisson likelihood

The log-likelihood contributions from follow-up of **one** individual:

$$d_i \log(\lambda(t)) - \lambda(t) y_t, \quad t = t_1, \dots, t_n$$

is also the log-likelihood from several independent Poisson observations with mean $\lambda(t) y_t$, i.e. \log -mean $\log(\lambda(t)) + \log(y_t)$

Analysis of the rates, (λ) can be based on a Poisson model with log-link applied to empirical rates where:

- ▶ d is the response variable.
- ▶ $\log(\lambda)$ is modelled by covariates
- ▶ $\log(y)$ is the offset variable.

Rates and Survival (surv-rate)

Survival models and Cox-regression
Bendix Carstensen

Rates and Survival

Lifetable estimators

Kaplan-Meier estimators

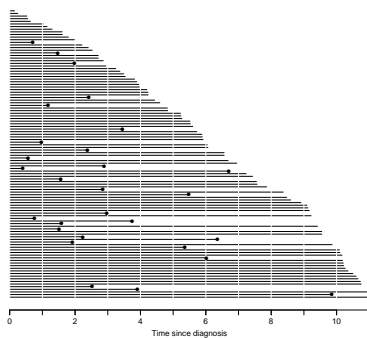
The Cox-model

Who needs the Cox-model anyway?

Multiple time scales and continuous rates

20 / 94

Empirical rates by time since diagnosis.



Rates and Survival (surv-rate)

Survival models and Cox-regression
Bendix Carstensen

Rates and Survival

Lifetable estimators

Kaplan-Meier estimators

The Cox-model

Who needs the Cox-model anyway?

Multiple time scales and continuous rates

17 / 94

Likelihood for follow-up of many persons

Adding empirical rates over the follow-up of persons:

$$D = \sum d \quad Y = \sum y \quad \Rightarrow \quad D \log(\lambda) - \lambda Y$$

- ▶ Persons are assumed independent
- ▶ Contribution from the same person are **conditionally** independent, hence give separate contributions to the log-likelihood.
- ▶ Therefore equivalent to likelihood for independent Poisson variates
- ▶ No need to correct for dependent observations; the likelihood is a product.

Rates and Survival (surv-rate)

Survival models and Cox-regression
Bendix Carstensen

Rates and Survival

Lifetable estimators

Kaplan-Meier estimators

The Cox-model

Who needs the Cox-model anyway?

Multiple time scales and continuous rates

21 / 94

Statistical inference: Likelihood

Two things needed:

- ▶ **Data** — what did we actually observe
Follow-up for each person:
Entry time, exit time, exit status, covariates
- ▶ **Model** — how was data generated
Rates as a function of time:
Probability machinery that generated data

Likelihood is the probability of observing the **data**, assuming the **model** is correct.

Maximum likelihood estimation is choosing **parameters** of the model that makes the likelihood maximal.

Rates and Survival (surv-rate)

Survival models and Cox-regression
Bendix Carstensen

Rates and Survival

Lifetable estimators

Kaplan-Meier estimators

The Cox-model

Who needs the Cox-model anyway?

Multiple time scales and continuous rates

18 / 94

Likelihood

Probability of the data and the parameter:

Assuming the rate (intensity) is constant, λ , the probability of observing 7 deaths in the course of 500 person-years:

$$\begin{aligned} P\{D = 7, Y = 500 | \lambda\} &= \lambda^D e^{-\lambda Y} \times K \\ &= \lambda^7 e^{-\lambda 500} \times K \\ &= L(\lambda | \text{data}) \end{aligned}$$

Best guess of λ is where this function is as large as possible.

Confidence interval is where it is not too far from the maximum

Rates and Survival (surv-rate)

Survival models and Cox-regression
Bendix Carstensen

Rates and Survival

Lifetable estimators

Kaplan-Meier estimators

The Cox-model

Who needs the Cox-model anyway?

Multiple time scales and continuous rates

22 / 94

Likelihood from one person

The likelihood from several empirical rates from one individual is a product of conditional probabilities:

$$\begin{aligned} P\{\text{event at } t_4 | t_0\} &= P\{\text{survive}(t_0, t_1) | \text{alive at } t_0\} \times \\ &P\{\text{survive}(t_1, t_2) | \text{alive at } t_1\} \times \\ &P\{\text{survive}(t_2, t_3) | \text{alive at } t_2\} \times \\ &P\{\text{event at } t_4 | \text{alive at } t_3\} \end{aligned}$$

Log-likelihood from one individual is a sum of terms.

Each term refers to one empirical rate (d, y)

— $y = t_i - t_{i-1}$ and mostly $d = 0$.

t_i is the timescale (covariate).

Rates and Survival (surv-rate)

Survival models and Cox-regression
Bendix Carstensen

Rates and Survival

Lifetable estimators

Kaplan-Meier estimators

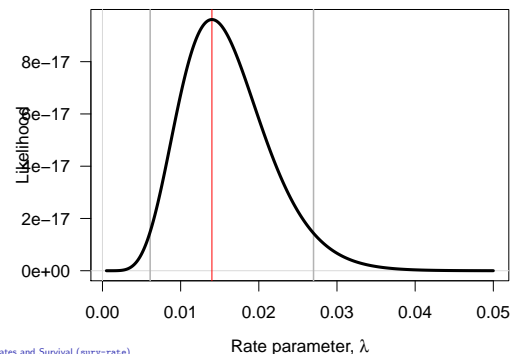
The Cox-model

Who needs the Cox-model anyway?

Multiple time scales and continuous rates

19 / 94

Likelihood function



Rates and Survival (surv-rate)

Survival models and Cox-regression
Bendix Carstensen

Rates and Survival

Lifetable estimators

Kaplan-Meier estimators

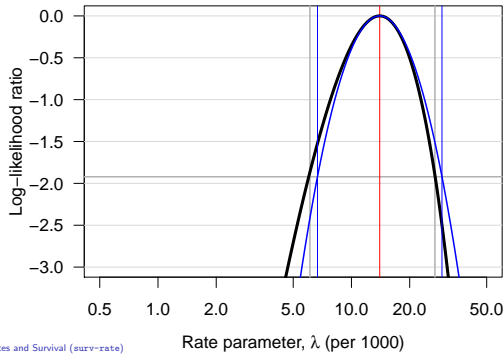
The Cox-model

Who needs the Cox-model anyway?

Multiple time scales and continuous rates

23 / 94

Likelihood function



Rates and Survival (surv-rate)

Survival models and Cox-regression
Bendix Carstensen

Rates and Survival
Lifetable estimators
Kaplan-Meier estimators
The Cox-model
Who needs the Cox-model anyway?
Multiple time scales and continuous rates

23 / 94

Survival analysis

- ▶ Response variable: Time to event, T
- ▶ Censoring time, Z
- ▶ We observe $(\min(T, Z), \delta = 1\{T < Z\})$.
- ▶ This gives time a special status, and mixes the response variable (risk)time with the covariate time(scale).
- ▶ Originates from clinical trials where everyone enters at time 0, and therefore $Y = T - 0 = T$

Lifetable estimators (1tab)

Survival models and Cox-regression
Bendix Carstensen

Rates and Survival
Lifetable estimators
Kaplan-Meier estimators
The Cox-model
Who needs the Cox-model anyway?
Multiple time scales and continuous rates

26 / 94

Example using R

Poisson likelihood, for one rate, based on 17 events in 843.7 PY:

```
library( Epi )
D <- 17 ; Y <- 843.7
m1 <- glm( D ~ 1, offset=log(Y/1000), family=poisson )
ci.exp( m1 )

exp(Est.)      2.5%      97.5%
(Intercept) 20.14934 12.52605 32.41213
```

Poisson likelihood, two rates, or one rate and RR:

```
D <- c(17,28) ; Y <- c(843.7,632.3) ; gg <- factor(0:1)
m2 <- glm( D ~ gg, offset=log(Y/1000), family=poisson )
ci.exp( m2 )

exp(Est.)      2.5%      97.5%
(Intercept) 20.149342 12.526051 32.412130
gg1          2.197728  1.202971  4.015068
```

Rates and Survival (surv-rate)

Survival models and Cox-regression
Bendix Carstensen

Rates and Survival
Lifetable estimators
Kaplan-Meier estimators
The Cox-model
Who needs the Cox-model anyway?
Multiple time scales and continuous rates

24 / 94

The life table method

The simplest analysis is by the "life-table method":

interval	alive	dead	cens.	
i	n_i	d_i	l_i	p_i
1	77	5	2	$5/(77 - 2/2) = 0.066$
2	70	7	4	$7/(70 - 4/2) = 0.103$
3	59	8	1	$8/(59 - 1/2) = 0.137$

$$p_i = P\{\text{death in interval } i\} = d_i / (n_i - l_i/2)$$

$$S(t) = (1 - p_1) \times \dots \times (1 - p_t)$$

Lifetable estimators (1tab)

Survival models and Cox-regression
Bendix Carstensen

Rates and Survival
Lifetable estimators
Kaplan-Meier estimators
The Cox-model
Who needs the Cox-model anyway?
Multiple time scales and continuous rates

