Multistate Models with Multiple Time Scales

Modern Demographic Methods in Epidemiology

Bendix Carstensen Steno Diabetes Center, Gentofte, Denmark
http://BendixCarstensen.com

27th IBC, Florence, 2014
6 July 2014
http://BendixCarstensen/AdvCoh/IBC2014

Plan of course

Mixture of lectures and demos — approximate times.
http://BendixCarstensen.com/AdvCoh/IBC2014

9:00–10:00 Introduction to survival and rates:

- Basic concepts
- Non-parametric and parametric models
- Practical estimation

10:00–11:00 Likelihood for and representation of multistate observations

- Data representation and overview
- Modles and reporting of rates

11:20–11:50 Simulation in multistate models.

12:15–13:30 A thorougly worked example:
Danish DM patients mortality

Rates and Survival

Sunday 5 July, morning

Bendix Carstensen

Multistate Models with Multiple Time Scales
Modern Demographic Methods in Epidemiology
6 July 2014
27th IBC, Florence, 2014
http://BendixCarstensen/AdvCoh/IBC2014

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”)

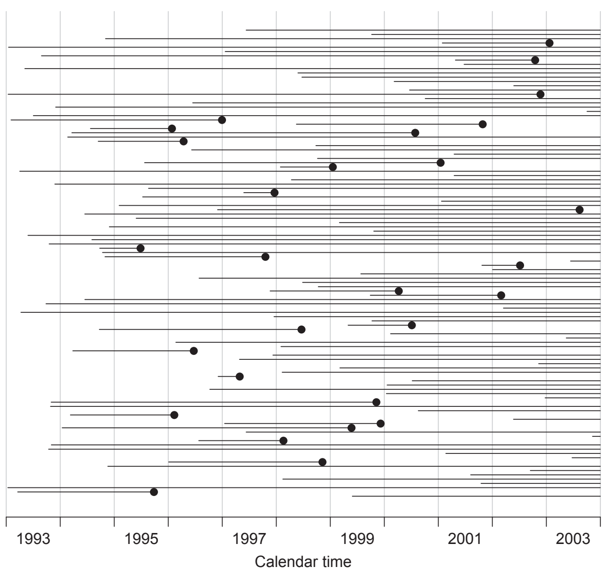
Examples of time-to-event measurements

- 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

Each line a person

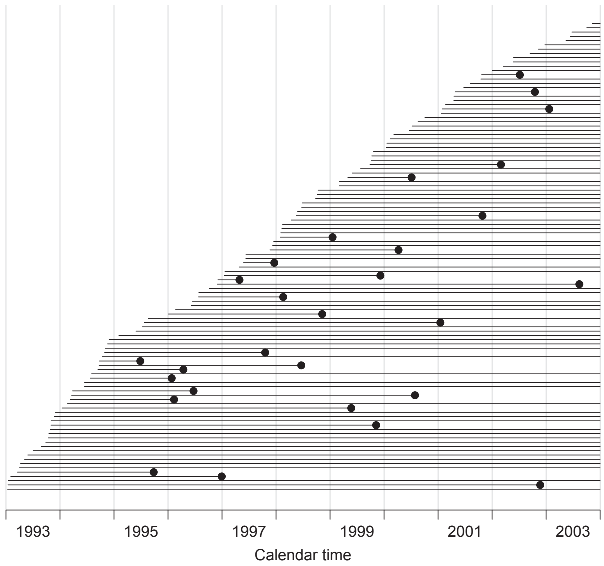
Each blob a death

Study ended at 31 Dec. 2003

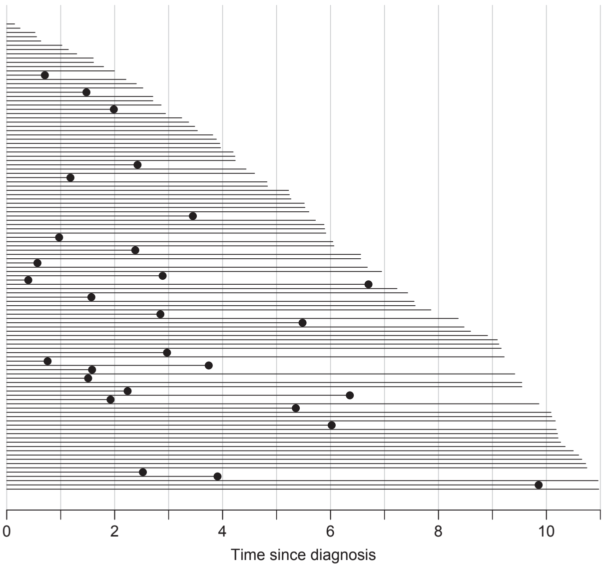


Ordered by date of entry

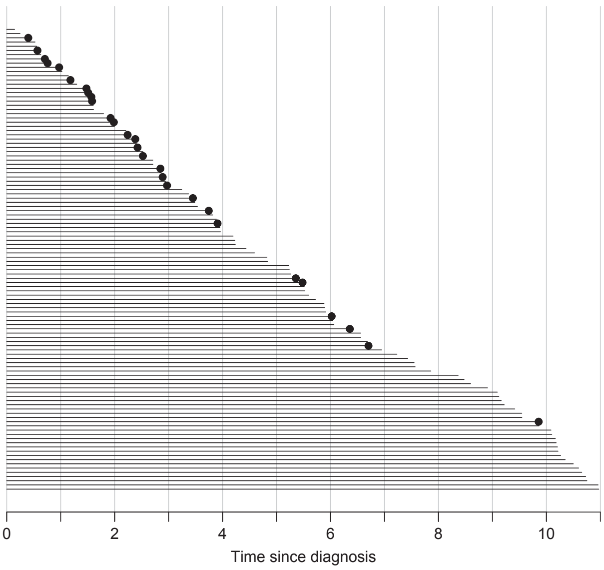
Most likely the order in your database.



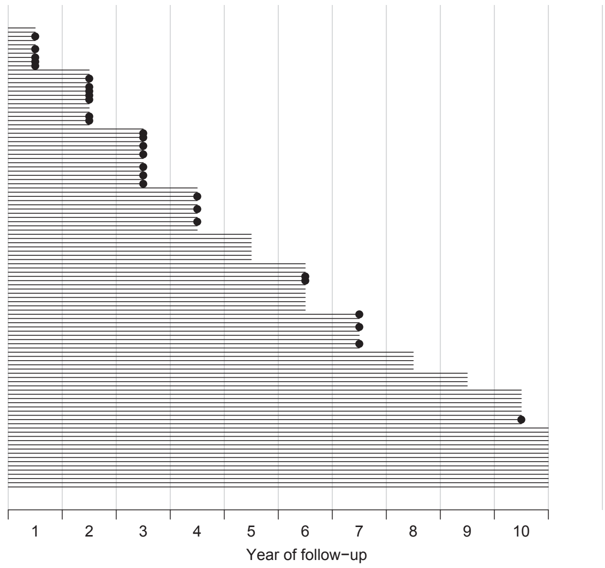
Timescale changed to "Time since diagnosis".



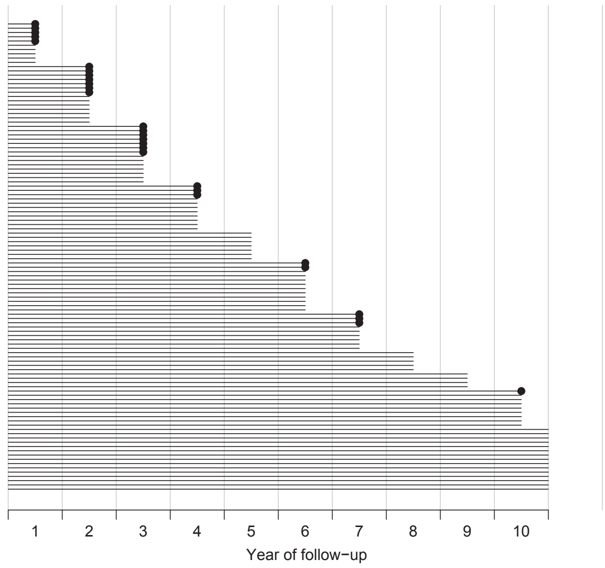
Patients ordered by survival time.



Survival times grouped into bands of survival.



Patients ordered by survival status within each band.



Survival after Cervix cancer

	Stage I			Stage II		
Year	n	d	l	n	d	l
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 $1 - 0.0465 = 0.9535$

Life-table estimator: $\hat{p}_i = d_i/(n_i - l_i/2)$

Life table estimators

- **Classical** lifetable estimator:
  - true probability of death in the ith interval is p_i
  - number of the l_i censored that are dead is $p_i l_i/2$
  - $p_i = (d_i + p_i l_i/2)/n_i \Leftrightarrow p_i = d_i/(n_i - l_i/2)$
- **Modified** liftetable estimator:
  - person years in interval of length ℓ_i: $\ell_i(n_i - d_i/2 - l_i/2)$
  - rate is $d_i/\ell_i(n_i - d_i/2 - l_i/2)$
  - culmulative rate is $\ell_i d_i/\ell_i(n_i - d_i/2 - l_i/2)$
  - $p_i = 1 - \exp\big(-d_i/(n_i - d_i/2 - l_i/2)\big)$
- Both cases: $S(t) = \prod_{i=0}^{i=t}(1 - p_i)$

Survival function

Persons enter at time 0:
Date of birth, date of randomization, date of diagnosis.

Survival time T — a stochastic variable.

Distribution is characterized by the survival function:

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

Note that the life-table estimator(s) **estimates** the distribution of the survival times. No restrictions on the relationship between p_is in different intervals.

Intensity or rate

$$\begin{aligned} &\mathrm{P}\{\text{event in } (t, t+h] \mid \text{alive at } t\}/h \\ &= \frac{F(t+h) - F(t)}{S(t) \times h} \\ &= -\frac{S(t+h) - S(t)}{S(t)h} \underset{h \to 0}{\longrightarrow} -\frac{\mathrm{d}\log S(t)}{\mathrm{d}t} \\ &= \lambda(t) \end{aligned}$$

This is the **intensity** or **hazard function** for the distribution. Characterizes the survival distribution as does f or F.

Theoretical counterpart of a **rate**.

Relationships

$$-\frac{\mathrm{d}\log S(t)}{\mathrm{d}t} = \lambda(t)$$

$$\Updownarrow$$

$$S(t) = \exp\left(-\int_0^t \lambda(u)\,\mathrm{d}u\right) = \exp\left(-\Lambda(t)\right)$$

$\Lambda(t) = \int_0^t \lambda(u)\,\mathrm{d}y$ is called

integrated intensity or **cumulative rate**
Not an intensity, it is dimensionless.

$$\lambda(t) = -\frac{\mathrm{d}\log(S(t))}{\mathrm{d}t} = -\frac{S'(t)}{S(t)} = \frac{F'(t)}{1-F(t)} = \frac{f(t)}{S(t)}$$

Rate and survival

$$S(t) = \exp\left(-\int_0^t \lambda(s)\,\mathrm{d}s\right) \qquad \lambda(t) = \frac{S'(t)}{S(t)}$$

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

Note:
A cumulative measure requires an **origin**!

Observed survival and rate

- Survival studies:
  Observe (right censored) survival time:

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

  — sometimes conditional on $T > t_0$
  (left truncated).
- Epidemiological studies:
  Observe (components of) a rate:

  $$D/Y$$

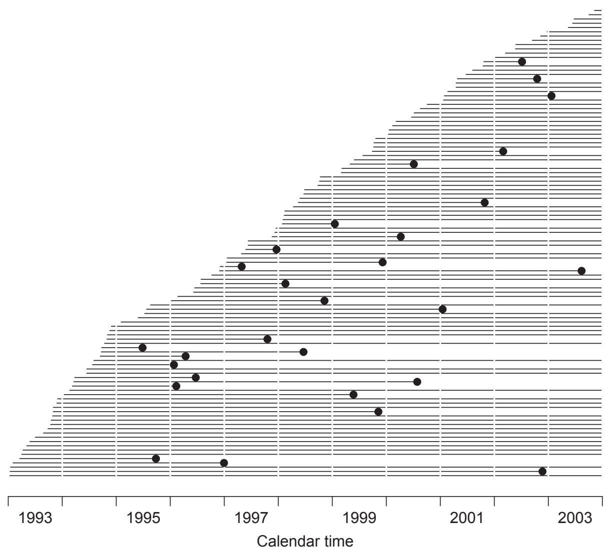
  D: no. events, Y no of person-years, in a prespecified time-frame.

