Statistical Analysis in the Lexis Diagram:
Age-Period-Cohort models

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

Max Planck Institut for Demographic Research, Rostock
May 2016

http://BendixCarstensen/APC/MPIDR-2016
Introduction

Statistical Analysis in the
Lexis Diagram:
Age-Period-Cohort models
May 2016
Max Planck Institut for Demographic Research, Rostock
http://BendixCarstensen/APC/MPIDR-2016
Welcome

- **Purpose of the course:**
  - knowledge about APC-models
  - technical knowledge of handling them
  - insight in the basic concepts of survival analysis

- **Remedies of the course:**
  - Lectures with handouts (BxC)
  - Practicals with suggested solutions (BxC)
  - Assignment for Thursday
Scope of the course

- Rates as observed in populations — disease registers for example.
- Understanding of survival analysis (statistical analysis of rates) — this is the content of much of the first day.
- Besides concepts, practical understanding of the actual computations (in R) are emphasized.
- There is a section in the practicals: “Basic concepts in analysis of rates and survival” — read it.
About the lectures

- Please interrupt:
  Most likely I did a mistake or left out a crucial argument.
- The handouts are not perfect
  — please comment on them,
  prospective students would benefit from it.
- There is a time-schedule in the practicals.
  It might need revision as we go.
About the practicals

- You should use your preferred R-environment.
- Epi-package for R is needed.
- Data are all on my website.
- Try to make a text version of the answers to the exercises — it is more rewarding than just looking at output. The latter is soon forgotten.
- An opportunity to learn emacs, ESS and Sweave?
Rates and Survival

Statistical Analysis in the Lexis Diagram:
Age-Period-Cohort models
May 2016
Max Planck Institut for Demographic Research, Rostock
http://BendixCarstensen/APC/MPIDR-2016
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 from jail release to re-offending.
Each line a person
Each blob a death
Study ended at 31 Dec. 2003

Rates and Survival (surv-rate)
Ordered by date of entry

Most likely the order in your database.
Timescale changed to “Time since diagnosis”.

Rates and Survival (surv-rate)

10/327
Patients ordered by survival time.
Survival times grouped into bands of survival.
Patients ordered by survival status within each band.
## Survival after Cervix cancer

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

Estimated risk in year 1 for Stage I women is $\frac{5}{107.5} = 0.0465$

Estimated 1 year survival is $1 - 0.0465 = 0.9535$ — Life-table estimator.
Survival function

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

How long they survive, survival time $T$ — a stochastic variable.
Distribution is characterized by the survival function:

\[
S(t) = P \{ \text{survival at least till } t \} = P \{ T > t \} = 1 - P \{ T \leq t \} = 1 - F(t)
\]
Intensity or rate

\[ \lambda(t) = \frac{P \{ \text{event in } (t, t + h] \mid \text{alive at } t \}}{h} = \frac{F(t + h) - F(t)}{S(t) \times h} \]

\[ = - \frac{S(t + h) - S(t)}{S(t)h} \xrightarrow{h \to 0} - \frac{d\log S(t)}{dt} \]

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

\[- \frac{d \log S(t)}{dt} = \lambda(t) \]

\[\Downarrow\]

\[S(t) = \exp \left( - \int_0^t \lambda(u) \, du \right) = \exp (-\Lambda(t)) \]

\[\Lambda(t) = \int_0^t \lambda(s) \, ds\] is called the **integrated intensity** or **cumulative hazard**.

\[\Lambda(t)\] is **not** an intensity — it is dimensionless.
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
- A rate is an **instantaneous** measure.
- **Note:** A cumulative measure requires an origin!
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 truncated).

- Epidemiological studies:
  Observation of (components of) a rate:

  \[ D, \quad Y, \quad D/Y \]

  \( D \): no. events, \( Y \) no of person-years.
Empirical rates for individuals

- At the individual level we introduce the empirical rate: \((d, y)\),
  - no. of events \((d \in \{0, 1\})\) during \(y\) risk time
- Each person may contribute several empirical \((d, y)\)
- Empirical rates are responses in survival analysis
- The timescale is a covariate:
  - varies across empirical rates from one individual: Age, calendar time, time since diagnosis
- Do not confuse timescale with
  - \(y\) — risk time (exposure in demography)
  - a difference between two points on any timescale
Empirical rates by calendar time.
Empirical rates by time since diagnosis.
Two timescales

Note that we actually have two timescales:

- Time since diagnosis (i.e. since entry into the study)
- Calendar time.

These can be shown simultaneously in a Lexis diagram.
Follow-up by calendar time and time since diagnosis:

A Lexis diagram!
Empirical rates by calendar time and time since diagnosis.
Likelihood for rates

Statistical Analysis in the Lexis Diagram:
Age-Period-Cohort models
May 2016
Max Planck Institut for Demographic Research, Rostock
http://BendixCarstensen/APC/MPIDR-2016
Likelihood contribution from one person

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

\[ P \{ \text{event at } t_4 \mid \text{alive at } t_0 \} = P \{ \text{event at } t_4 \mid \text{alive at } t_3 \} \times \]
\[ P \{ \text{survive } (t_2, t_3) \mid \text{alive at } t_2 \} \times \]
\[ P \{ \text{survive } (t_1, t_2) \mid \text{alive at } t_1 \} \times \]
\[ P \{ \text{survive } (t_0, t_1) \mid \text{alive at } t_0 \} \]

Likelihood contribution from one individual is a **product** of terms.

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

\[- y = t_i - t_{i-1} \text{ (mostly } d = 0).\]
Likelihood for an empirical rate

- Likelihood depends on **data** and the **model**
- Model: the rate is constant in the interval.
- The interval should sufficiently small for this assumption to be reasonable.

\[
L(\lambda|y, d) = P\{\text{survive } y\} \times P\{\text{event}\}^d = e^{-\lambda y} \times (\lambda dt)^d = \lambda^d e^{-\lambda y}
\]

\[
\ell(\lambda|y, d) = d \log(\lambda) - \lambda y
\]
Probability

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

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

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

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

log-Likelihood

\[ d \log(\lambda) - \lambda y \]

\[ = 0 \log(\lambda) - \lambda y_1 \]

\[ + 0 \log(\lambda) - \lambda y_2 \]

\[ + d \log(\lambda) - \lambda y_3 \]
Probability log-Likelihood

\[ P(\text{surv } t_0 \rightarrow t_x | \text{entry } t_0) = P(\text{surv } t_0 \rightarrow t_1 | \text{entry } t_0) \times P(\text{surv } t_1 \rightarrow t_2 | \text{entry } t_1) \times P(\text{surv } t_2 \rightarrow t_x | \text{entry } t_2) \]