27 / 94

Example using R

Poisson likelihood, two rates, or one rate and RR:

```
D <- c(17,28) ; Y <- c(843.7,632.3) ; gg <- factor(0:1)
m2 <- glm( D ~ gg, offset=log(Y/1000), family=poisson )
ci.exp( m2 )

exp(Est.)      2.5%      97.5%
(Intercept) 20.149342 12.526051 32.412130
gg1          2.197728  1.202971  4.015068

m3 <- glm( D ~ gg - 1, offset=log(Y/1000), family=poisson )
ci.exp( m3 )

exp(Est.)      2.5%      97.5%
gg0            20.14934 12.52605 32.41213
gg1            44.28278 30.57545 64.13525
```

Rates and Survival (surv-rate)

Survival models and Cox-regression
Bendix Carstensen

Rates and Survival
Lifetable estimators
Kaplan-Meier estimators
The Cox-model
Who needs the Cox-model anyway?
Multiple time scales and continuous rates

25 / 94

Population life table, DK 1997–98

a	Men			Women		
	$S(a)$	$\lambda(a)$	$E[l_{res}(a)]$	$S(a)$	$\lambda(a)$	$E[l_{res}(a)]$
0	1.00000	567	73.68	1.00000	474	78.65
1	0.99433	67	73.10	0.99526	47	78.02
2	0.99366	38	72.15	0.99479	21	77.06
3	0.99329	25	71.18	0.99458	14	76.08
4	0.99304	25	70.19	0.99444	14	75.09
5	0.99279	21	69.21	0.99430	11	74.10
6	0.99258	17	68.23	0.99419	6	73.11
7	0.99242	14	67.24	0.99413	3	72.11
8	0.99227	15	66.25	0.99410	6	71.11
9	0.99213	14	65.26	0.99404	9	70.12
10	0.99199	17	64.26	0.99395	17	69.12
11	0.99181	19	63.28	0.99378	15	68.14
12	0.99162	16	62.29	0.99363	11	67.15
13	0.99147	18	61.30	0.99352	14	66.15
14	0.99129	25	60.31	0.99338	11	65.16
15	0.99104	45	59.32	0.99327	10	64.17
16	0.99059	50	58.35	0.99317	18	63.18
17	0.99009	52	57.38	0.99299	29	62.19
18	0.98957	85	56.41	0.99270	35	61.21
19	0.98873	79	55.46	0.99235	30	60.23
20	0.98795	70	54.50	0.99205	35	59.24
21	0.98726	71	53.54	0.99170	31	58.27

Lifetable estimators (1tab)

Survival models and Cox-regression
Bendix Carstensen

Rates and Survival
Lifetable estimators
Kaplan-Meier estimators
The Cox-model
Who needs the Cox-model anyway?
Multiple time scales and continuous rates

28 / 94

Lifetable estimators

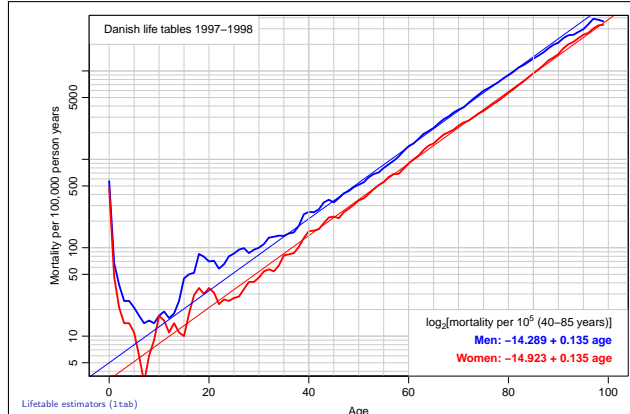
Bendix Carstensen

Senior Statistician, Steno Diabetes Center

Survival models and Cox-regression
IDEG 2017 training day, Abu Dhabi,
11 December 2017

<http://BendixCarstensen/Epi/Courses/IDEG2017>

1tab

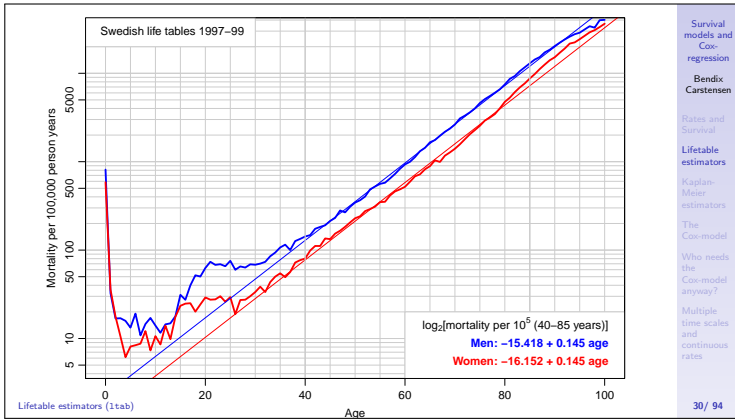


Lifetable estimators (1tab)

Survival models and Cox-regression
Bendix Carstensen

Rates and Survival
Lifetable estimators
Kaplan-Meier estimators
The Cox-model
Who needs the Cox-model anyway?
Multiple time scales and continuous rates

29 / 94



Survival models and Cox-regression
Bendix Carstensen

Rates and Survival
Lifetable estimators
Kaplan-Meier estimators
The Cox-model
Who needs the Cox-model anyway?
Multiple time scales and continuous rates

30/ 94

Life table approach

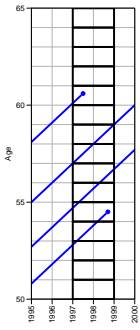
- ▶ The **population** experience:
 - D : Deaths (events).
 - Y : Person-years (risk time).
- ▶ The classical lifetable analysis compiles these for prespecified intervals of age, and computes age-specific mortality **rates**.
- ▶ Data are collected crosssectionally, but interpreted longitudinally.
- ▶ The **rates** are the basic building blocks — used for construction of:
 - ▶ RRs
 - ▶ cumulative measures (survival and risk)

Survival models and Cox-regression
Bendix Carstensen

Rates and Survival
Lifetable estimators
Kaplan-Meier estimators
The Cox-model
Who needs the Cox-model anyway?
Multiple time scales and continuous rates

34/ 94

Observations for the lifetable



Life table is based on person-years and deaths accumulated in a short period.
Age-specific rates — cross-sectional!
Survival function:
$$S(t) = e^{-\int_0^t \lambda(a) da} = e^{-\sum_0^t \lambda(a)}$$

— assumes stability of rates to be interpretable for actual persons.

Survival models and Cox-regression
Bendix Carstensen

Rates and Survival
Lifetable estimators
Kaplan-Meier estimators
The Cox-model
Who needs the Cox-model anyway?
Multiple time scales and continuous rates

31/ 94

Kaplan-Meier estimators

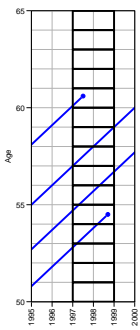
Bendix Carstensen
Senior Statistician, Steno Diabetes Center

Survival models and Cox-regression
IDEG 2017 training day, Abu Dhabi,
11 December 2017

<http://BendixCarstensen/Epi/Courses/IDEG2017>

km-na

Observations for the lifetable



This is a **Lexis** diagram.



Survival models and Cox-regression
Bendix Carstensen

Rates and Survival
Lifetable estimators
Kaplan-Meier estimators
The Cox-model
Who needs the Cox-model anyway?
Multiple time scales and continuous rates

32/ 94

The Kaplan-Meier Method

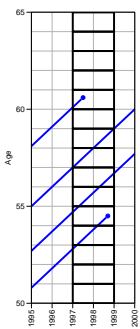
- ▶ The most common method of estimating the survival function.
- ▶ A non-parametric method.
- ▶ Divides time into small intervals where the intervals are defined by the unique times of failure (death).
- ▶ Based on conditional probabilities as we are interested in the probability a subject surviving the next time interval given that they have survived so far.

Survival models and Cox-regression
Bendix Carstensen

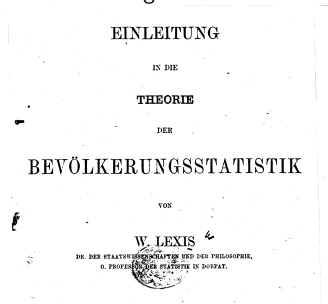
Rates and Survival
Lifetable estimators
Kaplan-Meier estimators
The Cox-model
Who needs the Cox-model anyway?
Multiple time scales and continuous rates

35/ 94

Observations for the lifetable



This is a **Lexis** diagram.