Empirical rates for individuals

- At the **individual** level we introduce the **empirical rate:** (d, y),
  — no. events ($d \in \{0, 1\}$) during y risk time.
- A person may contribute several observations of (d_t, y_t)
- Indexed by t - timescale(s) and other covariates
- Empirical rates are **responses** in survival analysis — note it's **bivariate**.
- The timescale is a **covariate** — varies across empirical rates from one individual:
  Age, calendar time, time since diagnosis.
- Time at risk, follow-up time, y is part of the response.

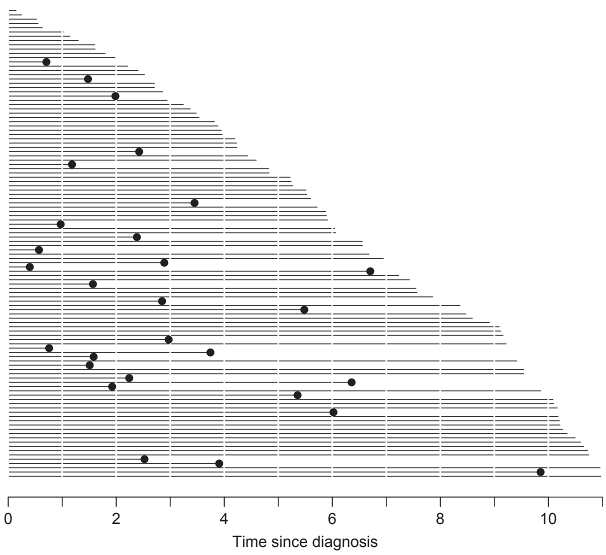
Empirical rates by calendar time.

. . . but each of these also has time since diagnosis and age included.



Empirical rates by time since diagnosis.

. . . but each of these also has calendar time and age included.



Likelihood from one person

...across several intervals (empirical rates) is a product of conditional probabilities:

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

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

Each term refers to one empirical rate (d_i, y_i) — $y_i = t_i - t_{i-1}$ and mostly $d_i = 0$.

t_i is the timescale (covariate).

Likelihood for an empirical rate

- Rate constant in (small) interval.
- $\pi = 1 - \mathrm{e}^{-\lambda y}$ is the death probability
- then:

$$\begin{aligned}
L(\lambda) = \mathrm{P}\left\{d \text{ events during } y \text{ time }\right\} &= \pi^d (1-\pi)^{1-d} \\
&= (1 - \mathrm{e}^{-\lambda y})^d (\mathrm{e}^{-\lambda y})^{1-d} \\
&= \left(\frac{1 - \mathrm{e}^{-\lambda y}}{\mathrm{e}^{-\lambda y}}\right)^d (\mathrm{e}^{-\lambda y}) \approx (\lambda y)^d \mathrm{e}^{-\lambda y}
\end{aligned}$$

  since the first term is equal to $\mathrm{e}^{\lambda y} - 1 \approx \lambda y$.
- $\ell(\lambda) \propto d\log(\lambda) - \lambda y$

"Poisson" likelihood

- Log-likelihood contributions from **one** individual:

$$\sum_t \left(d_t \log(\lambda_t) - \lambda_t y_t\right)$$

- the same as the log-likelihood from several **independent** Poisson observations, d_t, with mean $\lambda_t y_t$, i.e. log-mean:

$$\log\big(\mathrm{E}(d_t)\big) = \log\big(\lambda_t\big) + \log(y_t)$$

“Poisson” likelihood

- Muliplicative model for rates, $\log(\lambda_t) = X_t\beta$:
- Poisson observations, d_t, with mean $\lambda_t y_t$, i.e.:

$$\log\big(\mathrm{E}(d_t)\big) = \log\big(\lambda_t\big) + \log(y_t)$$
$$= X_t\beta + \log(y_t)$$

- Analysis of the rates, (λ_t) can be based on a Poisson model with log-link applied to empirical rates where:
  - d_t is the response variable.
  - $\log(y_t)$ is the offset variable.
  - X_t is the design matrix for describing rates in interval t

Likelihood for follow-up of many subjects

Adding empirical rates over the follow-up of persons:

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

- Persons are assumed independent
- Contribution from the same person are *conditionally* independent, hence give separate contributions to the log-likelihood.

The log-likelihood is maximal for:

$$\frac{\mathrm{d}\ell(\lambda)}{\mathrm{d}\lambda} = \frac{D}{\lambda} - Y = 0 \quad \Leftrightarrow \quad \hat{\lambda} = \frac{D}{Y}$$

Information about λ:

$$\ell(\lambda|D, Y) = D\log(\lambda) - \lambda Y, \quad \ell'_\lambda = D/\lambda - Y,$$
$$\ell''_\lambda = -D/\lambda^2$$

so $\mathrm{I}(\hat{\lambda}) = D/\hat{\lambda}^2 = Y^2/D$, hence $\mathrm{var}(\hat{\lambda}) = D/Y^2$
Standard error of a rate: $\sqrt{D}/Y$.

The log-likelihood is maximal for:

$$\frac{\mathrm{d}\ell(\lambda)}{\mathrm{d}\lambda} = \frac{D}{\lambda} - Y = 0 \quad \Leftrightarrow \quad \hat{\lambda} = \frac{D}{Y}$$

Information about $\theta = \log(\lambda)$:

$$\ell(\theta|D, Y) = D\theta - \mathrm{e}^{\theta} Y, \quad \ell'_{\theta} = D - \mathrm{e}^{\theta} Y,$$
$$\ell''_{\theta} = -\mathrm{e}^{\theta} Y$$

so $\mathrm{I}(\hat{\theta}) = \mathrm{e}^{\hat{\theta}} Y = \hat{\lambda} Y = D$, hence $\mathrm{var}(\hat{\theta}) = 1/D$

Standard error of log-rate: $1/\sqrt{D}$.

Note that this only depends on the no. events, **not** on the follow-up time.

Modelling a constant rate with `glm`

```
> D <- 12
> Y <- 1276.3/1000
> m0 <- glm( D   ~ 1, offset=log(Y), family=poisson )
> m1 <- glm( D/Y ~ 1, weights=Y, family=poisson )
> m2 <- glm( D/Y ~ 1, weights=Y, family=poisson(link=identity) )
> library( Epi )
> round( rbind( ci.lin( m0, E=T )[,c(1,2,5:7)],
+               ci.lin( m1, E=T )[,c(1,2,5:7)],
+               ci.lin( m2 )[,c(1,2,NA,5:6)] ), 3 )

     Estimate StdErr exp(Est.)  2.5%  97.5%
[1,]    2.241  0.289     9.402 5.340 16.556
[2,]    2.241  0.289     9.402 5.340 16.556
[3,]    9.402  2.714        NA 4.082 14.722

> round( c( 1/sqrt(D), sqrt(D)/Y ) , 3 )

[1] 0.289 2.714
```


Traditional survival analysis

Response variable: Time to event, T

Censoring at time Z

Observation $\big(\min(T, Z), \delta = 1\{T < Z\}\big)$.

Gives time a special status, because it mixes up:

- the response variable (risk)time
- the covariate time(scale).

Originates from clinical trials where everyone enters at time 0.

The life table method

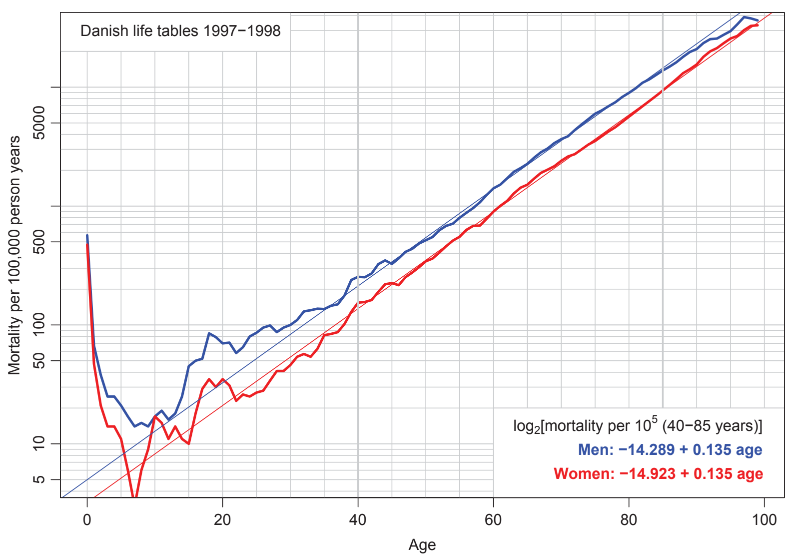
The simplest analysis is by the "life-table method":

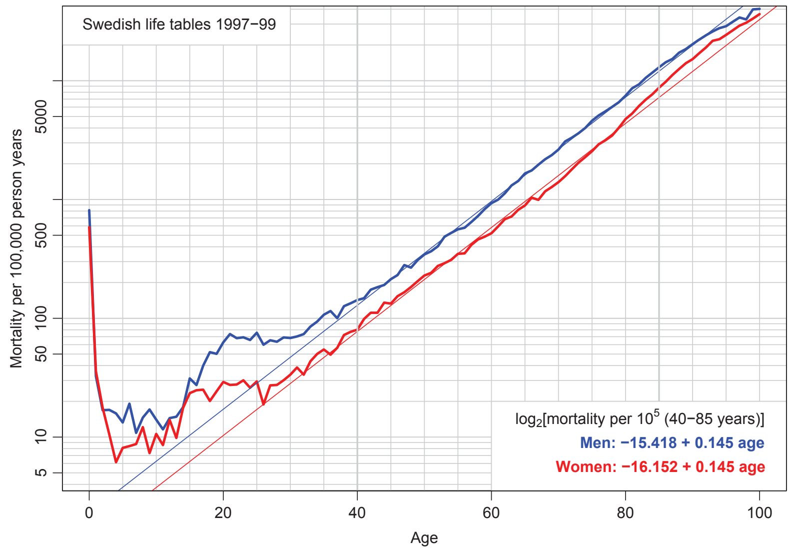
interval i	alive n_i	dead d_i	cens. l_i	p_i
1	77	5	2	$5/(77-2/2)=0.066$
2	70	7	4	$7/(70-4/2)=0.103$
3	59	8	1	$8/(59-1/2)=0.137$

$$p_i = \mathrm{P}\{\text{death in interval } i\} = 1 - d_i/(n_i - l_i/2)$$
$$S(t) = (1-p_1)\times\cdots\times(1-p_t)$$

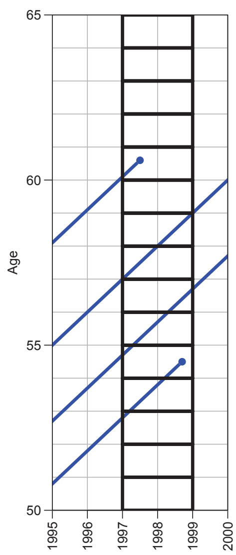
Population life table, DK 1997–98

	Men			Women		
a	$S(a)$	$\lambda(a)$	$\mathrm{E}[\ell_{\mathrm{res}}(a)]$	$S(a)$	$\lambda(a)$	$\mathrm{E}[\ell_{\mathrm{res}}(a)]$
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





Observations for the lifetable



Life table is based on person-years and deaths accumulated in a short period.

Age-specific rates — cross-sectional!

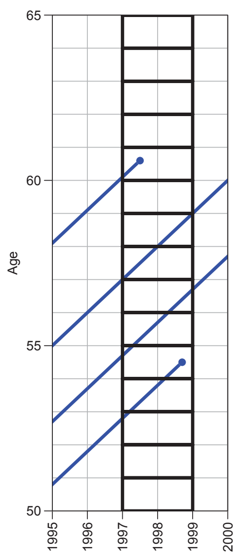
Survival function:

$$S(t) = \mathrm{e}^{-\int_0^t \lambda(a)\,\mathrm{d}a} = \mathrm{e}^{-\sum_0^t \lambda(a)}$$

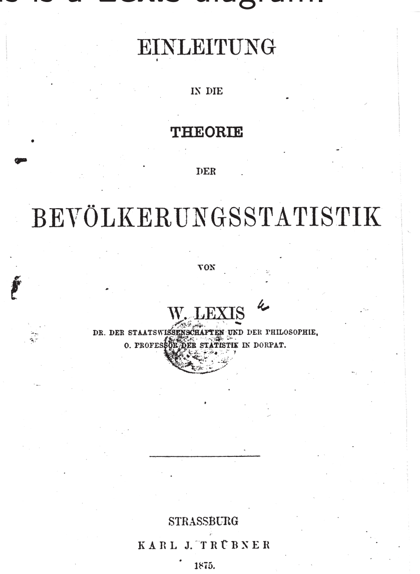
— assumes stability of rates to be interpretable for actual persons.

cross-sectional $\longleftrightarrow$ longitudinal

Observations for the lifetable



This is a **Lexis** diagram.