= \log(\lambda) - \lambda y

log-Likelihood

\[ = 0 \log(\lambda) - \lambda y_1 + 0 \log(\lambda) - \lambda y_2 + 0 \log(\lambda) - \lambda y_3 \]
Probability

\[ P(\text{event at } t_x|\text{entry } t_0) \]
\[ = P(\text{surv } t_0 \rightarrow t_1|\text{entry } t_0) \]
\[ \times P(\text{surv } t_1 \rightarrow t_2|\text{entry } t_1) \]
\[ \times P(\text{event at } t_x|\text{entry } t_2) \]

log-Likelihood

\[ 1 \log(\lambda) - \lambda y \]
\[ = 0 \log(\lambda) - \lambda y_1 \]
\[ + 0 \log(\lambda) - \lambda y_2 \]
\[ + 1 \log(\lambda) - \lambda y_3 \]
Aim of dividing time into bands:

- Compute rates in different bands of:
  - age
  - calendar time
  - disease duration
  - ...

- Allow rates to vary along the timescale:

\[
\begin{align*}
0 \log(\lambda) - \lambda y_1 &+ 0 \log(\lambda) - \lambda y_2 + d \log(\lambda) - \lambda y_3 \\
\rightarrow &+ 0 \log(\lambda_1) - \lambda_1 y_1 + 0 \log(\lambda_2) - \lambda_2 y_2 + d \log(\lambda_3) - \lambda_3 y_3
\end{align*}
\]
Log-likelihood from more persons

- One person, \( p \): \[ \sum_t \left( d_{pt} \log(\lambda_t) - \lambda_t y_{pt} \right) \]
- More persons: \[ \sum_p \sum_t \left( d_{pt} \log(\lambda_t) - \lambda_t y_{pt} \right) \]
- Collect terms with identical values of \( \lambda_t \):

\[
\sum_t \sum_p \left( d_{pt} \log(\lambda_t) - \lambda_t y_{pt} \right) = \sum_t \left( \left( \sum_t d_{pt} \right) \log(\lambda_t) - \lambda_t \left( \sum_t y_{pt} \right) \right) \\
= \sum_t \left( D_t \log(\lambda_t) - \lambda_t Y_t \right)
\]

- All events in interval \( t \) ("at" time \( t \)), \( D_t \)
- All exposure time in interval \( t \) ("at" time \( t \)), \( Y_t \)
Likelihood example

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

\[
P\{D = 7, Y = 500|\lambda\} = \lambda^D e^{\lambda Y} \times K = \lambda^7 e^{\lambda 500} \times K = L(\lambda|\text{data})
\]

- Best guess of \( \lambda \) is where this function is as large as possible.
- Confidence interval is where it is not too far from the maximum
Log-likelihood ratio

Rate parameter, \( \lambda \)

Likelihood for rates (likelihood)
Log-likelihood ratio

![Log-likelihood ratio graph](image)

Rate parameter, $\lambda$ (per 1000)
Log-likelihood ratio

Likelihood for rates (log-likelihood)

Rate parameter, $\lambda$ (per 1000)
\hat{\lambda} = \frac{7}{500} = 14
\hat{\lambda} \times \exp(1.96/\sqrt{7}) = (6.7, 29.4)
Poisson likelihood

Log-likelihood contribution from one individual, \( p \), say, is:

\[
\ell_{FU}(\lambda | d, y) = d_p t \log(\lambda(t)) - \lambda(t)y_{pt}, \quad t = 1, \ldots, t_p
\]

Log-likelihood from independent Poisson observations \( d_{pt} \) with mean \( \mu = \lambda(t)y_{pt} \):

\[
\ell_{\text{Poisson}}(\lambda y | d) = d_p t \log(\lambda(t)y_{pt}) - \lambda(t)y_{pt} = \ell_{FU}(\lambda | d, y) + d_p t \log(y_{pt})
\]

Extra term does not depend on the rate parameter \( \lambda \).
Poisson likelihood

Log-likelihood contribution from one individual, $p$, say, is:

$$\ell_{FU}(\lambda|d, y) = d_{pt}\log(\lambda(t)) - \lambda(t)y_{pt}, \quad t = 1, \ldots, t_p$$

- Terms are not independent,
- but the log-likelihood is a sum of Poisson-like terms,
- the same as a likelihood for independent Poisson variates, $d_{pt}$
- with mean $\mu = \lambda_t y_{py} \Leftrightarrow \log \mu = \log(\lambda_t) + \log(y_{py})$

⇒ Analyse rates $\lambda$ based on empirical rates $(d, y)$ Poisson model with log-link applied to where:

- $d$ is the response variable.
- $\log(y)$ is the offset variable.
Likelihood for follow-up of many subjects

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.
The log-likelihood is maximal for:

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

Information about the log-rate \( \theta = \log(\lambda) \):

\[
\ell(\theta|D, Y) = D\theta - e^\theta Y, \quad \ell'_\theta = D - e^\theta Y, \quad \ell''_\theta = -e^\theta Y
\]

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

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

Note that this only depends on the no. events, \textbf{not} on the follow-up time.
The log-likelihood is maximal for:

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

Information about the rate itself, \( \lambda \):

\[
\ell(\lambda| D, Y) = D \log(\lambda) - \lambda Y \quad \ell'_\lambda = \frac{D}{\lambda} - Y \quad \ell''_\lambda = -\frac{D}{\lambda^2}
\]

so \( I(\hat{\lambda}) = \frac{D}{\hat{\lambda}^2} = \frac{Y^2}{D} \), hence \( \text{var}(\hat{\lambda}) = D/Y^2 \)

Standard error of a rate: \( \sqrt{D}/Y \).
Confidence interval for a rate

A 95% confidence interval for the log of a rate is:

\[ \hat{\theta} \pm 1.96/\sqrt{D} = \log(\lambda) \pm 1.96/\sqrt{D} \]