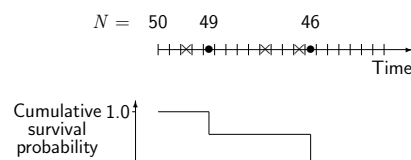
Survival models and Cox-regression
Bendix Carstensen

Rates and Survival
Lifetable estimators
Kaplan-Meier estimators
The Cox-model
Who needs the Cox-model anyway?
Multiple time scales and continuous rates

33/ 94

Kaplan-Meier method illustrated

(● = failure and × = censored):



- ▶ Steps caused by multiplying by $(1 - 1/49)$ and $(1 - 1/46)$ respectively
- ▶ Late entry can also be dealt with

Kaplan-Meier estimators (km-na)

Survival models and Cox-regression
Bendix Carstensen

Rates and Survival
Lifetable estimators
Kaplan-Meier estimators
The Cox-model
Who needs the Cox-model anyway?
Multiple time scales and continuous rates

36/ 94

Using R: Surv()

```
library( survival )
data( lung )
head( lung, 3 )

  inst time status age sex ph.ecog ph.karno pat.karno meal.cal wt.loss
1     3  306     2  74  1     1     90     100    1175     NA
2     3  455     2  68  1     0     90     90     1225     15
3     3 1010     1  56  1     0     90     90     NA     15

with( lung, Surv( time, status==2 ) )[1:10]
[1] 306 455 1010+ 210 883 1022+ 310 361 218 166
( s.km <- survfit( Surv( time, status==2 ) ~ 1, data=lung ) )
Call: survfit(formula = Surv(time, status == 2) ~ 1, data = lung)

      n  events median 0.95LCL 0.95UCL
228    165    310    285    363

plot( s.km )
abline( v=310, h=0.5, col="red" )
```

Kaplan-Meier estimators (km-na)

Survival models and Cox-regression
Bendix Carstensen

Rates and Survival
Lifetable estimators
Kaplan-Meier estimators
The Cox-model
Who needs the Cox-model anyway?
Multiple time scales and continuous rates

37 / 94

The proportional hazards model

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

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

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

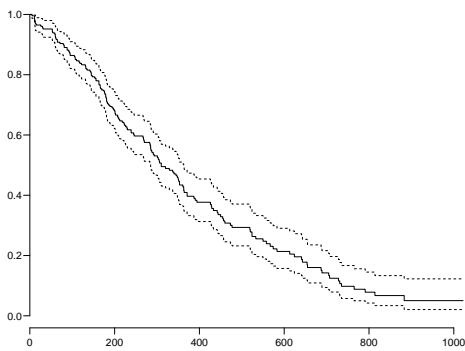
- ▶ This is David Cox's invention.
- ▶ Extremely efficient from a computational point of view.
- ▶ The baseline hazard $\lambda_0(t)$ is bypassed (profiled out).

The Cox-model (cox)

Survival models and Cox-regression
Bendix Carstensen

Rates and Survival
Lifetable estimators
Kaplan-Meier estimators
The Cox-model
Who needs the Cox-model anyway?
Multiple time scales and continuous rates

40 / 94



Kaplan-Meier estimators (km-na)

Survival models and Cox-regression
Bendix Carstensen

Rates and Survival
Lifetable estimators
Kaplan-Meier estimators
The Cox-model
Who needs the Cox-model anyway?
Multiple time scales and continuous rates

38 / 94

Proportional Hazards model

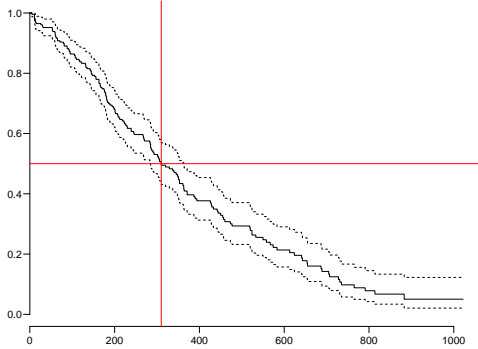
- ▶ The baseline hazard rate, $\lambda_0(t)$, is the hazard rate when all the covariates are 0.
- ▶ The form of the above equation means that covariates act **multiplicatively** on the baseline hazard rate.
- ▶ Time is a covariate (albeit modeled special).
- ▶ The baseline hazard is a function of time and thus varies with time.
- ▶ No assumption about the shape of the underlying hazard function.
- ▶ — but you will never see the shape...

The Cox-model (cox)

Survival models and Cox-regression
Bendix Carstensen

Rates and Survival
Lifetable estimators
Kaplan-Meier estimators
The Cox-model
Who needs the Cox-model anyway?
Multiple time scales and continuous rates

41 / 94



Kaplan-Meier estimators (km-na)

Survival models and Cox-regression
Bendix Carstensen

Rates and Survival
Lifetable estimators
Kaplan-Meier estimators
The Cox-model
Who needs the Cox-model anyway?
Multiple time scales and continuous rates

39 / 94

Interpreting Regression Coefficients

- ▶ If x_j is binary $\exp(\beta_j)$ is the estimated hazard ratio for subjects corresponding to $x_j = 1$ compared to those where $x_j = 0$.
- ▶ If x_j is continuous $\exp(\beta_j)$ is the estimated increase/decrease in the hazard rate for a unit change in x_j .
- ▶ With more than one covariate interpretation is similar, i.e. $\exp(\beta_j)$ is the hazard ratio for subjects who **only** differ with respect to covariate x_j .

The Cox-model (cox)

Survival models and Cox-regression
Bendix Carstensen

Rates and Survival
Lifetable estimators
Kaplan-Meier estimators
The Cox-model
Who needs the Cox-model anyway?
Multiple time scales and continuous rates

42 / 94

The Cox-model

Bendix Carstensen

Senior Statistician, Steno Diabetes Center

Survival models and Cox-regression
IDEG 2017 training day, Abu Dhabi,
11 December 2017

<http://BendixCarstensen/Epi/Courses/IDEG2017>

cox

Fitting a Cox- model in R

```
library( survival )
data(bladder)
bladder <- subset( bladder, enum<2 )
head( bladder )

  id rx number size stop event enum
1  1  1     1    3    1    0    1
5  2  1     2    4    1    0    1
9  3  1     1    1    7    0    1
13 4  1     5   10    0    1
17 5  1     4    6    1    1
21 6  1     1    1   14    0    1
```

The Cox-model (cox)

Survival models and Cox-regression
Bendix Carstensen

Rates and Survival
Lifetable estimators
Kaplan-Meier estimators
The Cox-model
Who needs the Cox-model anyway?
Multiple time scales and continuous rates

43 / 94

Fitting a Cox-model in R

```
c0 <- coxph( Surv(stop,event) ~ number + size, data=bladder )
c0
```

Call:
coxph(formula = Surv(stop, event) ~ number + size, data = bladder)

	coef	exp(coef)	se(coef)	z	p
number	0.2049	1.2274	0.0704	2.91	0.0036
size	0.0613	1.0633	0.1033	0.59	0.5525

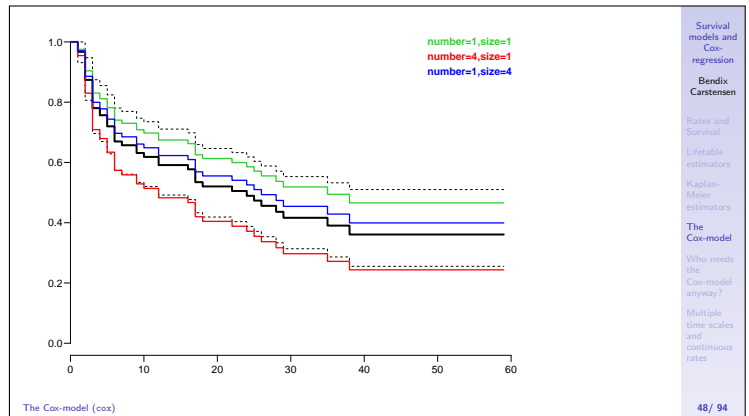
Likelihood ratio test=7.04 on 2 df, p=0.0296
n= 85, number of events= 47

The Cox-model (cox)

Survival models and Cox-regression
Bendix Carstensen

Rates and Survival
Lifetable estimators
Kaplan-Meier estimators
The Cox-model
Who needs the Cox-model anyway?
Multiple time scales and continuous rates

44 / 94



The Cox-model (cox)

Survival models and Cox-regression
Bendix Carstensen

Rates and Survival
Lifetable estimators
Kaplan-Meier estimators
The Cox-model
Who needs the Cox-model anyway?
Multiple time scales and continuous rates

48 / 94

Plotting the base survival in R

```
plot( survfit(c0) )
lines( survfit(c0), conf.int=F, lwd=3 )
```

The `plot.coxph` plots the survival curve for a person with an average covariate value

— which is **not** the average survival for the population considered...

— and not necessarily meaningful

The Cox-model (cox)

Survival models and Cox-regression
Bendix Carstensen

Rates and Survival
Lifetable estimators
Kaplan-Meier estimators
The Cox-model
Who needs the Cox-model anyway?
Multiple time scales and continuous rates

45 / 94

Who needs the Cox-model anyway?