EINLEITUNG

IN DIE

THEORIE

DER

BEVÖLKERUNGSSTATISTIK

VON

W. LEXIS

DR. DER STAATSWISSENSCHAFTEN UND DER PHILOSOPHIE,
O. PROFESSOR DER STATISTIK IN DORPAT.

STRASSBURG
KARL J. TRÜBNER
1875.

Life table approach

- The observation of interest is **not** the survival time of the **individual**.
- It is 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 crossectionally, but interpreted longitudinally.

Summary

- Likelihood for a constant rate is proportional to a Poisson likelihood
- Subdividing follow-up in small intervals does not alter the likelihood
- Likelihood contribution from one person is a product of **conditionally** independent terms; one for each interval
- Assuming constant rate in very small intervals effectively allows rates to vary along different timescales
- Flexible shapes of the rates allowed

Who needs the Cox-model anyway?

Sunday 5 July, morning

Bendix Carstensen

Multistate Models with Multiple Time Scales
Modern Demographic Methods in Epidemiology
6 July 2014
27th IBC, Florence, 2014
http://BendixCarstensen/AdvCoh/IBC2014

The proportional hazards 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.
- Conceptually it is less clear — t is but a covariate that varies within individual.

Cox-likelihood

The partial 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:

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

The Cox-likelihood as profile likelihood

Suppose the time scale has been divided into small intervals with at most one death in each — empirical rates (d_t, y_t)

Assume w.l.o.g. that the ys 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 who were at risk at time t.

Note: There is one contribution from each person at risk to this part of the log-likelihood (and exactly one is dead):

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

where η_{death} is the linear predictor for the person that died at t.

The derivative w.r.t. α_t is:

$$\mathrm{D}_{\alpha_t} \ell(\alpha_t, \beta) = 1 - \mathrm{e}_t^{\alpha} \sum_{i \in \mathcal{R}_t} \mathrm{e}^{\eta_i} = 0 \quad \Leftrightarrow \quad \mathrm{e}_t^{\alpha} = \frac{1}{\sum_{i \in \mathcal{R}_t} \mathrm{e}^{\eta_i}}$$

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

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

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

What the Cox-model really is

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

- dividing time as finely as possible,
- modelling one covariate, the time-scale, with one parameter per distinct value,
- profiling these parameters out, and only maximizing the profile likelihood

Subsequently, one may recover the effect of the timescale by smoothing an estimate of the cumulative sum of these.

Sensible modelling

Replace the α_ts by a parmetric function $f(t)$ with a limited number of parameters, for example:

- Piecewise constant
- Splines (linear, quadratic or cubic)
- Fractional polynomials

Use Poisson modelling software on a dataset of empirical rates for small intervals (ys).

... but the data set is going to be quite large.

Splitting the dataset

The Poisson approach needs a dataset of empirical rates with small values of y.

Larger than the original: each individual contributes many empirical rates.

From each empirical rate we get:

- Poisson-response d
- Risk time y
- Covariate value for the timescale (time since entry, current age, current date, ...)
- other covariates

Example: Mayo Clinic lung cancer

Code is in `lung-ex.R`.

```
> options( width=120 )
> library( survival )
> library( Epi )
> data( lung )
> head( lung )
```

```
  inst time status age sex ph.ecog ph.karno pat.karno meal.cal
1    3  306      2  74   1       1       90       100     1175
2    3  455      2  68   1       0       90        90     1225
3    3 1010      1  56   1       0       90        90       NA
4    5  210      2  57   1       1       90        60     1150
5    1  883      2  60   1       0      100        90       NA
6   12 1022      1  74   1       1       50        80      513
```

Example: Mayo Clinic lung cancer

```
> Lx <- Lexis( exit=list( tfd=time+runif(nrow(lung),-0.5,0.5)),
+              exit.status=(status==2),
+              data=lung )

NOTE: entry is assumed to be 0 on the tfd timescale.

> summary( Lx, scale=365.25 )

Transitions:
     To
From    FALSE TRUE  Records:  Events: Risk time:  Persons:
  FALSE    63  165       228       165     190.53       228

> head( Lx )

  tfd   lex.dur lex.Cst lex.Xst lex.id inst time status age se
1   0  305.8516   FALSE    TRUE      1    3  306      2  74
2   0  455.1188   FALSE    TRUE      2    3  455      2  68
3   0 1010.3961   FALSE   FALSE      3    3 1010      1  56
4   0  209.7926   FALSE    TRUE      4    5  210      2  57
5   0  882.6279   FALSE    TRUE      5    1  883      2  60
6   0 1021.5707   FALSE   FALSE      6   12 1022      1  74
```


Example: Mayo Clinic lung cancer

```
> Sx <- splitLexis( Lx, "tfd", breaks=c(0,unique(exit(Lx))) )
> summary( Sx, scale=365.25 )

Transitions:
     To
From    FALSE TRUE  Records:  Events: Risk time:  Persons:
  FALSE 25941  165     26106       165     190.53       228

> subset( Sx, lex.id==96 )

      lex.id       tfd     lex.dur lex.Cst lex.Xst inst time s
11844     96  0.000000  4.95782724   FALSE   FALSE   12   30
11845     96  4.957827  5.72230893   FALSE   FALSE   12   30
11846     96 10.680136  0.49538575   FALSE   FALSE   12   30
11847     96 11.175522  0.09471063   FALSE   FALSE   12   30
11848     96 11.270233  0.99979856   FALSE   FALSE   12   30
11849     96 12.270031  0.64096619   FALSE   FALSE   12   30
11850     96 12.910997  0.12029712   FALSE   FALSE   12   30
11851     96 13.031294  1.84800876   FALSE   FALSE   12   30
11852     96 14.879303 11.54554087   FALSE   FALSE   12   30
11853     96 26.424844  3.20993281   FALSE    TRUE   12   30
```


Example: Mayo Clinic lung cancer

```
> c1 <- coxph( Surv(time,status==2) ~ sex + pat.karno, data=lung
> c2 <- coxph( Surv(tfd,tfd+lex.dur,lex.Xst==TRUE) ~ sex + pat.k
> p1 <- glm( lex.Xst ~ factor(tfd) + sex + pat.karno,
+            offset = log(lex.dur), family=poisson,
+            data=Sx )
> p2 <- glm( lex.Xst ~ ns(tfd,df=6) + sex + pat.karno,
+            offset = log(lex.dur), family=poisson,
+            data=Sx )
> p3 <- glm( lex.Xst ~ ns(tfd,df=2) + sex + pat.karno,
+            offset = log(lex.dur), family=poisson,
+            data=Sx )
```

Example: Mayo Clinic lung cancer

...better to allocate knots explicitly:

```
> k7 <- c( 0, quantile( rep(Sx$tfd,Sx$lex.Xst), (1:7-0.5)/7 ) )
> k3 <- c( 0, quantile( rep(Sx$tfd,Sx$lex.Xst), (1:3-0.5)/3 ) )
> xtabs( lex.Xst ~ cut(tfd,breaks=c(k7,Inf)), data=Sx )

   cut(tfd, breaks = c(k7, Inf))
     (0,46.5] (46.5,111]  (111,176]  (176,225]  (225,308]  (308,4
           11         24         23         24         23

> xtabs( lex.Xst ~ cut(tfd,breaks=c(k3,Inf)), data=Sx )

   cut(tfd, breaks = c(k3, Inf))
     (0,91.7] (91.7,225]  (225,468]  (468,Inf]
           27         55         54         28

> p2 <- glm( lex.Xst ~ Ns(tfd,knots=k7) + sex + pat.karno,
+            offset = log(lex.dur), family=poisson,
+            data=Sx )
> p3 <- glm( lex.Xst ~ Ns(tfd,knots=k3) + sex + pat.karno,
+            offset = log(lex.dur), family=poisson,
+            data=Sx )
```

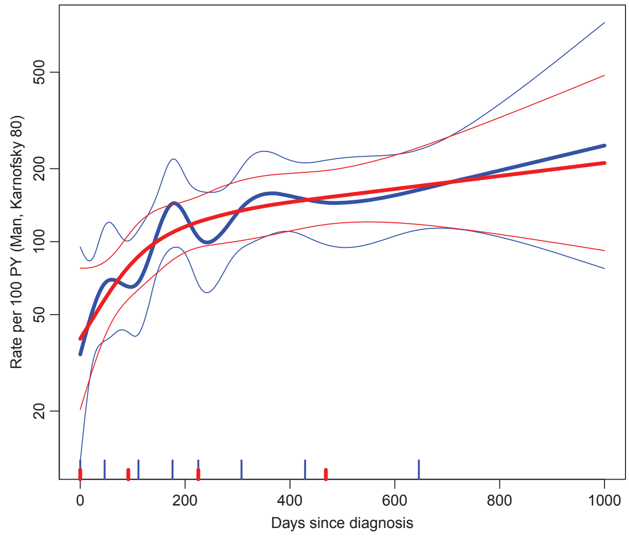

Example: Mayo Clinic lung cancer

```
> ee <- rbind( ci.exp( c1 ), ci.exp( c2 ),
+              ci.exp( p1, subset=c("sex","pat") ),
+              ci.exp( p2, subset=c("sex","pat") ),
+              ci.exp( p3, subset=c("sex","pat") ) )
> wh <- 1:5*2
> round( cbind( ee[wh-1,], ee[wh,] ), 4 )

       exp(Est.)   2.5%  97.5% exp(Est.)   2.5%  97.5%
   sex    0.5909 0.4244 0.8226    0.9801 0.9693 0.9909
   sex    0.5915 0.4249 0.8235    0.9800 0.9693 0.9909
   sex    0.5915 0.4249 0.8235    0.9800 0.9693 0.9909
   sex    0.5926 0.4256 0.8252    0.9798 0.9691 0.9907
   sex    0.5914 0.4248 0.8233    0.9797 0.9691 0.9906
```


Example: Mayo Clinic lung cancer

```
> range( Sx$tfd )

   [1]    0.000 1010.396

> nd <- data.frame( tfd=0:1000, lex.dur=36525,
+                   pat.karno=80, sex=1 )
> pr2 <- predict( p2, newdata=nd, se.fit=TRUE, type="link" )
> pr3 <- predict( p3, newdata=nd, se.fit=TRUE, type="link" )
> pr2 <- exp( cbind(pr2$fit,pr2$se.fit) %*% ci.mat() )
> pr3 <- exp( cbind(pr3$fit,pr3$se.fit) %*% ci.mat() )
> matplot( nd$tfd, cbind( pr2, pr3 ),
+          type="l", lty=1, lwd=c(4,1,1), col=rep(c("blue","red"
+          log="y", xlab="Days since diagnosis",
+          ylab="Rate per 100 PY (Man, Karnofsky 80)" )
> rug( k7, lwd=2, col="blue", ticksize=0.04 )
> rug( k3, lwd=4, col="red" , ticksize=0.02 )
```

Example: Mayo Clinic lung cancer



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:

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

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

δ-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 β is the total parameter vector in the model.
3. Variance-covariance matrix of $\hat{\beta}$: $\hat{\Sigma}$.
4. Variance-covariance of $\hat{f}(t_i)$: $\mathbf{B}\Sigma\mathbf{B}'$.
5. Transformation to the rates is the coordinate-wise exponential function, with derivative $\mathrm{diag}\bigl[\exp\bigl(\hat{f}(t_i)\bigr)\bigr]$

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

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

7. Transformation to cumulative hazard (ℓ is interval length):

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

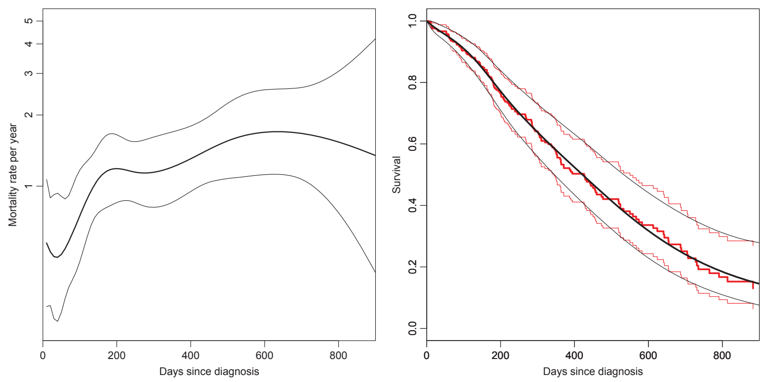
8. Variance-covariance matrix for the cumulative hazard is:

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

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

Mayo clinic lung cancer data

Smoothing by natural splines with 7 parameters; knots at $0, 25, 75, 150, 250, 500, 1000$ days



Summary

- All methods rely on some subdivision of the timescale(s):
  - Cox-modelling at the datapoints, implicitly in the algorithm
  - Poisson on an explicit pre-analysis division of data
- Based on the same form of the likelihood
- Poisson modelling gives easier access to the baseline hazard(s)
- Cox modelling is **much** faster, but misses the baseline hazard.