Take the exponential to get the confidence interval for the rate:

\[ \lambda \div \exp(1.96/\sqrt{D}) \]

error factor, erf

Alternatively do the c.i. directly on the rate scale:

\[ \lambda \pm 1.96\sqrt{D} / Y \]
Exercise

Suppose we have 17 deaths during 843.6 years of follow-up. Calculate the mortality rate with a 95% c.i.
Rates with `glm`

```r
> library(Epi)
> D <- 17
> Y <- 843.6/1000
> round( ci.exp( glm( D ~ 1, offset=log(Y), family=poisson ) ), 2 )

       exp(Est.)  2.5%  97.5%
(Intercept) 20.15 12.53 32.42

> round( ci.exp( glm( D/Y ~ 1, weight= Y , family=poisson ) ), 2 )

       exp(Est.)  2.5%  97.5%
(Intercept) 20.15 12.53 32.42

> round( ci.exp( glm( D/Y ~ 1, weight= Y , family=poisson(link="identity")),
+        Exp=FALSE), 2 )

          Estimate  2.5%  97.5%
(Intercept) 20.15 10.57 29.73
```
Ratio of two rates

If we have observations two rates \( \lambda_1 \) and \( \lambda_0 \), based on \((D_1, Y_1)\) and \((D_0, Y_0)\) the variance of the log of the ratio of the rates, \( \log(\text{RR}) \), is:

\[
\text{var}(\log(\text{RR})) = \text{var}(\log(\frac{\lambda_1}{\lambda_0})) = \text{var}(\log(\lambda_1)) + \text{var}(\log(\lambda_0)) = \frac{1}{D_1} + \frac{1}{D_0}
\]

As before, a 95% c.i. for the RR is then:

\[
\text{RR} \times \exp \left(1.96 \sqrt{\frac{1}{D_1} + \frac{1}{D_0}}\right)
\]

error factor
Exercise

Suppose we in group 0 have 17 deaths during 843.6 years of follow-up in one group, and in group 1 have 28 deaths during 632.3 years.

Calculate the rate-ratio between group 1 and 0 with a 95% c.i.
Lifetables

Statistical Analysis in the Lexis Diagram:
Age-Period-Cohort models
May 2016
Max Planck Institut for Demographic Research, Rostock
http://BendixCarstensen/APC/MPIDR-2016

lifetable
The life table method

The simplest analysis is by the “life-table method”:

\[
\begin{array}{cccccc}
\text{interval} & \text{alive} & \text{dead} & \text{cens.} & p_i \\
i & n_i & d_i & l_i & \\
1 & 77 & 5 & 2 & 5/(77 - 2/2) = 0.066 \\
2 & 70 & 7 & 4 & 7/(70 - 4/2) = 0.103 \\
3 & 59 & 8 & 1 & 8/(59 - 1/2) = 0.137 \\
\end{array}
\]

\[
p_i = P\{\text{death in interval } i\} = 1 - d_i/(n_i - l_i/2)
\]

\[
S(t) = (1 - p_1) \times \cdots \times (1 - p_t)
\]
The life table method

The life-table method computes survival probabilities for each time interval, in demography normally one year.

The rate is the number of deaths $d_i$ divided by the risk time $\left(n_i - d_i/2 - l_i/2\right) \times \ell_i$:

$$\lambda_i = \frac{d_i}{\left(n_i - d_i/2 - l_i/2\right) \times \ell_i}$$

and hence the death probability:

$$p_i = 1 - \exp\left(-\lambda_i \ell_i\right) = 1 - \exp\left(-\frac{d_i}{\left(n_i - d_i/2 - l_i/2\right)}\right)$$

The modified life-table estimator.
### Population life table, DK 1997–98

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

Lifetables (lifetable)
Danish life tables 1997−98

\[ \log_2 \left( \text{mortality per 10}^5 \text{ (40−85 years)} \right) \]

Men: \(-14.244 + 0.135 \text{ age}\)

Women: \(-14.877 + 0.135 \text{ age}\)
Swedish life tables 1997–98

\[ \log_2(\text{mortality per} \ 10^5 \ (40–85 \text{ years})) \]

Men: \[-15.453 + 0.146 \text{ age} \]

Women: \[-16.204 + 0.146 \text{ age} \]

Lifetables (lifetable)
Practical

Based on the previous slides answer the following for both Danish and Swedish lifetables:

- What is the doubling time for mortality?
- What is the rate-ratio between males and females?
- How much older should a woman be in order to have the same mortality as a man?
<table>
<thead>
<tr>
<th></th>
<th>Males</th>
<th>Females</th>
</tr>
</thead>
<tbody>
<tr>
<td>Denmark</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \log_2(\lambda(a)) )</td>
<td>(-14.244 + 0.135 \text{ age} )</td>
<td>(-14.877 + 0.135 \text{ age} )</td>
</tr>
<tr>
<td>Doubling time</td>
<td>(1/0.135 = 7.41 \text{ years})</td>
<td></td>
</tr>
<tr>
<td>M/F rate-ratio</td>
<td>(2^{-14.244+14.877} = 2^{0.633} = 1.55)</td>
<td></td>
</tr>
<tr>
<td>Age-difference</td>
<td>((-14.244 + 14.877)/0.135 = 4.69 \text{ years})</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sweden</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \log_2(\lambda(a)) )</td>
<td>(-15.453 + 0.146 \text{ age} )</td>
<td>(-16.204 + 0.146 \text{ age} )</td>
</tr>
<tr>
<td>Doubling time</td>
<td>(1/0.146 = 6.85 \text{ years})</td>
<td></td>
</tr>
<tr>
<td>M/F rate-ratio</td>
<td>(2^{-15.453+16.204} = 2^{0.751} = 1.68)</td>
<td></td>
</tr>
<tr>
<td>Age-difference</td>
<td>((-15.453 + 16.204)/0.146 = 5.14 \text{ years})</td>
<td></td>
</tr>
</tbody>
</table>
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.
Life table approach

The observation of interest is not the survival time of the individual.

It is the population experience:

\[ D: \text{ Deaths (events).} \]
\[ Y: \text{ Person-years (risk time).} \]

The classical lifetable analysis compiles these for prespecified intervals of age, and computes age-specific mortality rates.

Data are collected cross-sectionally, but interpreted longitudinally.
Rates vary over time:

Finnish life tables 1986

log$_2$ (mortality per 10$^5$ (40–85 years))

Men: $-14.061 + 0.138 \text{age}$

Women: $-15.266 + 0.138 \text{age}$
Rates vary over time:

Finnish life tables 1994

\[ \log_2(\text{mortality per } 10^5 \text{ (40–85 years)}) \]

Men: \(-14.275 + 0.137 \text{ age}\)

Women: \(-15.412 + 0.137 \text{ age}\)
Rates vary over time:

\[
\log_2(\text{mortality per } 10^5 \text{ (40–85 years)}) \\
\text{Men: } -14.339 + 0.134 \text{ age} \\
\text{Women: } -15.412 + 0.134 \text{ age}
\]
Who needs the Cox-model anyway?