Bendix Carstensen

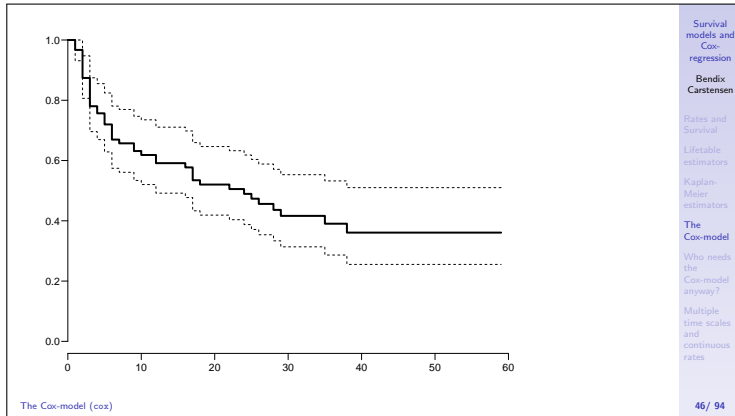
Senior Statistician, Steno Diabetes Center

Survival models and Cox-regression

IDEG 2017 training day, Abu Dhabi,
11 December 2017

<http://BendixCarstensen/Epi/Courses/IDEG2017>

KMCoX



The Cox-model (cox)

Survival models and Cox-regression
Bendix Carstensen

Rates and Survival
Lifetable estimators
Kaplan-Meier estimators
The Cox-model
Who needs the Cox-model anyway?
Multiple time scales and continuous rates

46 / 94

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 .

The covariate t has a special status:

- ▶ Computationally, because all individuals contribute to (some of) the range of t .
- ▶ ... the scale along which time is split (the risk sets)
- ▶ Conceptually t is just a covariate that varies within individual.
- ▶ Cox's approach profiles $\lambda_0(t)$ out from the model

Who needs the Cox-model anyway? (KMCoX)

Survival models and Cox-regression
Bendix Carstensen

Rates and Survival
Lifetable estimators
Kaplan-Meier estimators
The Cox-model
Who needs the Cox-model anyway?
Multiple time scales and continuous rates

49 / 94

Plotting the base survival in R

You can plot the survival curve for specific values of the covariates, using the `newdata=` argument:

```
plot( survfit(c0) )
lines( survfit(c0), conf.int=F, lwd=3 )
lines( survfit(c0, newdata=data.frame(number=1,size=1)),
      lwd=2, col="limegreen" )
text( par("usr")[2]*0.98, 1.00, "number=1,size=1",
      col="limegreen", font=2, adj=1 )
```

The Cox-model (cox)

Survival models and Cox-regression
Bendix Carstensen

Rates and Survival
Lifetable estimators
Kaplan-Meier estimators
The Cox-model
Who needs the Cox-model anyway?
Multiple time scales and continuous rates

47 / 94

The Cox-likelihood as profile likelihood

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

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

- ▶ Profile likelihood:
 - ▶ Derive estimates of α_t as function of data and β s — assuming constant rate between death times
 - ▶ Insert in likelihood, now only a function of data and β s
 - ▶ Turns out to be Cox's partial likelihood

Who needs the Cox-model anyway? (KMCoX)

Survival models and Cox-regression
Bendix Carstensen

Rates and Survival
Lifetable estimators
Kaplan-Meier estimators
The Cox-model
Who needs the Cox-model anyway?
Multiple time scales and continuous rates

50 / 94

The Cox-likelihood: mechanics of computing

- ▶ The likelihood is computed by summing over risk-sets at each event time t :

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

- ▶ 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

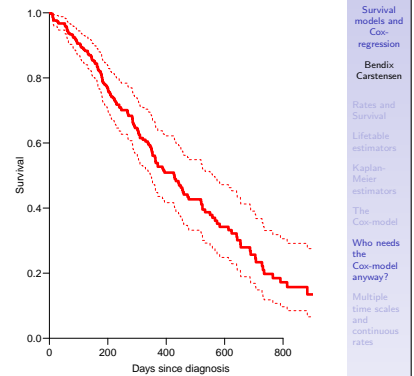
Who needs the Cox-model anyway? (R0Cox)

Survival models and Cox-regression
Bendix Carstensen

Rates and Survival
Lifetable estimators
Kaplan-Meier estimators
The Cox-model
Who needs the Cox-model anyway?
Multiple time scales and continuous rates

51 / 94

Mayo Clinic lung cancer 60 year old woman



Who needs the Cox-model anyway? (R0Cox)

Survival models and Cox-regression
Bendix Carstensen

Rates and Survival
Lifetable estimators
Kaplan-Meier estimators
The Cox-model
Who needs the Cox-model anyway?
Multiple time scales and continuous rates

55 / 94

$$\log(\lambda(t, x_i)) = \log(\lambda_0(t)) + \beta_1 x_{i1} + \dots + \beta_p x_{ip} = \alpha_t + \eta_i$$

- ▶ 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 $d_{it} = 0$.
- ▶ 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 α_t will be from:
 - ▶ the (only) empirical rate $(1, 1)$ with the death at time t .
 - ▶ all other empirical rates $(0, 1)$ from those at risk at time t .

Who needs the Cox-model anyway? (R0Cox)

Survival models and Cox-regression
Bendix Carstensen

Rates and Survival
Lifetable estimators
Kaplan-Meier estimators
The Cox-model
Who needs the Cox-model anyway?
Multiple time scales and continuous rates

52 / 94

Example: Mayo Clinic lung cancer I

```
> library( survival )
> library( Epi )
> 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.

Who needs the Cox-model anyway? (R0Cox)

Survival models and Cox-regression
Bendix Carstensen

Rates and Survival
Lifetable estimators
Kaplan-Meier estimators
The Cox-model
Who needs the Cox-model anyway?
Multiple time scales and continuous rates

56 / 94

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
 - ▶ Covariate value for the timescale (time since entry, current age, current date, ...)
 - ▶ other covariates

Who needs the Cox-model anyway? (R0Cox)

Survival models and Cox-regression
Bendix Carstensen

Rates and Survival
Lifetable estimators
Kaplan-Meier estimators
The Cox-model
Who needs the Cox-model anyway?
Multiple time scales and continuous rates

53 / 94

Example: Mayo Clinic lung cancer II

```
> mL.cox <- coxph( Surv( tfe, tfe+lex.dur, lex.Xst=="Dead" ) ~
+               age + factor( sex ),
+               method="breslow", eps=10^-8, iter.max=25, data=Lung )
> Lung.s <- splitLexis( Lung,
+                     breaks=c(0, sort(unique(Lung$time))),
+                     time.scale="tfe" )
> Lung.S <- splitLexis( Lung,
+                     breaks=c(0, sort(unique(Lung$time[Lung$lex.Xst=="Dead"]))),
+                     time.scale="tfe" )
> summary( Lung.s )
```

Transitions:

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

```
> summary( Lung.S )
```

Who needs the Cox-model anyway? (R0Cox)

Survival models and Cox-regression
Bendix Carstensen

Rates and Survival
Lifetable estimators
Kaplan-Meier estimators
The Cox-model
Who needs the Cox-model anyway?
Multiple time scales and continuous rates

57 / 94

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

Who needs the Cox-model anyway? (R0Cox)

Survival models and Cox-regression
Bendix Carstensen

Rates and Survival
Lifetable estimators
Kaplan-Meier estimators
The Cox-model
Who needs the Cox-model anyway?
Multiple time scales and continuous rates

54 / 94

Example: Mayo Clinic lung cancer III

Transitions:

From	To	Alive	Dead	Records:	Events:	Risk time:	Persons:
Alive	15916	165	16081	165	69593	228	

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

lex.id	tfe	lex.dur	lex.Cst	lex.Xst	inst	time	status	age	sex	ph.ecog
9235	96	0	5	Alive	Alive	12 30	2	72	1	2
9236	96	5	6	Alive	Alive	12 30	2	72	1	2
9237	96	11	1	Alive	Alive	12 30	2	72	1	2
9238	96	12	1	Alive	Alive	12 30	2	72	1	2
9239	96	13	2	Alive	Alive	12 30	2	72	1	2
9240	96	15	11	Alive	Alive	12 30	2	72	1	2
9241	96	26	4	Alive	Dead	12 30	2	72	1	2

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

```
[1] 186
```

Who needs the Cox-model anyway? (R0Cox)

Survival models and Cox-regression
Bendix Carstensen

Rates and Survival
Lifetable estimators
Kaplan-Meier estimators
The Cox-model
Who needs the Cox-model anyway?
Multiple time scales and continuous rates

58 / 94

Example: Mayo Clinic lung cancer IV

```
> system.time(
+ mLS.pois.fc <- glm( lex.Xst=="Dead" ~ - 1 + factor( tfe ) +
+                   age + factor( sex ),
+                   offset = log(lex.dur),
+                   family=poisson, data=Lung.s, eps=10^-8, maxit=25 )
+ )
user system elapsed
10.642 19.996 8.894