Representation of follow-up

Sunday 5 July, morning

Bendix Carstensen

Multistate Models with Multiple Time Scales
Modern Demographic Methods in Epidemiology
6 July 2014
27th IBC, Florence, 2014
`http://BendixCarstensen/AdvCoh/IBC2014`

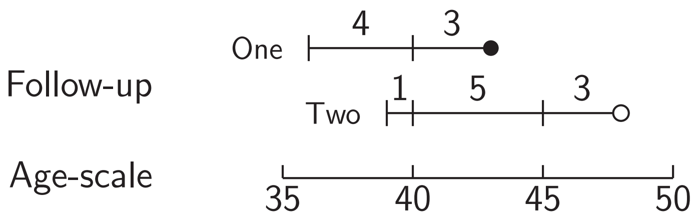
Follow-up and rates

- Follow-up studies:
  - D — events, deaths
  - Y — person-years
  - $\lambda = D/Y$ rates
- Rates differ between persons.
- Rates differ within persons:
  - By age
  - By calendar time
  - By disease duration
  - ...
- Multiple timescales.
- Multiple states (little boxes — later)

Stratification by age

If follow-up is rather short, age at entry is OK for age-stratification.

If follow-up is long, use stratification by categories of **current age**, both for:
No. of events, D, and Risk time, Y.



Representation of follow-up data

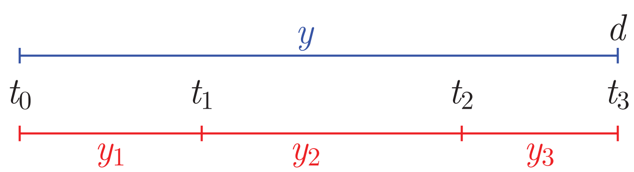
A cohort or follow-up study records:
Events and **Risk time**.

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

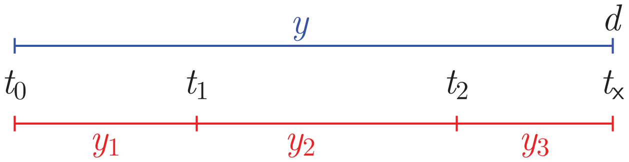
Follow-up **data** for each individual must therefore have (at least) three variables:

Date of entry	`entry`	date variable
Date of exit	`exit`	date variable
Status at exit	`fail`	indicator (0/1)

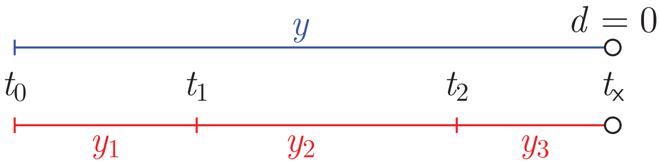
Specific for each **type** of outcome.



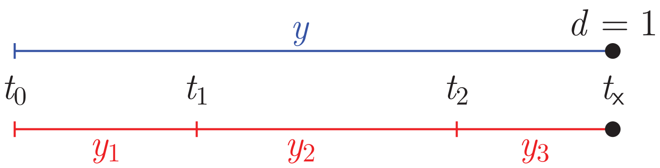
Probability	log-Likelihood
$\mathrm{P}(\text{event } t_3 \mid \text{entry } t_0)$	$d\log(\lambda) - \lambda y$
$= \mathrm{P}(\text{surv } t_0 \rightarrow t_1 \mid \text{entry } t_0)$	$= 0\log(\lambda) - \lambda y_1$
$\times\, \mathrm{P}(\text{surv } t_1 \rightarrow t_2 \mid \text{entry } t_1)$	$+\, 0\log(\lambda) - \lambda y_2$
$\times\, \mathrm{P}(\text{event } t_3 \mid \text{entry } t_2)$	$+\, d\log(\lambda) - \lambda y_3$



Probability	log-Likelihood
$\mathrm{P}(d \text{ at } t_x \mid \text{entry } t_0)$	$d\log(\lambda) - \lambda y$
$= \mathrm{P}(\text{surv } t_0 \rightarrow t_1 \mid \text{entry } t_0)$	$= 0\log(\lambda) - \lambda y_1$
$\times\, \mathrm{P}(\text{surv } t_1 \rightarrow t_2 \mid \text{entry } t_1)$	$+\, 0\log(\lambda) - \lambda y_2$
$\times\, \mathrm{P}(d \text{ at } t_x \mid \text{entry } t_2)$	$+\, d\log(\lambda) - \lambda y_3$



Probability	log-Likelihood
$\mathrm{P}(\text{surv } t_0 \rightarrow t_x \mid \text{entry } t_0)$	$0\log(\lambda) - \lambda y$
$= \mathrm{P}(\text{surv } t_0 \rightarrow t_1 \mid \text{entry } t_0)$	$= 0\log(\lambda) - \lambda y_1$
$\times\, \mathrm{P}(\text{surv } t_1 \rightarrow t_2 \mid \text{entry } t_1)$	$+\, 0\log(\lambda) - \lambda y_2$
$\times\, \mathrm{P}(\text{surv } t_2 \rightarrow t_x \mid \text{entry } t_2)$	$+\, 0\log(\lambda) - \lambda y_3$



Probability	log-Likelihood
$\mathrm{P}(\text{event at } t_x \mid \text{entry } t_0)$	$1\log(\lambda) - \lambda y$
$= \mathrm{P}(\text{surv } t_0 \rightarrow t_1 \mid \text{entry } t_0)$	$= 0\log(\lambda) - \lambda y_1$
$\times\, \mathrm{P}(\text{surv } t_1 \rightarrow t_2 \mid \text{entry } t_1)$	$+\, 0\log(\lambda) - \lambda y_2$
$\times\, \mathrm{P}(\text{event at } t_x \mid \text{entry } t_2)$	$+\, 1\log(\lambda) - \lambda y_3$

Dividing time into bands:

If we want to put D and Y into intervals on the timescale we must know:

Origin: The date where the time scale is 0:

- Age — 0 at date of birth
- Disease duration — 0 at date of diagnosis
- Occupation exposure — 0 at date of hire

Intervals: How should it be subdivided:

- 1-year classes? 5-year classes?
- Equal length?

Aim: Separate rate in each interval

Example: cohort with 3 persons:

```
Id      Bdate      Entry       Exit St
 1 14/07/1952 04/08/1965 27/06/1997  1
 2 01/04/1954 08/09/1972 23/05/1995  0
 3 10/06/1987 23/12/1991 24/07/1998  1
```

- Age bands: 10-years intervals of current age.
- Split Y for every subject accordingly
- Treat each segment as a separate unit of observation.
- Keep track of exit status in each interval.

Splitting the follow up

	subj. 1	subj. 2	subj. 3
Age at **E**ntry:	13.06	18.44	4.54
Age at e**X**it:	44.95	41.14	11.12
Status at exit:	Dead	Alive	Dead
Y	31.89	22.70	6.58
D	1	0	1

	subj. 1		subj. 2		subj. 3		$\sum$	
Age	Y	D	Y	D	Y	D	Y	D
0–	0.00	0	0.00	0	5.46	0	5.46	0
10–	6.94	0	1.56	0	1.12	1	8.62	1
20–	10.00	0	10.00	0	0.00	0	20.00	0
30–	10.00	0	10.00	0	0.00	0	20.00	0
40–	4.95	1	1.14	0	0.00	0	6.09	1
$\sum$	31.89	1	22.70	0	6.58	1	60.17	2

Splitting the follow-up

id	Bdate	Entry	Exit	St	risk	int
1	14/07/1952	03/08/1965	14/07/1972	0	6.9432	10
1	14/07/1952	14/07/1972	14/07/1982	0	10.0000	20
1	14/07/1952	14/07/1982	14/07/1992	0	10.0000	30
1	14/07/1952	14/07/1992	27/06/1997	1	4.9528	40
2	01/04/1954	08/09/1972	01/04/1974	0	1.5606	10
2	01/04/1954	01/04/1974	31/03/1984	0	10.0000	20
2	01/04/1954	31/03/1984	01/04/1994	0	10.0000	30
2	01/04/1954	01/04/1994	23/05/1995	0	1.1417	40
3	10/06/1987	23/12/1991	09/06/1997	0	5.4634	0
3	10/06/1987	09/06/1997	24/07/1998	1	1.1211	10

Keeping track of calendar time too?

Timescales

- A timescale is a variable that varies **deterministically** *within* each person during follow-up:
  - Age
  - Calendar time
  - Time since treatment
  - Time since relapse
- All timescales advance at the same pace (1 year per year . . .)
- Note: Cumulative exposure is **not** a timescale.

Follow-up on several timescales

- The risk-time is the same on all timescales
- Only need the entry point on each time scale:
  - Age at entry.
  - Date of entry.
  - Time since treatment at entry.
    — if time of treatment is the entry, this is 0 for all.
- Response variable in analysis of rates:

$$(d, y) \qquad (\text{event}, \text{duration})$$

- Covariates in analysis of rates:
  - timescales
  - other (fixed) measurements

Follow-up data in Epi — Lexis objects

A follow-up study:

```
> round( th, 2 )
    id sex birthdat contrast injecdat volume exitdat ex
1    1   2  1916.61        1  1938.79     22 1976.79
2  640   2  1896.23        1  1945.77     20 1964.37
3 3425   1  1886.97        2  1955.18      0 1956.59
4 4017   2  1936.81        2  1957.61      0 1992.14
...
```

Timescales of interest:

- Age
- Calendar time
- Time since injection

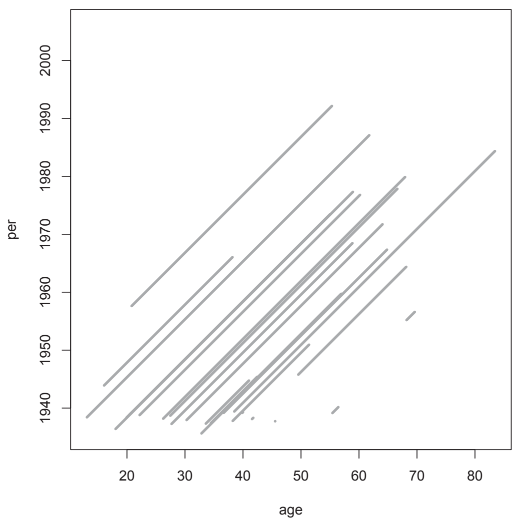
Definition of Lexis object

```
> thL <- Lexis( entry = list( age = injecdat-birthdat,
+                             per = injecdat,
+                             tfi = 0 ),
+               exit = list( per = exitdat ),
+        exit.status = as.numeric(exitstat==1),
+               data = th )
```

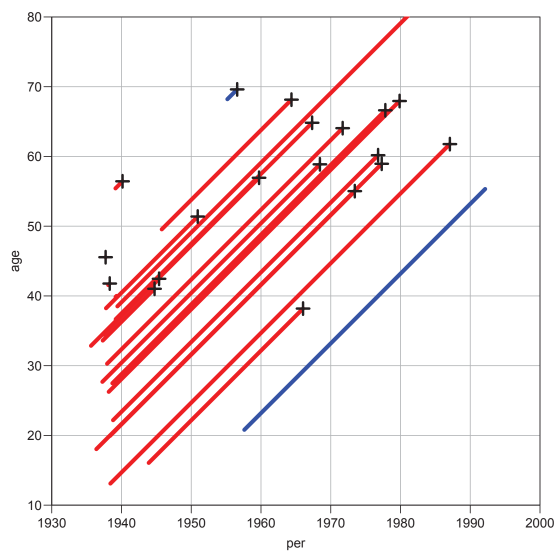
entry is defined on **three** timescales,
but exit is only defined on **one** timescale:
Follow-up time is the same on all timescales:
exitdat - injecdat

The looks of a `Lexis` object

```
> thL[,1:9]
     age     per tfi lex.dur lex.Cst lex.Xst lex.id
1 22.18 1938.79   0   37.99       0       1      1
2 49.54 1945.77   0   18.59       0       1      2
3 68.20 1955.18   0    1.40       0       1      3
4 20.80 1957.61   0   34.52       0       0      4
...

> summary( thL )
Transitions:
     To
From 0  1 Records:  Events:  Risk time:  Persons:
   0 3 20       23       20      512.59        23
```


```
> plot( thL, lwd=3 )
```