Statistical Analysis in the Lexis Diagram:
Age-Period-Cohort models
May 2016
Max Planck Institut for Demographic Research, Rostock
http://BendixCarstensen/APC/MPIDR-2016
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 it is less clear — \( t \) is but a covariate that varies within individual.
- Cox’s approach profiles \( \lambda_0(t) \) out.
Cox-likelihood

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

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

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

\[ \log(\lambda(t, x)) = \log(\lambda_0(t)) + x' \beta = \alpha_t + \eta \]
The Cox-likelihood as profile likelihood

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

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

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

Who needs the Cox-model anyway? (WntCma)
Suppose the time scale has been divided into small intervals with at most one death in each.

Assume w.l.o.g. the $y$s in the empirical rates all are 1.

Log-likelihood contributions that contain information on a specific time-scale parameter $\alpha_t$ will be from:

- the (only) empirical rate $(1, 1)$ with the death at time $t$.
- all other empirical rates $(0, 1)$ from those who were at risk at time $t$. 

Who needs the Cox-model anyway? (WntCma)
Note: There is one contribution from each person at risk to this part of the log-likelihood:

\[ \ell_t(\alpha_t, \beta) = \sum_{i \in R_t} d_i \log(\lambda_i(t)) - \lambda_i(t) y_i \]

\[ = \sum_{i \in R_t} \left\{ d_i (\alpha_t + \eta_i) - e^{\alpha_t + \eta_i} \right\} \]

\[ = \alpha_t + \eta_{\text{death}} - e^{\alpha_t} \sum_{i \in R_t} e^{\eta_i} \]

where \( \eta_{\text{death}} \) is the linear predictor for the person that died.
The derivative w.r.t. $\alpha_t$ is:

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

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

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

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

▸ The Poisson approach needs a dataset of empirical rates \((d, y)\)
  with suitably small values of \(y\).
▸ — much larger than the original dataset
▸ — 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 \(y\)
  ▸ Covariate value for the timescale
    (time since entry, current age, current date, \ldots)
  ▸ other covariates
▸ Modelling is by standard \texttt{glm} Poisson

Who needs the Cox-model anyway? (WntCma)
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? (WntCma)
Example: Mayo Clinic lung cancer I

```r
> round(cmp, 5)

<table>
<thead>
<tr>
<th></th>
<th>age</th>
<th></th>
<th>2.5%</th>
<th>97.5%</th>
<th>sex</th>
<th></th>
<th>2.5%</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cox</td>
<td>1.01716</td>
<td>0.99894</td>
<td>1.03571</td>
<td>0.59896</td>
<td>0.43137</td>
<td>0.83165</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poisson-factor</td>
<td>1.01716</td>
<td>0.99894</td>
<td>1.03571</td>
<td>0.59896</td>
<td>0.43137</td>
<td>0.83165</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poisson-spline</td>
<td>1.01619</td>
<td>0.99803</td>
<td>1.03468</td>
<td>0.59983</td>
<td>0.43199</td>
<td>0.83287</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
```

Who needs the Cox-model anyway? (WntCma)
Who needs the Cox-model anyway? (WntCma)
Who needs the Cox-model anyway? (WntCma)
```r
> 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))
```

Who needs the Cox-model anyway? (WntCma)
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.

⇒ difficult to access the baseline hazard.
⇒ uninitiated tempted to show survival curves where irrelevant
Modeling in this world

- Replace the \( \alpha_t \)'s by a parametric function \( f(t) \) with a limited number of parameters, for example:
  - Piecewise constant
  - Splines (linear, quadratic or cubic)
  - Fractional polynomials

- 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 in calculations of
  - expected residual life time
  - state occupancy probabilities in multistate models
  - ...
The baseline hazard and survival functions

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

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

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

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

1. Select timepoints \(t_i\) (fairly close).
2. Get estimates of log-rates \(f(t_i) = g(t_i) + x_0'\gamma\) for these points:
   \[
   \hat{f}(t_i) = B \hat{\beta}
   \]
   where \(\beta\) is the total parameter vector in the model.
3. Variance-covariance matrix of \(\hat{\beta}\): \(\hat{\Sigma}\).
4. Variance-covariance of \(\hat{f}(t_i)\): \(B\Sigma B'\).
5. Transformation to the rates is the coordinate-wise exponential function, with derivative \(\text{diag}\left[\exp(\hat{f}(t_i))\right]\)
6. Variance-covariance matrix of the rates at the points $t_i$:

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

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

$$\ell \times \begin{bmatrix}
1 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 \\
1 & 1 & 1 & 0 & 0 \\
1 & 1 & 1 & 1 & 0 \\
1 & 1 & 1 & 1 & 1 & 0
\end{bmatrix} \begin{bmatrix}
\hat{e}^f(t_1) \\
\hat{e}^f(t_2) \\
\hat{e}^f(t_3) \\
\hat{e}^f(t_4)
\end{bmatrix} = \mathbf{L} \begin{bmatrix}
\hat{e}^f(t_1) \\
\hat{e}^f(t_2) \\
\hat{e}^f(t_3) \\
\hat{e}^f(t_4)
\end{bmatrix}$$
8. Variance-covariance matrix for the cumulative hazard is:

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

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

Practical: Cox and Poisson modelling
(non)-Linear models: Estimates and predictions

Statistical Analysis in the Lexis Diagram:

Age-Period-Cohort models
May 2016
Max Planck Institut for Demographic Research, Rostock
http://BendixCarstensen/APC/MPIDR-2016
Linear models

```r
> library( Epi )
> data( diet )
> names( diet )

[1] "id"   "doe"  "dox"  "dob"  "y"  "fail"
[8] "month" "energy" "height" "weight" "fat" "fibre"
[15] "chd"

> with( diet, plot( weight ~ height, pch=16 ) )
> abline( lm( weight ~ height, data=diet ), col="red", lwd=2 )
```

(non)-Linear models: Estimates and predictions (lin-mod)
> with(diet, plot(weight ~ height, pch=16))
> abline(lm(weight ~ height, data=diet), col="red", lwd=2)
Linear models, extracting estimates

```r
> ml <- lm( weight ~ height, data=diet )
> summary( ml )

Call:
  lm(formula = weight ~ height, data = diet)

Residuals:
     Min      1Q  Median      3Q     Max
-24.7361 -7.4553  0.1608  6.9384  27.8130

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -59.91601   14.31557  -4.185 3.66e-05
height      0.76421     0.08252   9.261  < 2e-16

Residual standard error: 9.625 on 330 degrees of freedom
   (5 observations deleted due to missingness)
Multiple R-squared:  0.2063, Adjusted R-squared:  0.2039
F-statistic: 85.76 on 1 and 330 DF,  p-value: < 2.2e-16
```