> length( coef(mLS.pois.fc) )
[1] 188

> system.time(
+ mLS.pois.fc <- glm( lex.Xst=="Dead" ~ - 1 + factor( tfe ) +
+                   age + factor( sex ),
+                   offset = log(lex.dur),
+                   family=poisson, data=Lung.S, eps=10^-8, maxit=25 )
+ )
```

Who needs the Cox-model anyway? (RXCox)

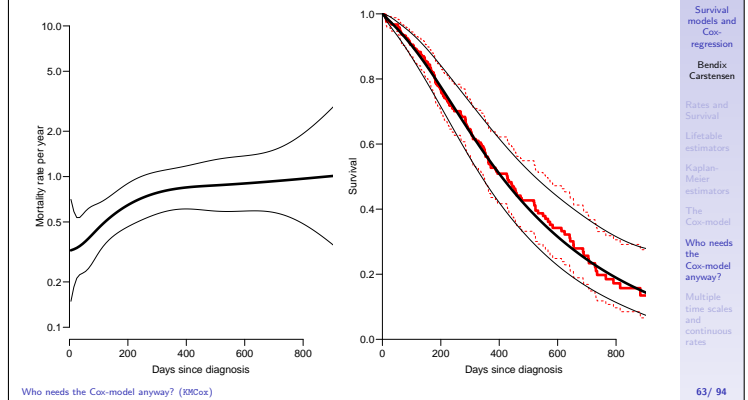
Survival models and Cox-regression
Bendix Carstensen

Rates and Survival
Lifetable estimators
Kaplan-Meier estimators
The Cox-model

Who needs the Cox-model anyway?

Multiple time scales and continuous rates

59 / 94



Who needs the Cox-model anyway? (RXCox)

Survival models and Cox-regression
Bendix Carstensen

Rates and Survival
Lifetable estimators
Kaplan-Meier estimators
The Cox-model

Who needs the Cox-model anyway?

Multiple time scales and continuous rates

63 / 94

Example: Mayo Clinic lung cancer V

```
user system elapsed
3.859 7.426 3.068

> length( coef(mLS.pois.fc) )
[1] 142

> 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( 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 )
+ )
```

Who needs the Cox-model anyway? (RXCox)

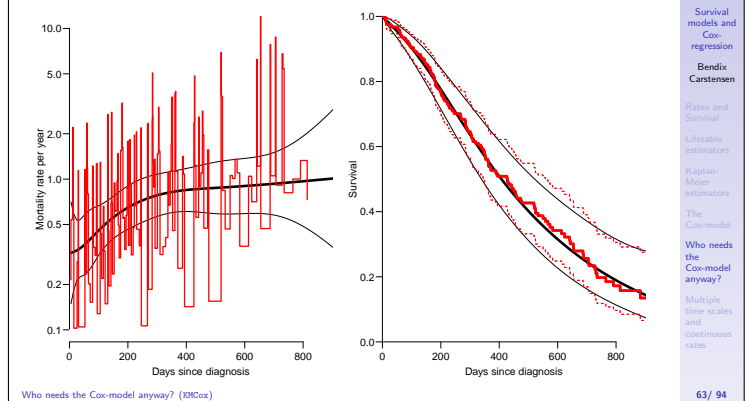
Survival models and Cox-regression
Bendix Carstensen

Rates and Survival
Lifetable estimators
Kaplan-Meier estimators
The Cox-model

Who needs the Cox-model anyway?

Multiple time scales and continuous rates

60 / 94



Who needs the Cox-model anyway? (RXCox)

Survival models and Cox-regression
Bendix Carstensen

Rates and Survival
Lifetable estimators
Kaplan-Meier estimators
The Cox-model

Who needs the Cox-model anyway?

Multiple time scales and continuous rates

63 / 94

Example: Mayo Clinic lung cancer VI

```
user system elapsed
0.413 0.642 0.341

> ests <-
+ rbind( ci.exp(mL.cox),
+        ci.exp(mLS.pois.fc, subset=c("age", "sex")),
+        ci.exp(mLS.pois.fc, subset=c("age", "sex")),
+        ci.exp(mLS.pois.sp, subset=c("age", "sex")) )
> cmp <- cbind( ests[c(1,3,5,7), ],
+              ests[c(1,3,5,7)+1, ] )
> rownames( cmp ) <- c("Cox", "Poisson-factor", "Poisson-factor (D)", "Poisson-spline")
> colnames( cmp )[c(1,4)] <- c("age", "sex")

> round( cmp, 7 )
```

Who needs the Cox-model anyway? (RXCox)

Survival models and Cox-regression
Bendix Carstensen

Rates and Survival
Lifetable estimators
Kaplan-Meier estimators
The Cox-model

Who needs the Cox-model anyway?

Multiple time scales and continuous rates

61 / 94

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 )

> CM <- cbind( 1, Ns( seq(10,1000,10)-5, knots=t.kn ), 60, 1 )
> lambda <- ci.exp( mLS.pois.sp, ctr.mat=CM )
> Lambda <- ci.cum( mLS.pois.sp, ctr.mat=CM, int1=10 )[, -4]
> survp <- exp(-rbind(0, Lambda))
```

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

Who needs the Cox-model anyway? (RXCox)

Survival models and Cox-regression
Bendix Carstensen

Rates and Survival
Lifetable estimators
Kaplan-Meier estimators
The Cox-model

Who needs the Cox-model anyway?

Multiple time scales and continuous rates

64 / 94

Example: Mayo Clinic lung cancer VII

	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-factor (D)	1.017332	0.9991211	1.035874	0.5984794	0.4310150	0.8310094
Poisson-spline	1.016189	0.9980329	1.034676	0.5998287	0.4319932	0.8328707

Who needs the Cox-model anyway? (RXCox)

Survival models and Cox-regression
Bendix Carstensen

Rates and Survival
Lifetable estimators
Kaplan-Meier estimators
The Cox-model

Who needs the Cox-model anyway?

Multiple time scales and continuous rates

62 / 94

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

Who needs the Cox-model anyway? (RXCox)

Survival models and Cox-regression
Bendix Carstensen

Rates and Survival
Lifetable estimators
Kaplan-Meier estimators
The Cox-model

Who needs the Cox-model anyway?

Multiple time scales and continuous rates

65 / 94

Models of this world

- Replace the α_i 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
 - ...

Who needs the Cox-model anyway? (R20x)

Survival models and Cox-regression
Bendix Carstensen
Rates and Survival
Lifetable estimators
Kaplan-Meier estimators
The Cox-model
Who needs the Cox-model anyway?
Multiple time scales and continuous rates
66/94

Linear effects in glm

How do rates depend on age?

```
> ml <- glm( D ~ A, offset=log(Y), family=poisson, data=testisDK )
> round( ci.lin( ml ), 4 )

              Estimate StdErr          z P      2.5% 97.5%
(Intercept) -9.7755 0.0207 -472.3164 0 -9.8160 -9.7349
A              0.0055 0.0005  11.3926 0  0.0045  0.0064

> round( ci.exp( ml ), 4 )

              exp(Est.)  2.5% 97.5%
(Intercept)  0.0001 0.0001 0.0001
A              1.0055 1.0046 1.0064
```

Linear increase of log-rates by age

Multiple time scales and continuous rates (crv-mod)

Survival models and Cox-regression
Bendix Carstensen
Rates and Survival
Lifetable estimators
Kaplan-Meier estimators
The Cox-model
Who needs the Cox-model anyway?
Multiple time scales and continuous rates
69/94

Multiple time scales and continuous rates

Bendix Carstensen

Senior Statistician, Steno Diabetes Center

Survival models and Cox-regression
IDEG 2017 training day, Abu Dhabi,
11 December 2017

<http://BendixCarstensen/Epi/Courses/IDEG2017>

crv-mod

Linear effects in glm

```
> nd <- data.frame( A=15:60, Y=10^-5 )
> pr <- ci.pred( ml, newdata=nd )
> head( pr )

              Estimate      2.5% 97.5%
1  6.170105  5.991630  6.353896
2  6.204034  6.028525  6.384652
3  6.238149  6.065547  6.415662
4  6.272452  6.102689  6.446937
5  6.306943  6.139944  6.478485
6  6.341624  6.177301  6.510319

> matplot( nd$A, pr,
+          type="l", lty=1, lwd=c(3,1,1), col="black", log="y" )
```

Multiple time scales and continuous rates (crv-mod)

Survival models and Cox-regression
Bendix Carstensen
Rates and Survival
Lifetable estimators
Kaplan-Meier estimators
The Cox-model
Who needs the Cox-model anyway?
Multiple time scales and continuous rates
70/94

Testis cancer

Testis cancer in Denmark:

```
> options( show.signif.stars=FALSE )
> library( Epi )
> data( testisDK )
> str( testisDK )

'data.frame': 4860 obs. of 4 variables:
 $ A: num  0 1 2 3 4 5 6 7 8 9 ...
 $ P: num  1943 1943 1943 1943 1943 ...
 $ D: num  1 1 0 1 0 1 0 0 0 0 ...
 $ Y: num  39650 36943 34588 33267 32614 ...

> head( testisDK )

  A  P  D      Y
1 0 1943 1 39649.50
2 1 1943 1 36942.83
3 2 1943 0 34588.33
4 3 1943 1 33267.00
5 4 1943 0 32614.00
6 5 1943 0 32020.33
```

Multiple time scales and continuous rates (crv-mod)

Survival models and Cox-regression
Bendix Carstensen
Rates and Survival
Lifetable estimators
Kaplan-Meier estimators
The Cox-model
Who needs the Cox-model anyway?
Multiple time scales and continuous rates
67/94