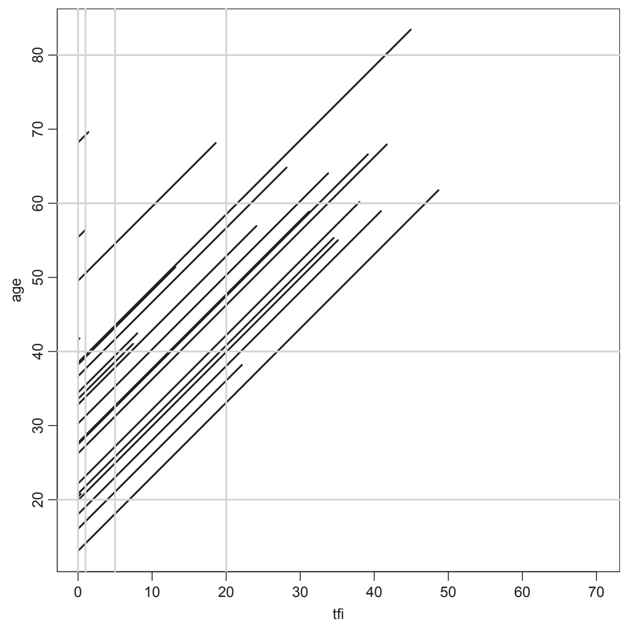

```
> plot( thL, 2:1, lwd=5, col=c("red","blue")[thL$contrast],
+        grid=TRUE, lty.grid=1, col.grid=gray(0.7),
+        xlim=1930+c(0,70), xaxs="i", ylim=  10+c(0,70), yaxs="i", las=1 )
> points( thL, 2:1, pch=c(NA,3)[thL$lex.Xst+1],lwd=3, cex=1.5 )
```

Splitting follow-up time

```
> spl1 <- splitLexis( thL, breaks=seq(0,100,20),
>                      time.scale="age" )
> round(spl1,1)
   age    per  tfi lex.dur lex.Cst lex.Xst   id sex birthdat con
1 22.2 1938.8  0.0    17.8       0       0    1   2   1916.6
2 40.0 1956.6 17.8    20.0       0       0    1   2   1916.6
3 60.0 1976.6 37.8     0.2       0       1    1   2   1916.6
4 49.5 1945.8  0.0    10.5       0       0  640   2   1896.2
5 60.0 1956.2 10.5     8.1       0       1  640   2   1896.2
6 68.2 1955.2  0.0     1.4       0       1 3425   1   1887.0
7 20.8 1957.6  0.0    19.2       0       0 4017   2   1936.8
8 40.0 1976.8 19.2    15.3       0       0 4017   2   1936.8
...
```


Split on another timescale

```
> spl2 <- splitLexis( spl1, time.scale="tfi",
                      breaks=c(0,1,5,20,100) )
> round( spl2, 1 )
   lex.id  age    per  tfi lex.dur lex.Cst lex.Xst   id sex birt
1       1 22.2 1938.8  0.0     1.0       0       0    1   2   19
2       1 23.2 1939.8  1.0     4.0       0       0    1   2   19
3       1 27.2 1943.8  5.0    12.8       0       0    1   2   19
4       1 40.0 1956.6 17.8     2.2       0       0    1   2   19
5       1 42.2 1958.8 20.0    17.8       0       0    1   2   19
6       1 60.0 1976.6 37.8     0.2       0       1    1   2   19
7       2 49.5 1945.8  0.0     1.0       0       0  640   2   18
8       2 50.5 1946.8  1.0     4.0       0       0  640   2   18
9       2 54.5 1950.8  5.0     5.5       0       0  640   2   18
10      2 60.0 1956.2 10.5     8.1       0       1  640   2   18
11      3 68.2 1955.2  0.0     1.0       0       0 3425   1   18
12      3 69.2 1956.2  1.0     0.4       0       1 3425   1   18
13      4 20.8 1957.6  0.0     1.0       0       0 4017   2   19
14      4 21.8 1958.6  1.0     4.0       0       0 4017   2   19
15      4 25.8 1962.6  5.0    14.2       0       0 4017   2   19
16      4 40.0 1976.8 19.2     0.8       0       0 4017   2   19
17      4 40.8 1977.6 20.0    14.5       0       0 4017   2   19
...
```


```
 age  tfi lex.dur
22.2  0.0     1.0
23.2  1.0     4.0
27.2  5.0    12.8
40.0 17.8     2.2
42.2 20.0    17.8
60.0 37.8     0.2
```

```
plot( spl2, c(1,3), col="black", lwd=2 )
```

Likelihood for a constant rate

- This setup is for a situation where it is assumed that rates are constant in each of the intervals.
- Each observation in the dataset contributes a term to a "Poisson" likelihood.
- Rates can vary along several timescales simultaneously.
- Models can include fixed covariates, as well as the timescales (the left end-points of the intervals) as continuous variables.

Analysis of results

- d_{pi} — events in the variable: `lex.Xst`:
  In the model as response: `lex.Xst==1`
- y_{pi} — risk time: `lex.dur` (duration):
  In the model as offset $\log(y)$, `log(lex.dur)`.
- Covariates are:
  - timescales (age, period, time in study)
  - other variables for this person (constant or *assumed* constant in each interval).
- Model rates using the covariates in `glm`:
  — no difference between time-scales and other covariates.

Likelihood for multistate follow-up

Sunday 5 July, morning

Bendix Carstensen

Multistate Models with Multiple Time Scales
Modern Demographic Methods in Epidemiology
6 July 2014
27th IBC, Florence, 2014
`http://BendixCarstensen/AdvCoh/IBC2014`

Likelihood for transition through states

$$\mathbf{A} \longrightarrow \mathbf{B} \longrightarrow \mathbf{C} \longrightarrow$$

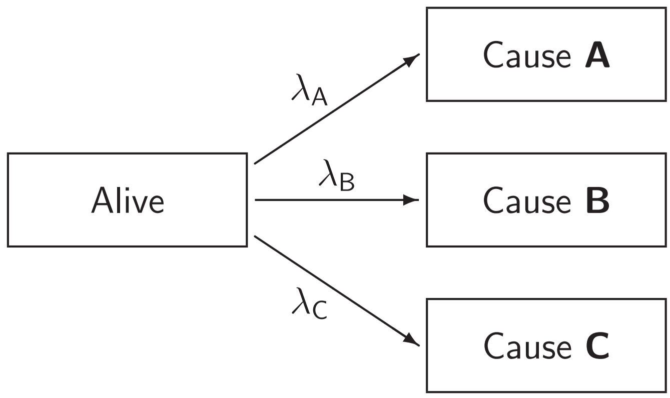
- given start of observation in **A** at time t_0
- transitions at times t_B and t_C
- survival in **C** till (at least) time t_x:

$$\begin{aligned} L = & \,\mathrm{P}\{\text{survive } t_0 \rightarrow t_B \text{ in } \mathbf{A}\} \\ & \times \mathrm{P}\{\text{transition } \mathbf{A} \rightarrow \mathbf{B} \text{ at } t_B | \text{ alive in } \mathbf{A}\} \\ & \times \mathrm{P}\{\text{survive } t_B \rightarrow t_C \text{ in } \mathbf{B} \mid \text{entered } \mathbf{B} \text{ at } t_B\} \\ & \times \mathrm{P}\{\text{transition } \mathbf{B} \rightarrow \mathbf{C} \text{ at } t_C | \text{ alive in } \mathbf{B}\} \\ & \times \mathrm{P}\{\text{survive } t_C \rightarrow t_x \text{ in } \mathbf{C} \mid \text{entered } \mathbf{C} \text{ at } t_C\} \end{aligned}$$

- Product of likelihoods for each transition
  — each one as for a survival model

Competing risks

But you may die from more than one cause
(or move to more than one state):



Cause-specific intensities

$$\begin{aligned} \lambda_A(t) &= \lim_{h\rightarrow 0} \frac{\mathrm{P}\{\text{death from cause A in } (t, t+h] \mid \text{alive at } t\}}{h} \\ \lambda_B(t) &= \lim_{h\rightarrow 0} \frac{\mathrm{P}\{\text{death from cause B in } (t, t+h] \mid \text{alive at } t\}}{h} \\ \lambda_C(t) &= \lim_{h\rightarrow 0} \frac{\mathrm{P}\{\text{death from cause C in } (t, t+h] \mid \text{alive at } t\}}{h} \end{aligned}$$

Total mortality rate:

$$\lambda_{\text{Total}}(t) = \lim_{h\rightarrow 0} \frac{\mathrm{P}\{\text{death from any cause in } (t, t+h] \mid \text{alive at } t\}}{h}$$

Cause-specific intensities

For small h, $\mathrm{P}\{2 \text{ events in } (t, t+h]\} \approx 0$, so:

$$\mathrm{P}\{\text{death from any cause in } (t, t+h] \mid \text{alive at } t\}$$

$$\begin{aligned} = \; & \mathrm{P}\{\text{death from cause A in } (t, t+h] \mid \text{alive at } t\} + \\ & \mathrm{P}\{\text{death from cause B in } (t, t+h] \mid \text{alive at } t\} + \\ & \mathrm{P}\{\text{death from cause C in } (t, t+h] \mid \text{alive at } t\} \end{aligned}$$

$$\Longrightarrow \qquad \lambda_{\text{Total}}(t) = \lambda_A(t) + \lambda_B(t) + \lambda_C(t)$$

Intensities are additive,
if they all refer to the
same risk set, in this case "Alive".

Likelihood for competing risks

Data:

Y	-	person years in "Alive"
D_A	-	deaths from cause A
D_B	-	deaths from cause B
D_C	-	deaths from cause C

Now, assume for a start that transition rates between states are constant.

Likelihood for competing risks

A survivor contributes to the log-likelihood:

$$\log(\mathrm{P}\{\text{Survival for a time of } y\}) = -(\lambda_A + \lambda_B + \lambda_C)y$$

A death from cause **A** contributes an additional $\log(\lambda_A)$, from cause **B** an additional $\log(\lambda_B)$ etc.

The total log-likelihood is then:

$$\begin{aligned} \ell(\lambda_A, \lambda_B, \lambda_C) = & D_A \log(\lambda_A) + D_B \log(\lambda_B) + D_C \log(\lambda_C) \\ & - (\lambda_A + \lambda_B + \lambda_C) Y \\ = & [D_A \log(\lambda_A) - \lambda_A Y] + \\ & [D_B \log(\lambda_B) - \lambda_B Y] + \\ & [D_C \log(\lambda_C) - \lambda_C Y] \end{aligned}$$

Components of the likelihood

The log-likelihood is made up of three contributions:

- one for cause A,
- one for cause B and
- one for cause C

Deaths are the cause-specific deaths,

but the **person-years** are the same in all contributions.

Likelihood for multiple states

- **Product** of likelihoods for each transition — each one as for a survival model
- **conditional** on being alive at (observed) entry to current state
- **Risk time** is the risk time in the current ("From") state
- **Events** are transitions to the "To" state
- All other transitions out of "From" are treated as **censorings** (but they are not)
- Fit models separately for each transition or jointly for all

Time varying rates:

- The same type of analysis as with a constant rates, but data must be
- split time in intervals sufficiently small to justify an assumption of constant rate (intensity)
- allow for a separate rate for each interval
- but constrained to follow model with a smooth effect of the time-scale values allocated to each interval.

Practical implications

- Empirical rates ((d, y) from each individual) will be the same for all analyses except for those where deaths occur.
- Analysis of cause **A**:
  - Contributions $(1, y)$ only for those intervals where a cause **A** death occurs.
  - Intervals with cause **B** or **C** deaths (or no deaths) contribute only $(0, y)$ treated as censorings.


```
             original                               expanded
------------------------------            ---------------------
id time cause     xx d.A d.B d.C          id time dd     xx    Tr
 1    1     B  0.50   0   1   0            1    1  0   0.50     A
 2    1    NA  1.00   0   0   0            2    1  0   1.00     A
 3    8     B -1.74   0   1   0            3    8  0  -1.74     A
 4    3     A -0.55   1   0   0            4    3  1  -0.55     A
 5    7    NA -0.58   0   0   0            5    7  0  -0.58     A
 6    7     C -0.04   0   0   1            6    7  0  -0.04     A

                                           1    1  1   0.50     B
                                           2    1  0   1.00     B
                                           3    8  1  -1.74     B
                                           4    3  0  -0.55     B
                                           5    7  0  -0.58     B
                                           6    7  0  -0.04     B

                                           1    1  0   0.50     C
                                           2    1  0   1.00     C
                                           3    8  0  -1.74     C
                                           4    3  0  -0.55     C
                                           5    7  0  -0.58     C
                                           6    7  1  -0.04     C
```

...accomplished by `stack.Lexis`

`Lexis` objects (data frame)

- Represents the **follow-up**
- `lex.dur` contains the total time at risk for (any) event
- `lex.Cst` is the state in which this time is spent
- `lex.Xst` is the state to which a transition occurs
  — if none, the same as `lex.Cst`.