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

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>StdErr</th>
<th>z</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-59.9160</td>
<td>14.3156</td>
<td>-4.19</td>
<td>0.0000</td>
</tr>
<tr>
<td>height</td>
<td>0.7642</td>
<td>0.0825</td>
<td>9.26</td>
<td>&lt;2.2e-16</td>
</tr>
</tbody>
</table>
```

(non)-Linear models: Estimates and predictions (lin-mod)
Linear models, prediction

\[
\text{ml} \leftarrow \text{lm( weight } \sim \text{ height, data=diet )}
\]

\[
\text{nd} \leftarrow \text{data.frame( height = 150:190 )}
\]

\[
\text{pr.co} \leftarrow \text{predict( ml, newdata=nd, interval="conf" )}
\]

\[
\text{pr.pr} \leftarrow \text{predict( ml, newdata=nd, interval="pred" )}
\]

\[
\text{with( diet, plot( weight } \sim \text{ height, pch=16 ) )}
\]

\[
\text{matlines( nd$height, pr.co, lty=1, lwd=c(5,2,2), col="blue" )}
\]

\[
\text{matlines( nd$height, pr.pr, lty=2, lwd=c(5,2,2), col="blue" )}
\]
non-Linear models, prediction

```r
> mq <- lm( weight ~ height + I(height^2), data=diet )
> pr.co <- predict( mq, newdata=nd, interval="conf" )
> pr.pr <- predict( mq, newdata=nd, interval="pred" )
> with( diet, plot( weight ~ height, pch=16 ) )
> matlines( nd$height, pr.co, lty=1, lwd=c(5,2,2), col="blue" )
> matlines( nd$height, pr.pr, lty=2, lwd=c(5,2,2), col="blue" )
```
Testis cancer

Testis cancer in Denmark:

```r
> 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 0 0 0 0 0 ...
  $ Y: num 39650 36943 34588 33267 32614 ...

> head( testisDK )

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

(non)-Linear models: Estimates and predictions (lin-mod)
## Cases, PY and rates

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

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>10.00</td>
<td>7.00</td>
<td>16.00</td>
<td>18.00</td>
<td>9.00</td>
<td>10.00</td>
<td>70.00</td>
</tr>
<tr>
<td></td>
<td>2604.66</td>
<td>4037.31</td>
<td>3884.97</td>
<td>3820.88</td>
<td>3070.87</td>
<td>2165.54</td>
<td>19584.22</td>
</tr>
<tr>
<td></td>
<td>0.38</td>
<td>0.17</td>
<td>0.41</td>
<td>0.47</td>
<td>0.29</td>
<td>0.46</td>
<td>0.36</td>
</tr>
<tr>
<td>10</td>
<td>13.00</td>
<td>27.00</td>
<td>37.00</td>
<td>72.00</td>
<td>97.00</td>
<td>75.00</td>
<td>321.00</td>
</tr>
<tr>
<td></td>
<td>2135.73</td>
<td>3505.19</td>
<td>4004.13</td>
<td>3906.08</td>
<td>3847.40</td>
<td>2260.97</td>
<td>19659.48</td>
</tr>
<tr>
<td></td>
<td>0.61</td>
<td>0.77</td>
<td>0.92</td>
<td>1.84</td>
<td>2.52</td>
<td>3.32</td>
<td>1.63</td>
</tr>
</tbody>
</table>

(non)-Linear models: Estimates and predictions ($\text{lin-mod}$)
Linear effects in \texttt{glm}

How do rates depend on age?

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

<table>
<thead>
<tr>
<th>Estimate</th>
<th>StdErr</th>
<th>z</th>
<th>P</th>
<th>2.5%</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-9.7755</td>
<td>0.0207</td>
<td>-472.3164</td>
<td>0</td>
<td>-9.8160</td>
</tr>
<tr>
<td>A</td>
<td>0.0055</td>
<td>0.0005</td>
<td>11.3926</td>
<td>0</td>
<td>0.0045</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>exp(Est.)</th>
<th>2.5%</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td>A</td>
<td>1.0055</td>
<td>1.0046</td>
</tr>
</tbody>
</table>
\end{verbatim}

Linear increase of log-rates by age

(non)-Linear models: Estimates and predictions (lin-mod)
Linear effects in `glm`

```r
> nd <- data.frame(A=15:60, Y=10^5)
> pr <- predict(ml, newdata=nd, type="link", se.fit=TRUE)
> str(pr)

List of 3
$ fit : Named num [1:46] 1.82 1.83 1.83 1.84 1.84 ...  
..- attr(*, "names")= chr [1:46] "1" "2" "3" "4" ...  
$ se.fit : Named num [1:46] 0.015 0.0146 0.0143 0.014 0.0137 ...  
..- attr(*, "names")= chr [1:46] "1" "2" "3" "4" ...  
$ residual.scale: num 1

> ci.mat()

<table>
<thead>
<tr>
<th>Estimate</th>
<th>2.5%</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>[1,]</td>
<td>1</td>
<td>1.000000 1.000000</td>
</tr>
<tr>
<td>[2,]</td>
<td>0</td>
<td>-1.959964 1.959964</td>
</tr>
</tbody>
</table>

> matplot(nd$A, exp( cbind(pr$fit,pr$se) %*% ci.mat() ),
+         type="l", lty=1, lwd=c(3,1,1), col="black", log="y" )
```

(non)-Linear models: Estimates and predictions (`lin-mod`)
Linear effects in glm

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

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>StdErr</th>
<th>z</th>
<th>P</th>
<th>2.5%</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-9.7755</td>
<td>0.0207</td>
<td>-472.3164</td>
<td>0</td>
<td>-9.8160</td>
<td>-9.7349</td>
</tr>
<tr>
<td>A</td>
<td>0.0055</td>
<td>0.0005</td>
<td>11.3926</td>
<td>0</td>
<td>0.0045</td>
<td>0.0064</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>[,1]</th>
<th>[,2]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>15</td>
</tr>
<tr>
<td>1</td>
<td>16</td>
</tr>
<tr>
<td>1</td>
<td>17</td>
</tr>
<tr>
<td>1</td>
<td>18</td>
</tr>
<tr>
<td>1</td>
<td>19</td>
</tr>
<tr>
<td>1</td>
<td>20</td>
</tr>
</tbody>
</table>