Linear effects in glm

```
> round( ci.lin( ml ), 4 )

              Estimate StdErr          z P      2.5% 97.5%
(Intercept) -9.7755 0.0207 -472.3164 0 -9.8160 -9.7349
A              0.0055 0.0005  11.3926 0  0.0045  0.0064

> CI <- cbind( 1, nd$A )
> head( CI )

      [,1] [,2]
[1,] 1    15
[2,] 1    16
[3,] 1    17
[4,] 1    18
[5,] 1    19
[6,] 1    20

> matplot( nd$A, ci.exp( ml, ctr.mat=CI ),
+          type="l", lty=1, lwd=c(3,1,1), col="black", log="y" )
```

Multiple time scales and continuous rates (crv-mod)

Survival models and Cox-regression
Bendix Carstensen
Rates and Survival
Lifetable estimators
Kaplan-Meier estimators
The Cox-model
Who needs the Cox-model anyway?
Multiple time scales and continuous rates
71/94

Cases, PY and rates

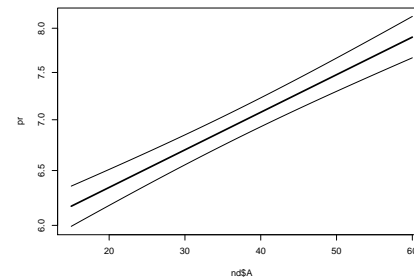
```
> stat.table( list(A=floor(A/10)*10,
+                 P=floor(P/10)*10,
+                 list( D=sum(D),
+                       Y=sum(Y/1000),
+                       rate=rate(D,Y,10^5) ),
+            margins=TRUE, data=testisDK )
```

A	P						
	1940	1950	1960	1970	1980	1990	Total
0	10.00	7.00	16.00	18.00	9.00	10.00	70.00
	2604.66	4037.31	3884.97	3820.88	3070.87	2165.54	19584.22
	0.38	0.17	0.41	0.47	0.29	0.46	0.36
10	13.00	27.00	37.00	72.00	97.00	75.00	321.00
	2135.73	3505.19	4004.13	3906.08	3847.40	2260.97	19659.48
	0.61	0.77	0.92	1.84	2.52	3.32	1.63
20	124.00	221.00	280.00	535.00	724.00	557.00	2441.00
	2225.55	2923.42	3401.65	4028.57	3941.18	2824.58	19344.74
	5.77	7.56	8.92	12.98	18.27	10.70	10.69

Multiple time scales and continuous rates (crv-mod)

Survival models and Cox-regression
Bendix Carstensen
Rates and Survival
Lifetable estimators
Kaplan-Meier estimators
The Cox-model
Who needs the Cox-model anyway?
Multiple time scales and continuous rates
68/94

Linear effects in glm

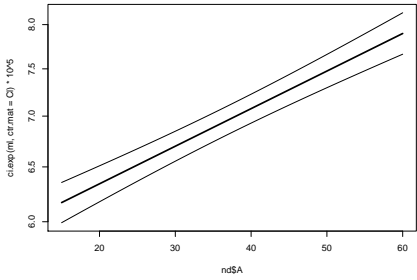


```
> matplot( nd$A, pr,
+          type="l", lty=1, lwd=c(3,1,1), col="black", log="y" )
```

Multiple time scales and continuous rates (crv-mod)

Survival models and Cox-regression
Bendix Carstensen
Rates and Survival
Lifetable estimators
Kaplan-Meier estimators
The Cox-model
Who needs the Cox-model anyway?
Multiple time scales and continuous rates
72/94

Linear effects in glm



```
> matplot( nd$A, ci.exp( ml, ctr.mat=C1 ) * 10^5,
+         type="l", lty=1, lwd=c(3,1,1), col="black", log="y" )
```

Multiple time scales and continuous rates (crrv-mod)

Survival models and Cox-regression

Bendix Carstensen

Rates and Survival

Lifetable estimators

Kaplan-Meier estimators

The Cox-model

Who needs the Cox-model anyway?

Multiple time scales and continuous rates

73 / 94

Spline effects in glm

```
> library( splines )
> ms <- glm( D ~ Ns(A,knots=seq(15,65,10)),
+         offset=log(Y), family=poisson, data=testisDK )
> round( ci.exp( ms ), 3 )
```

	exp(Est.)	2.5%	97.5%
(Intercept)	0.000	0.000	0.000
Ns(A, knots = seq(15, 65, 10))1	8.548	7.650	9.551
Ns(A, knots = seq(15, 65, 10))2	5.706	4.998	6.514
Ns(A, knots = seq(15, 65, 10))3	1.002	0.890	1.128
Ns(A, knots = seq(15, 65, 10))4	14.402	11.896	17.436
Ns(A, knots = seq(15, 65, 10))5	0.466	0.429	0.505

```
> aa <- 15:65
> As <- Ns( aa, knots=seq(15,65,10) )
> head( As )
```

	1	2	3	4	5
[1,]	0.0000000000	0.00000000	0.00000000	0.00000000	0.00000000
[2,]	0.0001666667	0.00000000	-0.02527011	0.07581034	-0.05054022
[3,]	0.0013333333	0.00000000	-0.05003313	0.15009940	-0.10006626
[4,]	0.0045000000	0.00000000	-0.07378197	0.22134590	-0.14756393

Multiple time scales and continuous rates (crrv-mod)

Survival models and Cox-regression

Bendix Carstensen

Rates and Survival

Lifetable estimators

Kaplan-Meier estimators

The Cox-model

Who needs the Cox-model anyway?

Multiple time scales and continuous rates

77 / 94

Quadratic effects in glm

How do rates depend on age?

```
> mq <- glm( D ~ A + I(A^2),
+         offset=log(Y), family=poisson, data=testisDK )
> round( ci.lin( mq ), 4 )
```

	Estimate	StdErr	z	P	2.5%	97.5%
(Intercept)	-12.3656	0.0596	-207.3611	0	-12.4825	-12.2487
A	0.1806	0.0033	54.8290	0	0.1741	0.1871
I(A^2)	-0.0023	0.0000	-53.7006	0	-0.0024	-0.0022

```
> round( ci.exp( mq ), 4 )
```

	exp(Est.)	2.5%	97.5%
(Intercept)	0.0000	0.0000	0.0000
A	1.1979	1.1902	1.2057
I(A^2)	0.9977	0.9976	0.9978

Multiple time scales and continuous rates (crrv-mod)

Survival models and Cox-regression

Bendix Carstensen

Rates and Survival

Lifetable estimators

Kaplan-Meier estimators

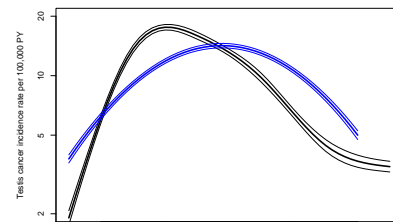
The Cox-model

Who needs the Cox-model anyway?

Multiple time scales and continuous rates

74 / 94

Spline effects in glm



```
> matplot( aa, ci.exp( ms, ctr.mat=cbind(1,As) ) * 10^5,
+         log="y", xlab="Age", ylab="Testis cancer incidence rate per 100,000 PY",
+         type="l", lty=1, lwd=c(3,1,1), col="black", ylim=c(2,20) )
> matlines( nd$A, ci.exp( mq, ctr.mat=Cq ) * 10^5,
+         type="l", lty=1, lwd=c(3,1,1), col="blue" )
```

Multiple time scales and continuous rates (crrv-mod)

Survival models and Cox-regression

Bendix Carstensen

Rates and Survival

Lifetable estimators

Kaplan-Meier estimators

The Cox-model

Who needs the Cox-model anyway?

Multiple time scales and continuous rates

78 / 94

Quadratic effect in glm

```
> round( ci.lin( mq ), 4 )
```

	Estimate	StdErr	z	P	2.5%	97.5%
(Intercept)	-12.3656	0.0596	-207.3611	0	-12.4825	-12.2487
A	0.1806	0.0033	54.8290	0	0.1741	0.1871
I(A^2)	-0.0023	0.0000	-53.7006	0	-0.0024	-0.0022

```
> Cq <- cbind( 1, 15:60, (15:60)^2 )
> head( Cq, 4 )
```

	[,1]	[,2]	[,3]
[1,]	1	15	225
[2,]	1	16	256
[3,]	1	17	289
[4,]	1	18	324

```
> matplot( nd$A, ci.exp( mq, ctr.mat=Cq ) * 10^5,
+         type="l", lty=1, lwd=c(3,1,1), col="black", log="y" )
```

Multiple time scales and continuous rates (crrv-mod)

Survival models and Cox-regression

Bendix Carstensen

Rates and Survival

Lifetable estimators

Kaplan-Meier estimators

The Cox-model

Who needs the Cox-model anyway?