This is used for modelling of single transitions between states — and multiple transitions with no two originating in the same state.

`stacked.Lexis` objects (data frame)

- Represents the **likelihood** contributions
- `lex.dur` contains the total time at risk for (any) event
- `lex.Tr` is the transition to which the record contributes
- `lex.Fail` is the event (failure) indicator for the transition in question.

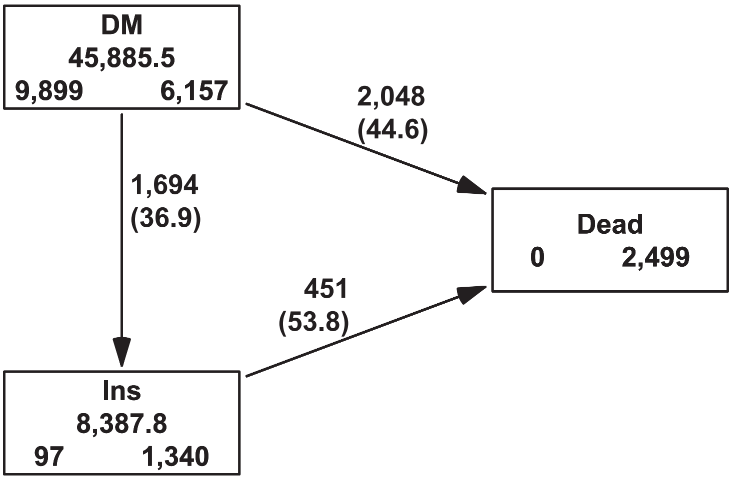
This is used for joint modelling of **all** transition in a multistate set-up. Particularly with several rates oriinating in the **same** state.

Implemented in the `stack.Lexis` function:

```
> library( Epi )
> data(DMlate)
> head(DMlate)

       sex    dobth     dodm    dodth    dooad doins      dox
50185    F 1940.256 1998.917       NA       NA    NA 2009.997
307563   M 1939.218 2003.309       NA 2007.446    NA 2009.997
294104   F 1918.301 2004.552       NA       NA    NA 2009.997
336439   F 1965.225 2009.261       NA       NA    NA 2009.997
245651   M 1932.877 2008.653       NA       NA    NA 2009.997
216824   F 1927.870 2007.886 2009.923       NA    NA 2009.923

> dml <- Lexis( entry = list(Per = dodm,
+                           Age = dodm-dobth,
+                         DMdur = 0 ),
+               exit = list(Per = dox ),
+        exit.status = factor(!is.na(dodth),
+                          labels=c("DM","Dead")),
+               data = DMlate )

NOTE: entry.status has been set to "DM" for all.
```


Implemented in the `stack.Lexis` function:

```
> dmi <- cutLexis( dml, cut = dml$doins,
+                new.state = "Ins",
+                precursor = "DM" )
> summary( dmi )

Transitions:
     To
From   DM  Ins Dead  Records:  Events: Risk time:  Persons:
  DM 6157 1694 2048      9899     3742   45885.49      9899
  Ins   0 1340  451      1791      451    8387.77      1791
  Sum 6157 3034 2499    11690     4193   54273.27      9996

> boxes( dmi, boxpos = list(x=c(20,20,80),
+                           y=c(80,20,50)),
+             scale.R=1000, show.BE=TRUE, hmult=1.2, wmult=1.1 )
```



Implemented in the stack.Lexis function:

```
> options( digits=3, width=200 )
> st.dmi <- stack( dmi )
> print( st.dmi[1:6,], row.names=F )

  Per  Age DMdur lex.dur lex.Cst lex.Xst  lex.Tr lex.Fail lex
 1999 58.7     0  11.080      DM      DM DM->Ins    FALSE
 2003 64.1     0   6.689      DM      DM DM->Ins    FALSE
 2005 86.3     0   5.446      DM      DM DM->Ins    FALSE
 2009 44.0     0   0.736      DM      DM DM->Ins    FALSE
 2009 75.8     0   1.344      DM      DM DM->Ins    FALSE
 2008 80.0     0   2.037      DM    Dead DM->Ins    FALSE

> str( st.dmi )

Classes 'stacked.Lexis' and 'data.frame': 21589 obs. of  16 va
 $ Per     : num  1999 2003 2005 2009 2009 ...
 $ Age     : num  58.7 64.1 86.3 44 75.8 ...
 $ DMdur   : num  0 0 0 0 0 0 0 0 0 0 ...
 $ lex.dur : num  11.08 6.689 5.446 0.736 1.344 ...
 $ lex.Cst : Factor w/ 3 levels "DM","Ins","Dead": 1 1 1 1 1 1
 $ lex.Xst : Factor w/ 3 levels "DM","Ins","Dead": 1 1 1 1 1 3
 $ lex.Tr  : Factor w/ 3 levels "DM->Ins","DM->Dead",..: 1 1 1
 $ lex.Fail: logi  FALSE FALSE FALSE FALSE FALSE FALSE ...
 $ lex.id  : int  1 2 3 4 5 6 7 8 9 10 ...
```


Implemented in the stack.Lexis function:

```
> print( subset(    dmi, lex.id %in% c(13,15,28) ), row.names=FA

  Per  Age DMdur lex.dur lex.Cst lex.Xst lex.id sex dobth dodm
 1997 59.4   0.0   0.890      DM    Dead     13   M  1938 1997
 2003 58.1   0.0   2.804      DM     Ins     15   M  1944 2003
 2005 60.9   2.8   4.643     Ins     Ins     15   M  1944 2003
 1999 73.7   0.0   8.701      DM     Ins     28   F  1925 1999
 2007 82.4   8.7   0.977     Ins    Dead     28   F  1925 1999

> print( subset( st.dmi, lex.id %in% c(13,15,28) ), row.names=FA

  Per  Age DMdur lex.dur lex.Cst lex.Xst    lex.Tr lex.Fail le
 1997 59.4   0.0   0.890      DM    Dead   DM->Ins    FALSE
 2003 58.1   0.0   2.804      DM     Ins   DM->Ins     TRUE
 1999 73.7   0.0   8.701      DM     Ins   DM->Ins     TRUE
 1997 59.4   0.0   0.890      DM    Dead  DM->Dead     TRUE
 2003 58.1   0.0   2.804      DM     Ins  DM->Dead    FALSE
 1999 73.7   0.0   8.701      DM     Ins  DM->Dead    FALSE
 2005 60.9   2.8   4.643     Ins     Ins Ins->Dead    FALSE
 2007 82.4   8.7   0.977     Ins    Dead Ins->Dead     TRUE
```

Analysis of rates in multistate models

- Interactions between all covariates (including time) and state (`lex.Cst`):
  ⇒ separate analyses of all transition rates.
- Only interaction between state (`lex.Cst`) and time(scales):
  ⇒ same covariate effects for all causes transitions, but separate baseline hazards — "stratified model".
- Main effect of state only (`lex.Cst`):
  ⇒ proportional hazards
- No effect of state:
  ⇒ identical baseline hazards — hardly ever relevant.

Analysis approaches and data representation

- `Lexis` objects represents the precise follow-up in the cohort, in states and along timescales
- — used for analysis of single transition rates.
- `stacked.Lexis` objects represents contributions to the total likelihood
- — used for joint analysis of (all) rates in a multistate setup
- ...which is the case if you want to specify common effects between different transitions.

Assumptions in competing risks

"Classical" way of looking at survival data: description of the distribution of time to death.

For competing risks that would require three variables:
T_A, T_B and T_C, representing times to death from each of the three causes.
But at most one of these is observed.

Often it is stated that these must be assumed independent in order to make the likelihood machinery work

1. It is not necessary.
2. Independence can never be assessed from data.

An account of these problems is given in:

PK Andersen, SZ Abildstrøm & S Rosthøj:
Competing risks as a multistate model,
Statistical Methods in Medical Research; **11**, 2002: pp. 203–215

Per Kragh Andersen, Ronald B Geskus, Theo de Witte & Hein Putter:
Competing risks in epidemiology: possibilities and pitfalls,
International Journal of Epidemiology; 2012: pp. 1–10

Contains examples where both dependent and independent "cause specific survival times" gives rise to the same set of cause specific rates.

Lifetime risk
Sunday 5 July, morning

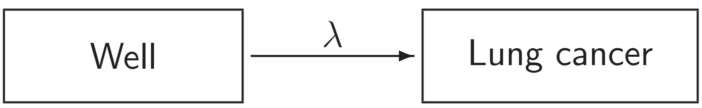
Bendix Carstensen

Multistate Models with Multiple Time Scales
Modern Demographic Methods in Epidemiology
6 July 2014
27th IBC, Florence, 2014
http://BendixCarstensen/AdvCoh/IBC2014

Competing risk interpretation

The problems with competing risk models **only** comes when estimated intensities (rates) are used to produce probability statements.

Classical set-up in cancer-registries:

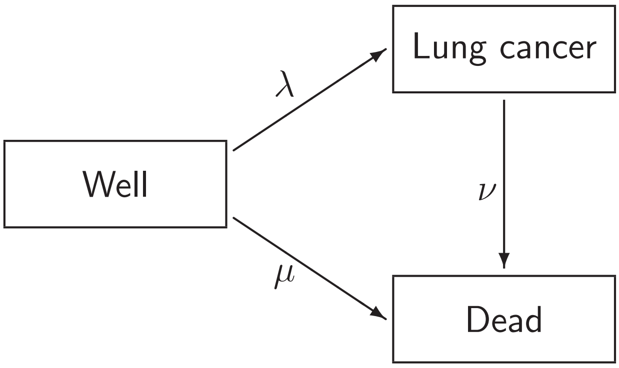


Common statement:

$$\mathrm{P}\left\{\text{Lung cancer before age 75}\right\} = 1 - \mathrm{e}^{-\Lambda(75)}$$

This is not quite right.

How the world really looks



Illness-death model, mortality of lung cancer patients (ν) not relevant here, we only want to find out how many pass through "Lung cancer"

How many get lung cancer before age a?

- $$\mathrm{P}\left\{\text{Lung cancer before age 75}\right\} \neq 1 - \mathrm{e}^{-\Lambda(75)}$$

  the r.h.s. does not take the possibility of death prior to lung cancer into account.
- $1 - \mathrm{e}^{-\Lambda(75)}$ often stated as the probability of lung cancer before age 75, assuming all other acuses of death absent.
- Lung cancer rates are however observed in a mortal population.
- If all other causes of death were absent, this would assume that lung cancer rates remained the same.