> matplot( nd$A, ci.exp( ml, ctr.mat=Cl ),
+       type="l", lty=1, lwd=c(3,1,1), col="black", log="y" )
Linear effects in \texttt{glm}

\begin{verbatim}
> matplot( nd$A, exp( cbind(pr$fit,pr$se) %*% ci.mat() ),
+          type="l", lty=1, lwd=c(3,1,1), col="black", log="y" )
\end{verbatim}
Linear effects in \texttt{glm}

\begin{verbatim}
> matplot( nd$A, ci.exp( ml, ctr.mat=Cl )*10^5,
+         type="l", lty=1, lwd=c(3,1,1), col="black", log="y" )
\end{verbatim}

(non)-Linear models: Estimates and predictions (lin-mod)
Quadratic effects in `glm`

How do rates depend on age?

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

<table>
<thead>
<tr>
<th>Estimate</th>
<th>StdErr</th>
<th>z</th>
<th>P</th>
<th>2.5%</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-12.3656</td>
<td>0.0596</td>
<td>-207.3611</td>
<td>0</td>
<td>-12.4825</td>
</tr>
<tr>
<td>A</td>
<td>0.1806</td>
<td>0.0033</td>
<td>54.8290</td>
<td>0</td>
<td>0.1741</td>
</tr>
<tr>
<td>I(A^2)</td>
<td>-0.0023</td>
<td>0.0000</td>
<td>-53.7006</td>
<td>0</td>
<td>-0.0024</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>exp(Est.)</th>
<th>2.5%</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>A</td>
<td>1.1979</td>
<td>1.1902</td>
</tr>
<tr>
<td>I(A^2)</td>
<td>0.9977</td>
<td>0.9976</td>
</tr>
</tbody>
</table>
```
Quadratic effect in `glm`

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

<table>
<thead>
<tr>
<th>Estimate</th>
<th>StdErr</th>
<th>z</th>
<th>P</th>
<th>2.5%</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-12.3656</td>
<td>0.0596</td>
<td>-207.3611</td>
<td>0</td>
<td>-12.4825</td>
</tr>
<tr>
<td>A</td>
<td>0.1806</td>
<td>0.0033</td>
<td>54.8290</td>
<td>0</td>
<td>0.1741</td>
</tr>
<tr>
<td>I(A^2)</td>
<td>-0.0023</td>
<td>0.0000</td>
<td>-53.7006</td>
<td>0</td>
<td>-0.0024</td>
</tr>
</tbody>
</table>

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

```
[,1] [,2] [,3]
[1,] 1  15  225
[2,] 1  16  256
[3,] 1  17  289
[4,] 1  18  324
[5,] 1  19  361
[6,] 1  20  400
```

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

(non)-Linear models: Estimates and predictions (`lin-mod`)
Quadratic effect in \textit{glm}

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

(non)-Linear models: Estimates and predictions (lin-mod)
Quadratic effect in glm

\[
de+c.e(mq, ctr.mat = Cq) \times 10^5
\]

> 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" )

(non)-Linear models: Estimates and predictions (lin-mod)
Spline effects in \texttt{glm}

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

(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

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

[1,] 0.000 0 0.000 0.000 0.000
[2,] 0.001 -0.025 0.076 -0.050 0.000
[3,] 0.001 -0.050 0.150 -0.100 -0.148
[4,] 0.004 0.074 0.221 0.147 0.000
[5,] 0.011 0.228 0.349 0.175 0.003
\end{verbatim}
Spline effects in \texttt{glm}

\begin{verbatim}
> 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" )
\end{verbatim}
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 )

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

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

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>[1,]</td>
<td>1</td>
<td>0.00000000000</td>
<td>0.000000000</td>
<td>0.000000000</td>
</tr>
<tr>
<td>[2,]</td>
<td>1</td>
<td>0.00016666667</td>
<td>0.02527011</td>
<td>0.07581034</td>
</tr>
<tr>
<td>[3,]</td>
<td>1</td>
<td>0.00133333333</td>
<td>0.05003313</td>
<td>0.15009940</td>
</tr>
<tr>
<td>[4,]</td>
<td>1</td>
<td>0.00450000000</td>
<td>0.07378197</td>
<td>0.22134590</td>
</tr>
<tr>
<td>[5,]</td>
<td>1</td>
<td>0.01066666667</td>
<td>0.09600952</td>
<td>0.28802857</td>
</tr>
</tbody>
</table>
Adding a linear period effect

```r
> 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) )
```
The period effect

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

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

> pp <- 1945:1995
> Cp <- cbind( pp ) - 1970
> head( Cp )

pp
[1,] -25
[2,] -24
[3,] -23
[4,] -22
[5,] -21
[6,] -20

(non)-Linear models: Estimates and predictions (lin-mod)
Period effect

> 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 )
A quadratic period effect

```r
> 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
```

> pp <- 1945:1995
> Cq <- cbind( pp-1970, pp^2-1970^2 )
> head( Cq )

```
[,1] [,2]
[1,]  -25 -97875
[2,]  -24 -93984
[3,]  -23 -90091
[4,]  -22 -86196
[5,]  -21 -82299
[6,]  -20 -78400
```

> ci.exp( mspq, subset="P" )

```
exp(Est.) 2.5%  97.5%
P       2.1893078 1.4566021 3.2905821
I(P^2)   0.9998075 0.9997042 0.9999107
```

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

(non)-Linear models: Estimates and predictions (lin-mod)
A quadratic period effect

> matplot( pp, ci.exp( mspq, subset="P", ctr.mat=Cq ),
+       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 )

(non)-Linear models: Estimates and predictions (lin-mod)
A spline period effect

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

(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))1  1.711  1.526  1.918
Ns(P, knots = seq(1950, 1990, 10))2  2.190  2.028  2.364
Ns(P, knots = seq(1950, 1990, 10))3  3.222  2.835  3.661
Ns(P, knots = seq(1950, 1990, 10))4  2.299  2.149  2.459
```

(non)-Linear models: Estimates and predictions (lin-mod)
A spline period effect

> pp <- 1945:1995
> Cs <- Ns( pp , knots=seq(1950,1990,10))
> Cr <- Ns(rep(1970,length(pp)),knots=seq(1950,1990,10))
> head( Cs )

```
 1  2  3  4
[1,] 0 0.12677314 -0.38031941 0.25354628
[2,] 0 0.10141851 -0.30425553 0.20283702
[3,] 0 0.07606388 -0.22819165 0.15212777
[4,] 0 0.05070926 -0.15212777 0.10141851
[5,] 0 0.02535463 -0.07606388 0.05070926
[6,] 0 0.00000000 0.00000000 0.00000000
```