Multiple time scales and continuous rates

75 / 94

Adding a linear period effect

```
> msp <- glm( D ~ Ns(A,knots=seq(15,65,10)) + P,
+         offset=log(Y), family=poisson, data=testisDK )
> round( ci.lin( msp ), 3 )
```

	Estimate	StdErr	z	P	2.5%	97.5%
(Intercept)	-58.105	1.444	-40.229	0.000	-60.935	-55.274
Ns(A, knots = seq(15, 65, 10))1	2.120	0.057	37.444	0.000	2.009	2.231
Ns(A, knots = seq(15, 65, 10))2	1.700	0.068	25.157	0.000	1.567	1.832
Ns(A, knots = seq(15, 65, 10))3	0.007	0.060	0.110	0.913	-0.112	0.125
Ns(A, knots = seq(15, 65, 10))4	2.596	0.097	26.631	0.000	2.405	2.787
Ns(A, knots = seq(15, 65, 10))5	-0.780	0.042	-18.748	0.000	-0.861	-0.698
P	0.024	0.001	32.761	0.000	0.023	0.025

```
> Ca <- cbind( 1, Ns( aa, knots=seq(15,65,10) ), 1970 )
> head( Ca )
```

	1	2	3	4	5	
[1,]	1	0.0000000000	0.00000000	0.00000000	0.00000000	1970
[2,]	1	0.0001666667	0.00000000	-0.02527011	0.07581034	-0.05054022
[3,]	1	0.0013333333	0.00000000	-0.05003313	0.15009940	-0.10006626
[4,]	1	0.0045000000	0.00000000	-0.07378197	0.22134590	-0.14756393
[5,]	1	0.0106666667	0.00000000	-0.09600952	0.28802857	-0.19201905

Multiple time scales and continuous rates (crrv-mod)

Survival models and Cox-regression

Bendix Carstensen

Rates and Survival

Lifetable estimators

Kaplan-Meier estimators

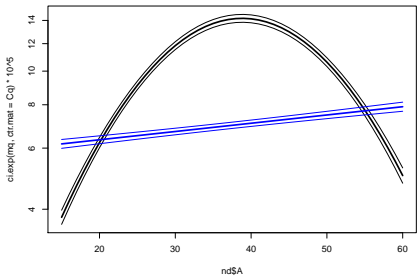
The Cox-model

Who needs the Cox-model anyway?

Multiple time scales and continuous rates

79 / 94

Quadratic effect in glm



```
> matplot( nd$A, ci.exp( mq, ctr.mat=Cq ) * 10^5,
+         type="l", lty=1, lwd=c(3,1,1), col="black", log="y" )
> matlines( nd$A, ci.exp( ml, ctr.mat=C1 ) * 10^5,
+         type="l", lty=1, lwd=c(3,1,1), col="blue" )
```

Multiple time scales and continuous rates (crrv-mod)

Survival models and Cox-regression

Bendix Carstensen

Rates and Survival

Lifetable estimators

Kaplan-Meier estimators

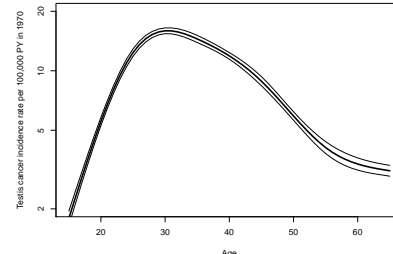
The Cox-model

Who needs the Cox-model anyway?

Multiple time scales and continuous rates

76 / 94

Adding a linear period effect



```
> matplot( aa, ci.exp( msp, ctr.mat=Ca ) * 10^5,
+         log="y", xlab="Age",
+         ylab="Testis cancer incidence rate per 100,000 PY in 1970",
+         type="l", lty=1, lwd=c(3,1,1), col="black", ylim=c(2,20) )
```

Multiple time scales and continuous rates (crrv-mod)

Survival models and Cox-regression

Bendix Carstensen

Rates and Survival

Lifetable estimators

Kaplan-Meier estimators

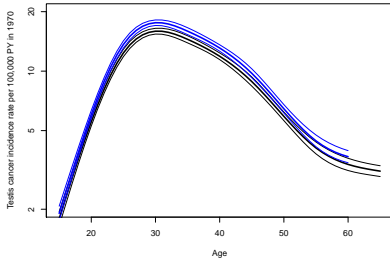
The Cox-model

Who needs the Cox-model anyway?

Multiple time scales and continuous rates

80 / 94

Adding a linear period effect



```
> matplot( aa, ci.exp( msp, ctr.mat=Ca ) * 10^5,
+         log="y", xlab="Age",
+         ylab="Testis cancer incidence rate per 100,000 PY in 1970",
+         type="l", lty=1, lwd=c(3,1,1), col="black", ylim=c(2,20) )
> matlines( nd$A, ci.pred( ms, newdata=nd ),
+         type="l", lty=1, lwd=c(3,1,1), col="blue" )
```

Multiple time scales and continuous rates (crrv-mod)

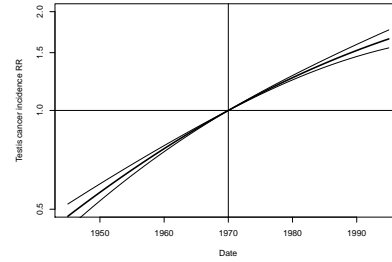
Survival models and Cox-regression
Bendix Carstensen

Rates and Survival
Lifetable estimators
Kaplan-Meier estimators
The Cox-model
Who needs the Cox-model anyway?

Multiple time scales and continuous rates

81 / 94

A quadratic period effect



```
> matplot( pp, ci.exp( mspq, subset="P", ctr.mat=Cq ),
+         log="y", ylim=c(0.5,2), xlab="Date",
+         ylab="Testis cancer incidence RR",
+         type="l", lty=1, lwd=c(3,1,1), col="black" )
> abline( h=1, v=1970 )
```

Multiple time scales and continuous rates (crrv-mod)

Survival models and Cox-regression
Bendix Carstensen

Rates and Survival
Lifetable estimators
Kaplan-Meier estimators
The Cox-model
Who needs the Cox-model anyway?

Multiple time scales and continuous rates

85 / 94

The period effect

```
> round( ci.lin( msp ), 3 )
```

	Estimate	StdErr	z	P	2.5%	97.5%
(Intercept)	-58.105	1.444	-40.229	0.000	-60.935	-55.274
Ns(A, knots = seq(15, 65, 10))1	2.120	0.057	37.444	0.000	2.009	2.231
Ns(A, knots = seq(15, 65, 10))2	1.700	0.068	25.157	0.000	1.567	1.832
Ns(A, knots = seq(15, 65, 10))3	0.007	0.060	0.110	0.913	-0.112	0.125
Ns(A, knots = seq(15, 65, 10))4	2.596	0.097	26.631	0.000	2.405	2.787
Ns(A, knots = seq(15, 65, 10))5	-0.780	0.042	-18.748	0.000	-0.861	-0.698
P	0.024	0.001	32.761	0.000	0.023	0.025

```
> pp <- seq(1945,1995,0.2)
> Cp <- cbind( pp ) - 1970
> head( Cp )
```

```
      pp
[1,] -25.0
[2,] -24.8
[3,] -24.6
[4,] -24.4
[5,] -24.2
[6,] -24.0
```

Multiple time scales and continuous rates (crrv-mod)

Survival models and Cox-regression
Bendix Carstensen

Rates and Survival
Lifetable estimators
Kaplan-Meier estimators
The Cox-model
Who needs the Cox-model anyway?

Multiple time scales and continuous rates

82 / 94

A spline period effect

Because we have the age-effect with the rate dimension, the period effect is a RR

```
> msp <- glm( D ~ Ns(A,knots=seq(15,65,10)) +
+           Ns(P,knots=seq(1950,1990,10),ref=1970),
+           offset=log(Y), family=poisson, data=testisDK )
> round( ci.exp( msp ), 3 )
```

	exp(Est.)	2.5%	97.5%
(Intercept)	0.000	0.000	0.000
Ns(A, knots = seq(15, 65, 10))1	8.327	7.452	9.305
Ns(A, knots = seq(15, 65, 10))2	5.528	4.842	6.312
Ns(A, knots = seq(15, 65, 10))3	1.007	0.894	1.133
Ns(A, knots = seq(15, 65, 10))4	13.447	11.107	16.279
Ns(A, knots = seq(15, 65, 10))5	0.458	0.422	0.497
Ns(P, knots = seq(1950, 1990, 10), ref = 1970)1	1.711	1.526	1.918
Ns(P, knots = seq(1950, 1990, 10), ref = 1970)2	2.190	2.028	2.364
Ns(P, knots = seq(1950, 1990, 10), ref = 1970)3	3.222	2.835	3.661
Ns(P, knots = seq(1950, 1990, 10), ref = 1970)4	2.299	2.149	2.459

Multiple time scales and continuous rates (crrv-mod)

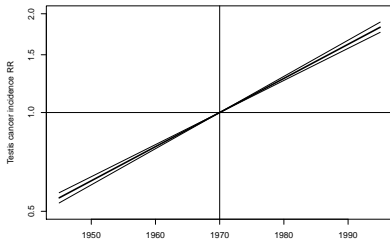
Survival models and Cox-regression
Bendix Carstensen

Rates and Survival
Lifetable estimators
Kaplan-Meier estimators
The Cox-model
Who needs the Cox-model anyway?

Multiple time scales and continuous rates

86 / 94

Period effect



```
> matplot( pp, ci.exp( msp, subset="P", ctr.mat=Cp ),
+         log="y", ylim=c(0.5,2), xlab="Date",
+         ylab="Testis cancer incidence RR",
+         type="l", lty=1, lwd=c(3,1,1), col="black" )
> abline( h=1, v=1970 )
```

Multiple time scales and continuous rates (crrv-mod)

Survival models and Cox-regression
Bendix Carstensen

Rates and Survival
Lifetable estimators
Kaplan-Meier estimators
The Cox-model
Who needs the Cox-model anyway?