How it really is:

$\mathrm{P}\left\{\text{Lung cancer diagnosis before age } a\right\}$

$$= \int_0^a \mathrm{P}\left\{\text{Lung cancer at age } u\right\} \mathrm{d}u$$

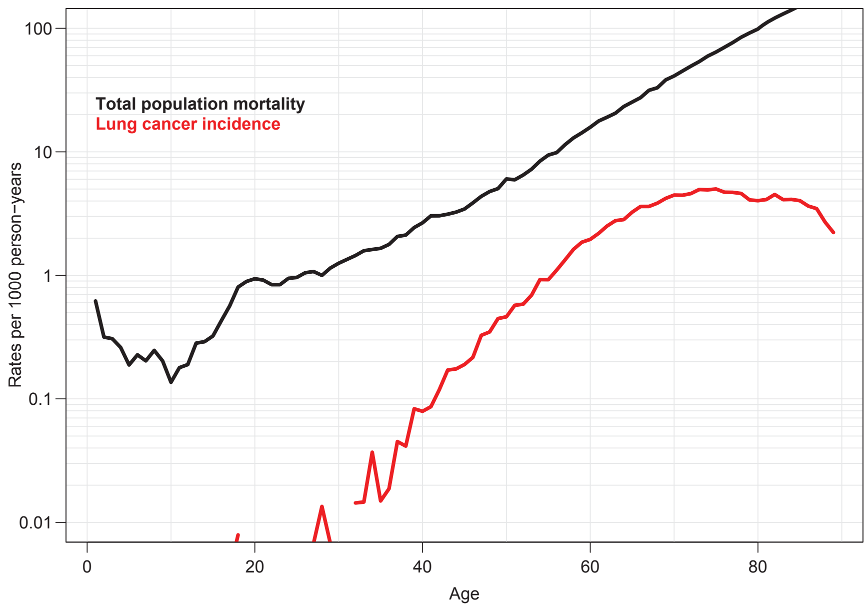
$$= \int_0^a \mathrm{P}\left\{\text{Lung cancer in age } (u, u + \mathrm{d}u] \mid \text{alive at } u\right\} \times \mathrm{P}\left\{\text{alive at } u \text{ without lung cancer}\right\} \mathrm{d}u$$

$$= \int_0^a \lambda(u) \exp\left(-\int_0^u \mu(s) + \lambda(s)\,\mathrm{d}s\right) \mathrm{d}u$$

Probability of lungcancer

The rates are easily plotted for inspection in R:

```
matplot( age, 1000*cbind( D/Y, lung/Y ),
         log="y", type="l", lty=1, lwd=3,
         ylim=c(0.01,100), xlab="Age",
         ylab="Rates per 1000 person-years" )
```

The probablility that a person contracts lung cancer before age a is:

$$\int_0^a \lambda(u) \exp\left(-\int_0^u \mu(s) + \lambda(s)\,\mathrm{d}s\right)\,\mathrm{d}u$$

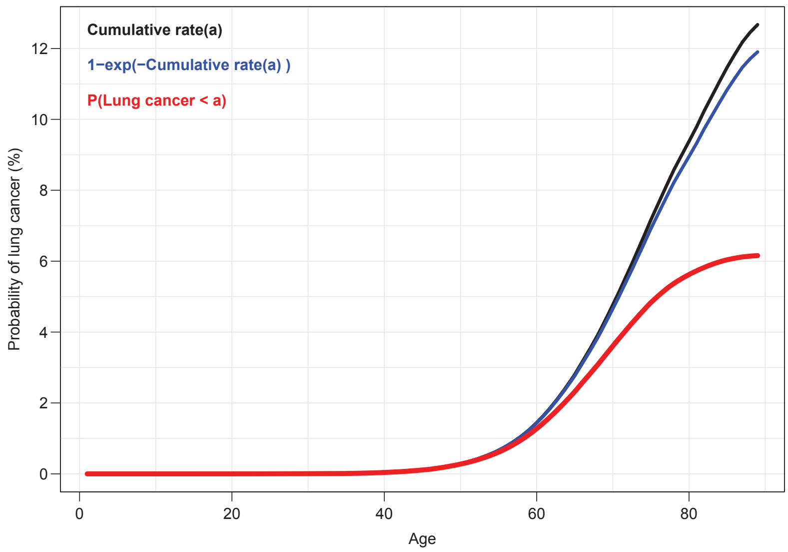
$$= \int_0^a \lambda(u) \exp\Big(-\big(\mathrm{M}(u) + \Lambda(u)\big)\Big)\,\mathrm{d}u$$

$\mathrm{M}(u)$ is the cumulative mortality rate.

$\Lambda(u)$ is the cumulative lung cancer incidence rate.

R-commands needed to do the calculations:

```
cr.death <- cumsum( D/Y )
cr.lung <- cumsum( lung/Y )
p.simple <- 1 - exp( -cr.lung )
p.lung <- cumsum( lung/Y *
                   exp( -(cr.death+cr.lung) ) )
matlines( age, 100*cbind( cr.lung, p.simple, p.lung ),
          type="l", lty=1, lwd=2*c(2,2,3),
          col=c("black","blue","red") )
```

Assumptions

- The calculation and the statement "6% of Danish males will get lung cancer" assumess that the lung cancer rates and the mortality rates in the file apply to a cohort of men.
- But they are cross-sectional rates, so the assumption is one of steady state of:
    1. mortality rates (which is dubious)
    2. lung cancer incidence rates (which is appalling).
- However, the machinery can be applied to any set of rates for competing risks, regardless of how they were estimated.

Interactions and timescales
Sunday 5 July, morning

Bendix Carstensen

Multistate Models with Multiple Time Scales
Modern Demographic Methods in Epidemiology
6 July 2014
27th IBC, Florence, 2014

`http://BendixCarstensen/AdvCoh/IBC2014`

Computational aspects of fitting models

- Cox model:
  - Only one timescale
  - Each person contributes one (or very few) records
  - Computationally simple, because time (risk / covariate) is profiled out in the estimation
  - Partial model, invariant under monotone transformation of the timescale
- Poisson modelling:
  - Many records per person
  - Very large datasets
  - Any number of timescales
  - Timeconsuming due to the large data sets
  - Full modelling of the rates as continuous functions of timescales
- Both are based on the same type of likelihood: small intervals with constant rate

Historical aspects

Whitehead J: Fitting Cox's regression model to survival data using GLIM. Applied Statistics, 29(3):268–275, 1980.[**?**][1]

Set up tables of event counts and person-years, classified by event times and covariate patterns.

Even with moderate datasets this can be large, albeit smaller than some 100 separate records per person.

[1] Recall **Keiding's law**: "Any result was published earlier than you think, even if you take Keiding's law into account."

Computational practicalities

Early 1980s: Fitting of Poisson models on datasets with 50,000 records were out of the question.
In particular with 100+ parameters.

Computationally feasible approaches to cohort studies were:

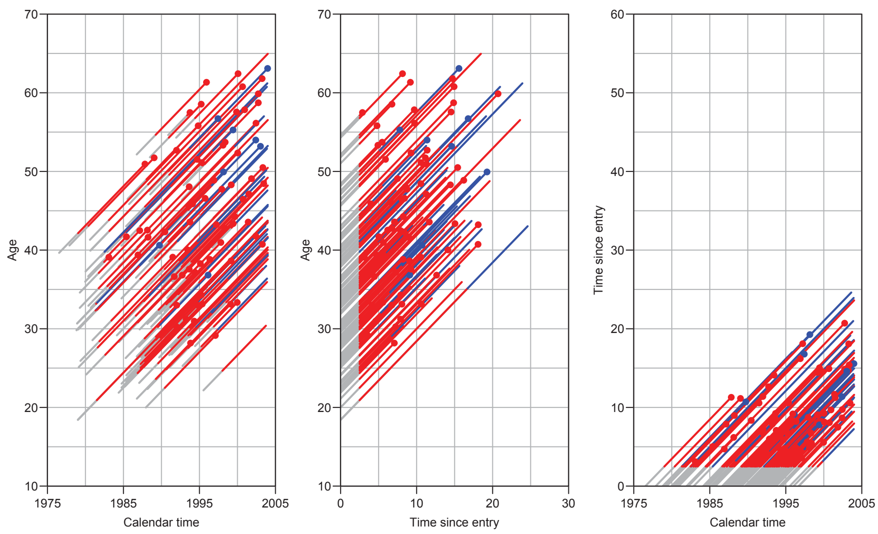
- Cox modelling — thanks to computational elegance.
- Time-splitting and tabulation in broad intervals before modelling.

The tabulation legacy (curse)

The **computational** need for tabulation has influenced thinking in epidemiology / demography:

- Life-tables in 1-year intervals.
- Rates are regarded in 5-year age by period intervals. Used for analysis of mortality and incidence rates based on registers.
  Age-period-cohort models with one parameter per level of the age/period factor.
- Yet, survival analysis is largely based on "time to event" methods (Kaplan-Meier, Cox), even from cancer registries — only one timescale.

Representation of follow-up



Age at entry as covariate

t: time since entry
e: age at entry
$a = e + t$: current age

$$\log\big(\lambda(a, t)\big) = f(t) + \beta e = (f(t) - \beta t) + \beta a$$

Immaterial whether a or e is used as (log)-*linear* covariate as long as t is in the model.

In a Cox-model with time since entry as time-scale, only the baseline hazard will change if age at entry is replaced by current age (a time-dependent variable).

"Controlling for age"

Including age at entry:

- Linear effect.
- Grouped variable.
- Parametric function.

— still only controls for the **linear** effect of **current age**.

Non-linear effects of time-scales

Arbitrary effects of the three variables t, a and e: Genuine extension of the model.

$$\log\big(\lambda(a, t, x_i)\big) = f(t) + g(a) + h(e) + \eta_i$$

Three quantities can be arbitrarily moved between the three functions:

$$\begin{aligned}\tilde{f}(t) &= f(a) - \mu_a - \mu_e + \gamma t \\ \tilde{g}(a) &= g(p) + \mu_a \qquad\quad - \gamma a \\ \tilde{h}(e) &= h(c) \qquad\quad + \mu_a + \gamma e\end{aligned}$$

because $t - a + e = 0$.
How many timescales in this model?

"Controlling for age"

— is not a well defined statement.

- Mostly it means that age **at entry** is included in the model.
- But ideally one would check whether there were non-linear effects of age at entry and current age.
- Requires modelling of multiple timescales.
- . . . and test of which ones are the relevant ones

⇒ splitting follow-up and modelling the timescales explicitly.

An worked example is in [**?**].

Several timescales: Caveat

As an example, consider:
t: time since entry
e: age at entry
$a = e + t$: current age

The relation: $a = t + e$ must hold for all units of analysis.

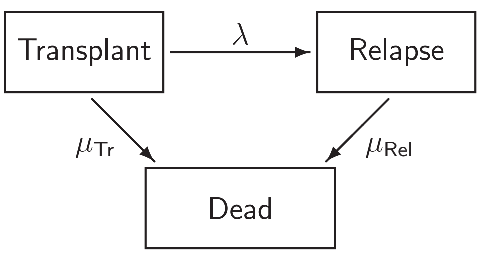
In general:
The difference between two time-scales must be constant within individuals.

Time dependent variable (new state)

How does relapse influence the mortality?

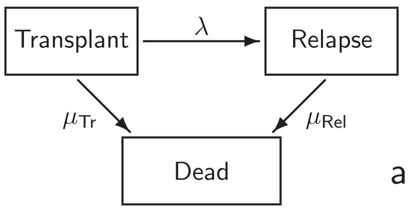
$$\lambda(t) = \lambda_0(t)\exp\big(1\{\text{relapse}\}(t) \times \beta\big)$$

i.e. when remission occurs, mortality increase by e^{β}.



What transitions are modelled here?

Time-dependent variable



If we take

$$1\{\text{Relapse}\}(t)$$

as time-dependent variable, we assume that μ_{r} and μ_{Rel} are proportional on the same timescale — no disease duration! — and λ is not modelled at all.

Fullt pobability statements require also modellng if the realpse rate λ

Stratified model

A popular version of the Cox-model allowing for non-proportionality is the **stratified model**:

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

where s refers to levels of a factor S.

- This is but a completely general **interaction** between the factor S and the chosen timescale.
- A better approach to interactions would be to specify a clinically founded form of interaction, so that test for interaction is against a specific (and sensible) alternative.

Time varying coefficients

This is a concept introduced by letting (some of) the parameters depend on time:

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

- This is also an interaction, but restricted: The effect of a covariate is linear for any value of t.
- If the covariate is a factor, then we just have a reparametrization of the stratified model.

Poisson modelling of interactions

When interactions are needed (or desired):

- use the familiar terminology of interaction as known from (generalized) linear models.
- use clinical judgement of which interactions are relevant.
- use clinical judgement of which forms of interaction are relevant.
- are interactions with time of special interest?

Poisson model for time-split data

- Clarifies the destinction between (risk) time as response variable and time(scales) as covariates.
- Multiple timescales easily handled.
- Smooth hazard rates by standard methods.
- More credible estimates of survival functions.
- Sensible modelling of interactions between timescales and other variables — for example **states**
- Interactions are called interactions.