> head( Cr )

```
 1   2   3   4
[1,] 0.6666667 0.1125042 0.1624874 -0.1083249
[2,] 0.6666667 0.1125042 0.1624874 -0.1083249
[3,] 0.6666667 0.1125042 0.1624874 -0.1083249
[4,] 0.6666667 0.1125042 0.1624874 -0.1083249
[5,] 0.6666667 0.1125042 0.1624874 -0.1083249
[6,] 0.6666667 0.1125042 0.1624874 -0.1083249
```

\(\text{(non)-Linear models: Estimates and predictions (Lin-mod)}\)

107/ 327
Period effect

> matplot( pp, ci.exp( msps, subset="P", ctr.mat=Cs-Cr ),
+       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 )

(non)-Linear models: Estimates and predictions (lin-mod)
Period effect

```r
> par( mfrow=c(1,2) )
> Cap <- cbind( 1, Ns( aa ,knots=seq(15,65,10)),
+ Ns(rep(1970,length(aa)),knots=seq(1950,1990,10)) )
> matplot( aa, ci.exp( msps, ctr.mat=Cap )*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( msps, subset="P", ctr.mat=Cs-Cr ),
+ 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 )
```
Age and period effect

(non)-Linear models: Estimates and predictions (lin-mod)
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 reference value.
- 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 prediction points and the reference point.
Recap of Monday — rates

- Rate, intensity: \( \lambda(t) = P\{\text{event in } (t, t+h)| \text{ alive at } t\} / h \)
- Observe empirical rates \((d, y)\) — possibly many per person.
- \( \ell_{\text{FU}} = d \log(\lambda) - \lambda y \), obs: \((d, y)\), rate par: \(\lambda\)
- \( \ell_{\text{Poisson}} = d \log(\lambda y) - \lambda y \), obs: \(d\), mean par: \(\mu = \lambda y\)
- \( \ell_{\text{Poisson}} - \ell_{\text{FU}} = d \log(y) \) does not involve \(\lambda\) — use either to find m.l.e. of \(\lambda\)
- Poisson model is for \(\log(\mu) = \log(\lambda y) = \log(\lambda) + \log(y)\) hence offset=\(\log(Y)\)
- Once rates are known, we can construct survival curves and derivatives of that.
Recap Monday — models

- Empirical rate \((d_t, y_t)\) relates to a **time** \(t\)
- Many for the same person — different times
- Not independent, but likelihood is a product
- One parameter per interval \(\Rightarrow\) exchangeable times
- Use scaling of \(t\): \(\Rightarrow\) smooth continuous effects of time
- ...technically complicated:
  - Construct \(CA \leftarrow Ns(a.pt, knots=a.kn)\)
  - \(ci.exp(\ model, \ ctr.mat=CA )\)
  - \(RR\) by period: \(CP \leftarrow Ns(p.pt, knots=p.kn)\)
    and: \(CR \leftarrow Ns(rep(p.ref, nrow(CP)), knots=p.kn)\)
  - \(ci.exp(\ model, \ ctr.mat=CP-CR)\)
  - ...actually: \(CP \leftarrow Ns(p.pt, knots=p.kn, ref=p.ref)\)

(non)-Linear models: Estimates and predictions (lin-mod)
Follow-up data

Statistical Analysis in the Lexis Diagram:
Age-Period-Cohort models
May 2016
Max Planck Institut for Demographic Research, Rostock
http://BendixCarstensen/APC/MPIDR-2016

FU-rep-Lexis
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:
  - Along age
  - Along calendar time
- Multiple timescales.
Representation of follow-up data

In a cohort study we have records of: 
**Events** and **Risk time**.

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

- Date of entry — date variable.
- Date of exit — date variable
- Status at exit — indicator-variable (0/1)

Specific for each *type* of outcome.
Aim of dividing time into bands:

Put $D$ — events

$Y$ — risk time

in intervals on the timescale:

**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?
Cohort with 3 persons:

<table>
<thead>
<tr>
<th>Id</th>
<th>Bdate</th>
<th>Entry</th>
<th>Exit</th>
<th>St</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>14/07/1952</td>
<td>04/08/1965</td>
<td>27/06/1997</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>01/04/1954</td>
<td>08/09/1972</td>
<td>23/05/1995</td>
<td>0</td>
</tr>
</tbody>
</table>

- Define strata: 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

<table>
<thead>
<tr>
<th></th>
<th>subj. 1</th>
<th>subj. 2</th>
<th>subj. 3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age at Entry:</strong></td>
<td>13.06</td>
<td>18.44</td>
<td>4.54</td>
</tr>
<tr>
<td><strong>Age at Exit:</strong></td>
<td>44.95</td>
<td>41.14</td>
<td>11.12</td>
</tr>
<tr>
<td><strong>Status at exit:</strong></td>
<td>Dead</td>
<td>Alive</td>
<td>Dead</td>
</tr>
<tr>
<td><strong>Y</strong></td>
<td>31.89</td>
<td>22.70</td>
<td>6.58</td>
</tr>
<tr>
<td><strong>D</strong></td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Follow-up data (FU-rep-Lexis)
<table>
<thead>
<tr>
<th>Age</th>
<th>subj. 1</th>
<th>subj. 2</th>
<th>subj. 3</th>
<th>( \sum )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Y</td>
<td>D</td>
<td>Y</td>
<td>D</td>
</tr>
<tr>
<td>0–</td>
<td>0.00</td>
<td>0</td>
<td>0.00</td>
<td>0</td>
</tr>
<tr>
<td>10–</td>
<td>6.94</td>
<td>0</td>
<td>1.56</td>
<td>0</td>
</tr>
<tr>
<td>20–</td>
<td>10.00</td>
<td>0</td>
<td>10.00</td>
<td>0</td>
</tr>
<tr>
<td>30–</td>
<td>10.00</td>
<td>0</td>
<td>10.00</td>
<td>0</td>
</tr>
<tr>
<td>40–</td>
<td>4.95</td>
<td>1</td>
<td>1.14</td>
<td>0</td>
</tr>
<tr>
<td>( \sum )</td>
<td>31.89</td>
<td>1</td>
<td>22.70</td>
<td>0</td>
</tr>
</tbody>
</table>
Splitting the follow-up

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

- but what if we want to keep track of calendar time too?