Multiple time scales and continuous rates

83 / 94

A spline period effect

```
> Cp <- Ns( pp, knots=seq(1950,1990,10),ref=1970 )
> head( Cp, 4 )
```

```
      1          2          3          4
[1,] -0.6666667  0.0142689462 -0.5428068  0.3618712
[2,] -0.6666667  0.0091980207 -0.5275941  0.3517294
[3,] -0.6666667  0.0041270951 -0.5123813  0.3415875
[4,] -0.6666667 -0.0009438304 -0.4971685  0.3314457
```

```
> ci.exp( msp, subset="P" )
```

	exp(Est.)	2.5%	97.5%
Ns(P, knots = seq(1950, 1990, 10), ref = 1970)1	1.710808	1.525946	1.918065
Ns(P, knots = seq(1950, 1990, 10), ref = 1970)2	2.189650	2.027898	2.364303
Ns(P, knots = seq(1950, 1990, 10), ref = 1970)3	3.221563	2.835171	3.660614
Ns(P, knots = seq(1950, 1990, 10), ref = 1970)4	2.298946	2.149148	2.459186

```
> matplot( pp, ci.exp( msp, subset="P", ctr.mat=Cp ),
+         log="y", ylim=c(0.5,2), xlab="Date",
+         ylab="Testis cancer incidence RR",
+         type="l", lty=1, lwd=c(3,1,1), col="black" )
```

Multiple time scales and continuous rates (crrv-mod)

Survival models and Cox-regression
Bendix Carstensen

Rates and Survival
Lifetable estimators
Kaplan-Meier estimators
The Cox-model
Who needs the Cox-model anyway?

Multiple time scales and continuous rates

87 / 94

A quadratic period effect

```
> mspq <- glm( D ~ Ns(A,knots=seq(15,65,10)) + P + I(P^2),
+           offset=log(Y), family=poisson, data=testisDK )
> round( ci.exp( mspq ), 3 )
```

	exp(Est.)	2.5%	97.5%
(Intercept)	0.000	0.000	0.000
Ns(A, knots = seq(15, 65, 10))1	8.356	7.478	9.337
Ns(A, knots = seq(15, 65, 10))2	5.513	4.829	6.295
Ns(A, knots = seq(15, 65, 10))3	1.006	0.894	1.133
Ns(A, knots = seq(15, 65, 10))4	13.439	11.101	16.269
Ns(A, knots = seq(15, 65, 10))5	0.458	0.422	0.497
P	2.189	1.457	3.291
I(P^2)	1.000	1.000	1.000

```
> Cq <- cbind( pp-1970, pp^2-1970^2 )
> head( Cq )
```

```
      [,1]      [,2]
[1,] -25.0 -97875.00
[2,] -24.8 -97096.96
[3,] -24.6 -96318.84
[4,] -24.4 -95540.64
```

Multiple time scales and continuous rates (crrv-mod)

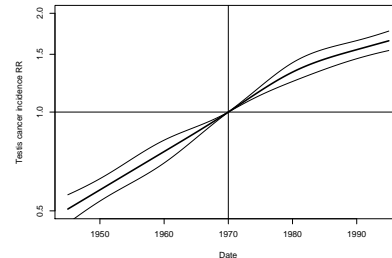
Survival models and Cox-regression
Bendix Carstensen

Rates and Survival
Lifetable estimators
Kaplan-Meier estimators
The Cox-model
Who needs the Cox-model anyway?

Multiple time scales and continuous rates

84 / 94

Period effect



```
> matplot( pp, ci.exp( msp, subset="P", ctr.mat=Cp ),
+         log="y", ylim=c(0.5,2), xlab="Date",
+         ylab="Testis cancer incidence RR",
+         type="l", lty=1, lwd=c(3,1,1), col="black" )
> abline( h=1, v=1970 )
```

Multiple time scales and continuous rates (crrv-mod)

Survival models and Cox-regression
Bendix Carstensen

Rates and Survival
Lifetable estimators
Kaplan-Meier estimators
The Cox-model
Who needs the Cox-model anyway?

Multiple time scales and continuous rates

88 / 94

Period effect

```
> par( mfrow=c(1,2) )
> matplot( aa, ci.pred( msp, newdata=data.frame(A=aa,P=1970,Y=10^5) ),
+         log="y", xlab="Age",
+         ylab="Testis cancer incidence rate per 100,000 PY in 1970",
+         type="l", lty=1, lwd=c(3,1,1), col="black" )
> matplot( pp, ci.exp( msp, subset="P", ctr.mat=Cp ),
+         log="y", xlab="Date", ylab="Testis cancer incidence RR",
+         type="l", lty=1, lwd=c(3,1,1), col="black" )
> abline( h=1, v=1970 )
```

Multiple time scales and continuous rates (crv-mod)

Survival models and Cox-regression

Bendix Carstensen

Rates and Survival

Lifetable estimators

Kaplan-Meier estimators

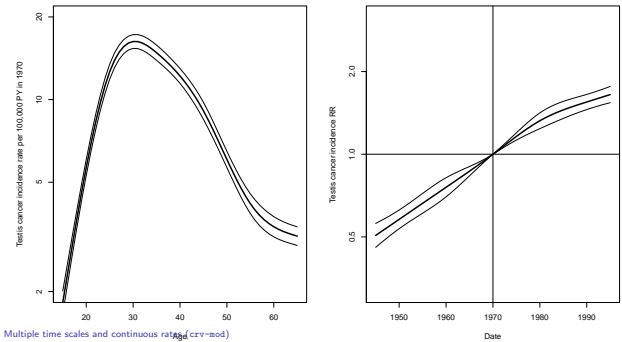
The Cox-model

Who needs the Cox-model anyway?

Multiple time scales and continuous rates

89 / 94

Age and period effect



Multiple time scales and continuous rates (crv-mod)

Survival models and Cox-regression

Bendix Carstensen

Rates and Survival

Lifetable estimators

Kaplan-Meier estimators

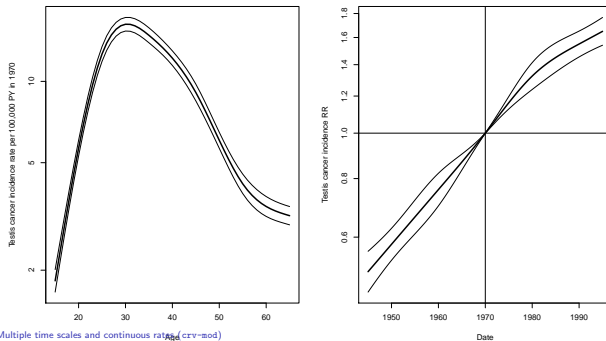
The Cox-model

Who needs the Cox-model anyway?

Multiple time scales and continuous rates

92 / 94

Age and period effect



Multiple time scales and continuous rates (crv-mod)

Survival models and Cox-regression

Bendix Carstensen

Rates and Survival

Lifetable estimators

Kaplan-Meier estimators

The Cox-model

Who needs the Cox-model anyway?

Multiple time scales and continuous rates

90 / 94

Age and period effect with ci.exp

- ▶ In rate models there is always one term with the **rate** dimension — usually **age**
- ▶ But it must refer to a specific **reference** value for **all other** variables (P).
- ▶ **All** parameters must be used in computing rates, at some reference value(s).
- ▶ For the “other” variables, report the RR **relative** to the reference point.
- ▶ Only parameters relevant for the variable (P) used.
- ▶ Contrast matrix is a **difference** between (splines at) the prediction points and the reference point.

Multiple time scales and continuous rates (crv-mod)

Survival models and Cox-regression

Bendix Carstensen

Rates and Survival

Lifetable estimators

Kaplan-Meier estimators

The Cox-model

Who needs the Cox-model anyway?

Multiple time scales and continuous rates

93 / 94

Period effect

```
> par( mfrow=c(1,2) )
> matplot( aa, ci.pred( msp, newdata=data.frame(A=aa,P=1970,Y=10^5) ),
+         log="y", xlab="Age",
+         ylim=c(2,20), xlim=c(15,65),
+         ylab="Testis cancer incidence rate per 100,000 PY in 1970",
+         type="l", lty=1, lwd=c(3,1,1), col="black" )
> matplot( pp, ci.exp( msp, subset="P", ctr.mat=Cp ),
+         log="y", xlab="Date",
+         ylim=c(2,20)/sqrt(2*20), xlim=c(15,65)+1930,
+         ylab="Testis cancer incidence RR",
+         type="l", lty=1, lwd=c(3,1,1), col="black" )
> abline( h=1, v=1970 )
```

Multiple time scales and continuous rates (crv-mod)

Survival models and Cox-regression

Bendix Carstensen

Rates and Survival

Lifetable estimators

Kaplan-Meier estimators

The Cox-model

Who needs the Cox-model anyway?

Multiple time scales and continuous rates

91 / 94