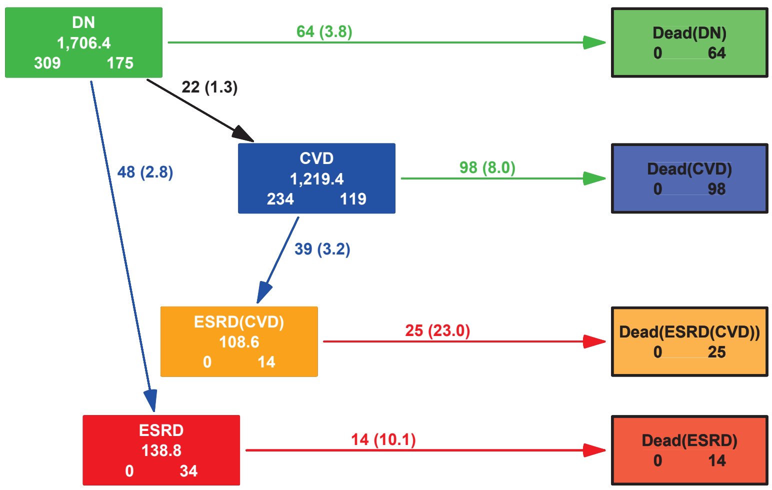
Simulation of follow-up

Sunday 5 July, morning

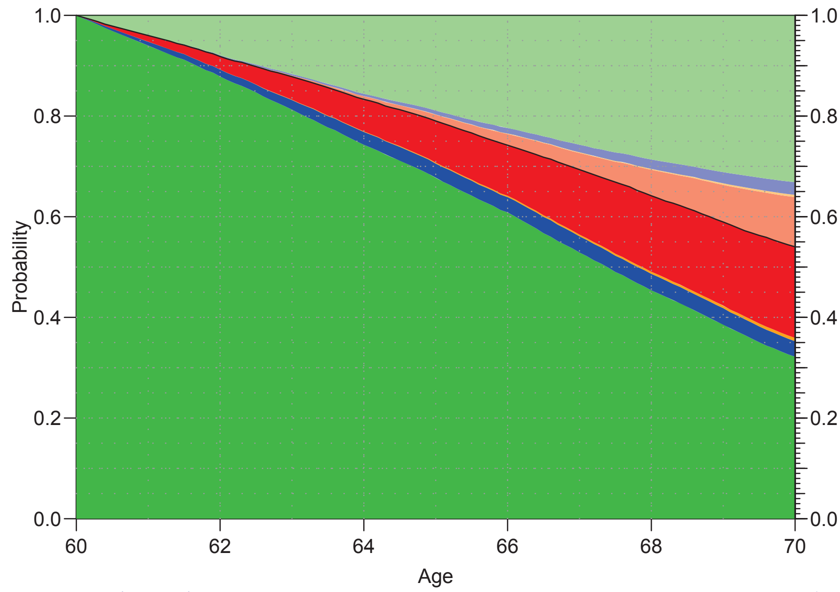
Bendix Carstensen

Multistate Models with Multiple Time Scales
Modern Demographic Methods in Epidemiology
6 July 2014
27th IBC, Florence, 2014
`http://BendixCarstensen/AdvCoh/IBC2014`

A more complicated multistate model



A more complicated multistate model



State probabilities

How do we get from rates to probabilities:

- 1: Analytical calculations:
  - immensely complicated formulae
  - computationally fast (once implemented)
  - difficult to generalize
- 2: Simulation of persons' histories
  - conceptually simple
  - computationally not quite simple
  - easy to generalize
- In the example the analytical option is effectively intractable

Simulation of a survival time

- For a rate function $\lambda(t)$, $\Lambda(t) = \int_0^t \lambda(s)\,\mathrm{d}s$:

$$S(t) = \exp\big(-\Lambda(t)\big)$$

- Simulate a survival probability $u \in [0,1]$:

$$u = S(t) \quad \Leftrightarrow \quad \Lambda(t) = -\log(u)$$

- Knowledge of $\Lambda(t)$ makes it easy to find a survival time.

Simulation of a survival time

Simulated random variate: u:

$$u = 0.853 \quad \Leftrightarrow \quad -\log(u) = 0.159$$

Look up 0.159 in the
table of the cumulative rates $\Lambda(t)$:

```
time  Lambda
 ...
 1.2   0.131
 1.3   0.151
 1.4   0.165
 1.5   0.181
 ...
```

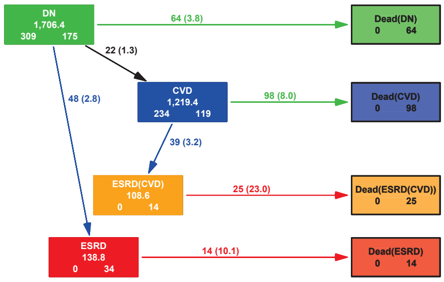
Linear interpolation gives:

$$t = 1.3{+}0.1{\times}(0.159{-}0.151)/(0.165{-}0.151) = 1.357$$

Simulation of one survival time

- Cumulative rates as a function of time
- Obtained from a model for the mortality rates:
  - Cox-model:
    Cumulative incidence directly — the Breslow estimator
  - Poisson model:
    Estimated incidence rates cumulated
  - ...
- Simulate survival probability
- Invert to time by look-up in table

Simulation in a multistate model

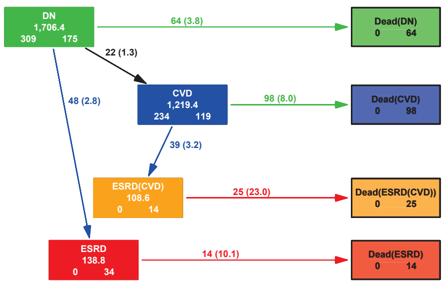


- Simulate a "survival time" for each possible transition **out** of a state.
- The smallest of these is the transition time.
- Choose the corresponding transition type as transition.

Multiple timescales

- The simulation just needs the cumulative rate (or survival function) for a person entering a state
- Therefore multiple timescales are easily accommodated, they just appear as variables in the model
- The tricky thing is to **update** the time-scales at every transition
- That is why a `Lexis` object is needed — the timescales are defined in the object

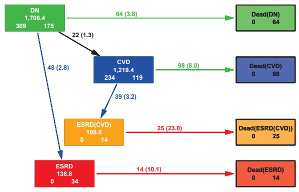
Transition object are `glms`




```
Tr <- list( "DN" = list( "Dead(DN)"   = E1d,
                          "CVD"        = E1c,
                          "ESRD"       = E1e ),
           "CVD" = list( "Dead(CVD)"  = E1d,
                          "ESRD(CVD)"  = E1e ),
          "ESRD" = list( "Dead(ESRD)"= E1n ),
     "ESRD(CVD)" = list( "Dead(ESRD(CVD))"= E1n ) )
```

Construction of the glms

```
E1d <- glm( lex.Xst %in% c("Dead(DN)","Dead(CVD)") ~
                                  Ns( age, kn=a.kn ) +
                                  Ns( dur, kn=d.kn ) +
                                  Ns( tfn, kn=n.kn ) +
                                  (...) +
                                  I(lex.Cst=="CVD"),
            offset = log(lex.dur),
            family = poisson,
   data = subset( S5, lex.Cst %in% c("DN","CVD") ) )

E1c <- update( E1d, (lex.Xst=="CVD") ~ .,
                    data = subset( S5, lex.Cst=="DN" )
```



simLexis

Input required:

- A Lexis object representing the initial state of the persons to be simulated. (lex.dur and lex.Xst will be ignored.)
- A transition object with the estimated Poisson models collected in a list of lists.

Output produced:

- A Lexis object with simulated event histories.
- Use nState to count how many persons in each state at different times.

Using simLexis

Put one record a new Lexis object (init, say). representing a person with the desired covariates.

Must have same structure as the one used for estimation:

```
init <- subset( S5, FALSE,
                select=c(timeScales(S5),"lex.Cst",
                      "dm.type","sex","hba1c",
                      "sys.bt","tchol","alb",
                      "smoke","bmi","gfr","hmgb",
                      "ins.kg") )
init[1,"sex"] <- "M"
init[1,"age"] <- 60
...

sim1 <- simLexis( Tr1, init,
                  time.pts=seq(0,25,0.2),
                  N=500 ) )
```

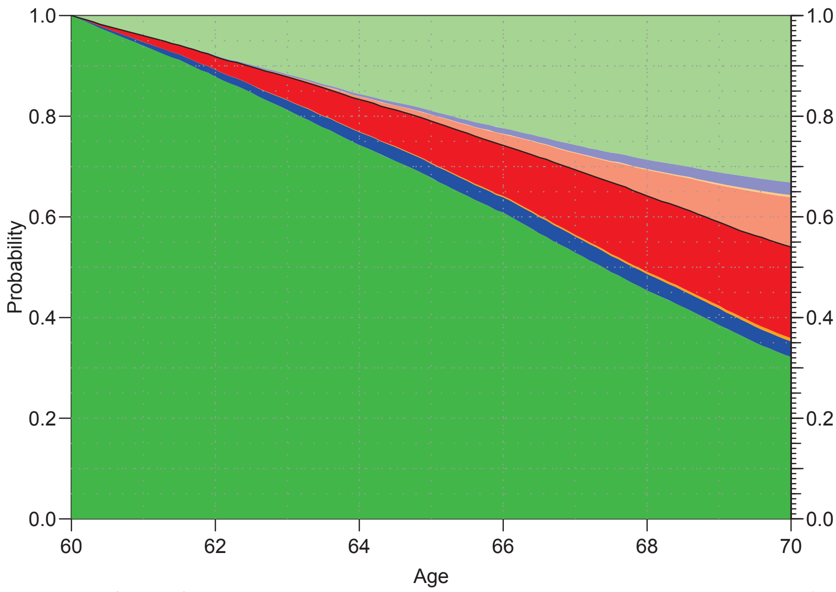
Output from `simLexis`

```
> summary( sim1 )

Transitions:
     To
From         DN CVD ES(CVD)  ES Dead(CVD) Dead(ES(CVD)) Dead(ES) Dead(DN)
  DN        212  81       0 145         0             0        0       62
  CVD         0  50       7   0        24             0        0        0
  ESRD(CVD)   0   0       3   0         0             4        0        0
  ESRD        0   0       0  70         0             0       75        0
  Sum       212 131      10 215        24             4       75       62

Transitions:
     To
From         Records:  Events: Risk time:  Persons:
  DN              500      288    9245.95       500
  CVD              81       31     667.90        81
  ESRD(CVD)         7        4      45.72         7
  ESRD            145       75     891.11       145
  Sum             733      398   10850.67       500
```


Using a simulated `Lexis` object

```
nw1 <- pState( nState( sim1,
                       at = seq(0,15,0.1),
                       from = 60,
                       time.scale = "age" ),
               perm = c(1:4,7:5,8) ) )
head( pState )
when       DN    CVD ES(CVD)     ES Dead(ES) Dead(ES(CVD)) Dead(
  60   1.0000 1.0000  1.0000 1.0000  1.0000        1.0000    1.
  60.1 0.9983 0.9986  0.9986 0.9997  0.9997        0.9997    0.
  60.2 0.9954 0.9964  0.9964 0.9990  0.9990        0.9990    0.
  60.3 0.9933 0.9947  0.9947 0.9981  0.9981        0.9981    0.
  60.4 0.9912 0.9929  0.9929 0.9973  0.9973        0.9973    0.
  60.5 0.9894 0.9913  0.9913 0.9964  0.9964        0.9964    0.

plot( pState )
```


Simulated probabilities



How many persons should you simulate?

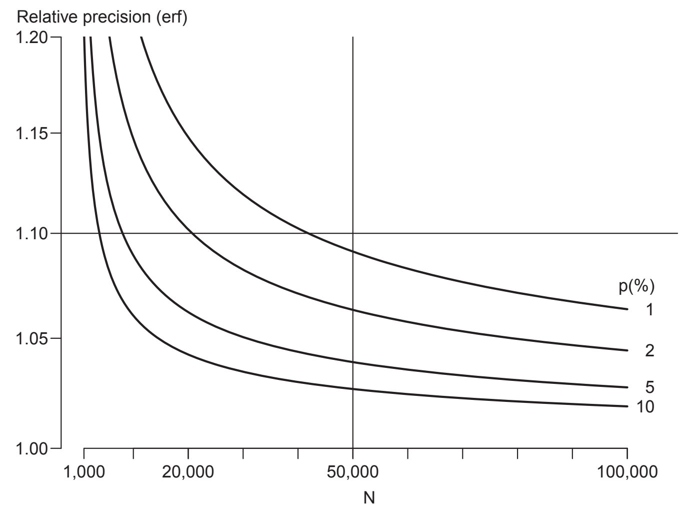
- All probabilities have the same denominator — the initial number of persons in the simulation, N, say.
- Thus, any probability will be of the form $p = x/N$
- For small probabilities we have that:

$$\text{s.e.}\big(\log(\hat{p})\big) = (1-p)/\sqrt{Np(1-p)}$$

- So c.i. of the form $p \overset{\times}{\div} \text{erf}$ where:

$$\text{erf} = \exp\big(1.96 \times (1-p)/\sqrt{Np(1-p)}\big)$$

Precision of simulated probabilities



Your turn: the `sim-Lexis` exercise / demo

Multistate model overview

- Clarify what the relevant states are
- Allows proper estimation of transition rates
- — and relationships between them
- Separate model for each transition (arrow)
- The usual survival methodology to compute probabilities breaks down
- Simulation allows estimation of cumulative probabilities:
  - Estimate transition rates (as usual)
  - Simulate probabilities (**not** as usual)