Follow-up data (FU-rep-Lexis)
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.
Representation of follow-up on several timescales

- The time followed is the same on all timescales.
- Only use 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.
Follow-up data in Epi: Lexis objects

A follow-up study:

```r
> round( th, 2 )

   id sex birthdat contrast injecdat volume exitdat exitstat
1  1  1   2.00     1 1916.61   1  1938.79    22  1976.79    1
2 640  2 1896.23   1  1945.77   2  1964.37    20  1964.37    1
3 3425  1 1886.97   2  1955.18   0  1956.59    19  1956.59    1
```

Timescales of interest:

- Age
- Calendar time
- Time since injection
Definition of *Lexis* object

```r
> thL <- Lexis( entry = list( age=injecdat-birthdat,
+ per=injecdat,
+ tfi=0 ),
+ exit = list( per=exitdat ),
+ exit.status = (exitstat==1)*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.
The looks of a **Lexis** object

```r
> round(thL[,c(1:8,14,15)], 2)

<table>
<thead>
<tr>
<th>id</th>
<th>age</th>
<th>per</th>
<th>tfi</th>
<th>lex.dur</th>
<th>lex.Cst</th>
<th>lex.Xst</th>
<th>lex.id</th>
<th>id</th>
<th>exitdat</th>
<th>exitstat</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>22.18</td>
<td>1938.79</td>
<td>0</td>
<td>38.00</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1976.79</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>49.55</td>
<td>1945.77</td>
<td>0</td>
<td>18.60</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>640</td>
<td>1964.37</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>68.21</td>
<td>1955.18</td>
<td>0</td>
<td>1.40</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>3425</td>
<td>1956.59</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>20.80</td>
<td>1957.61</td>
<td>0</td>
<td>34.52</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>4017</td>
<td>1992.14</td>
<td>2</td>
</tr>
</tbody>
</table>
```

Follow-up data (FU-rep-Lexis)
> plot( thL, lwd=3 )

Follow-up data (FU-rep-Lexis)
Follow-up data (FU-rep-Lexis)
Follow-up data (FU-rep-Lexis) 128/ 327

```r
> 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**

```r
> spl1 <- splitLexis( thL, "age", breaks=seq(0,100,20) )
> round( spl1, 2 )

<table>
<thead>
<tr>
<th>lex.id</th>
<th>age</th>
<th>per</th>
<th>tfi</th>
<th>lex.dur</th>
<th>lex.Cst</th>
<th>lex.Xst</th>
<th>id</th>
<th>sex</th>
<th>birthdat</th>
<th>contrast</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>22.18</td>
<td>1938.79</td>
<td>0.00</td>
<td>17.82</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>1916.61</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>40.00</td>
<td>1956.61</td>
<td>17.82</td>
<td>20.00</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>1916.61</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>60.00</td>
<td>1976.61</td>
<td>37.82</td>
<td>0.18</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1916.61</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>49.55</td>
<td>1945.77</td>
<td>0.00</td>
<td>10.45</td>
<td>0</td>
<td>0</td>
<td>640</td>
<td>2</td>
<td>1896.23</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>60.00</td>
<td>1956.23</td>
<td>10.45</td>
<td>8.14</td>
<td>0</td>
<td>1</td>
<td>640</td>
<td>2</td>
<td>1896.23</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>68.21</td>
<td>1955.18</td>
<td>0.00</td>
<td>1.40</td>
<td>0</td>
<td>1</td>
<td>3425</td>
<td>1</td>
<td>1886.97</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>20.80</td>
<td>1957.61</td>
<td>0.00</td>
<td>19.20</td>
<td>0</td>
<td>0</td>
<td>4017</td>
<td>2</td>
<td>1936.81</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>40.00</td>
<td>1976.81</td>
<td>19.20</td>
<td>15.33</td>
<td>0</td>
<td>0</td>
<td>4017</td>
<td>2</td>
<td>1936.81</td>
<td>2</td>
</tr>
</tbody>
</table>
```

Follow-up data (FU-rep-Lexis)
Split on a second timescale

```r
> # Split further on tfi:
> spl2 <- splitLexis( spl1, "tfi", breaks=c(0,1,5,20,100) )
> round( spl2, 2 )

<table>
<thead>
<tr>
<th>lex.id</th>
<th>age</th>
<th>per</th>
<th>tfi</th>
<th>lex.dur</th>
<th>lex.Cst</th>
<th>lex.Xst</th>
<th>id</th>
<th>sex</th>
<th>birthdat</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>22.18</td>
<td>0.00</td>
<td>1.00</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>1916.61</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>23.18</td>
<td>1.00</td>
<td>4.00</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>1916.61</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>27.18</td>
<td>5.00</td>
<td>12.82</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>1916.61</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>40.00</td>
<td>17.82</td>
<td>2.18</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>1916.61</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>42.18</td>
<td>20.00</td>
<td>17.82</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>1916.61</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>60.00</td>
<td>37.82</td>
<td>0.18</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1916.61</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>49.55</td>
<td>0.00</td>
<td>1.00</td>
<td>0</td>
<td>0</td>
<td>640</td>
<td>2</td>
<td>1896.23</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>50.55</td>
<td>1.00</td>
<td>4.00</td>
<td>0</td>
<td>0</td>
<td>640</td>
<td>2</td>
<td>1896.23</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>54.55</td>
<td>5.00</td>
<td>5.45</td>
<td>0</td>
<td>0</td>
<td>640</td>
<td>2</td>
<td>1896.23</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>60.00</td>
<td>10.45</td>
<td>8.14</td>
<td>0</td>
<td>1</td>
<td>640</td>
<td>2</td>
<td>1896.23</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>68.21</td>
<td>0.00</td>
<td>1.00</td>
<td>0</td>
<td>0</td>
<td>3425</td>
<td>1</td>
<td>1886.97</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>69.21</td>
<td>0.40</td>
<td>0.40</td>
<td>0</td>
<td>1</td>
<td>3425</td>
<td>1</td>
<td>1886.97</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>20.80</td>
<td>0.00</td>
<td>1.00</td>
<td>0</td>
<td>0</td>
<td>4017</td>
<td>2</td>
<td>1936.81</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>21.80</td>
<td>1.00</td>
<td>4.00</td>
<td>0</td>
<td>0</td>
<td>4017</td>
<td>2</td>
<td>1936.81</td>
<td></td>
</tr>
</tbody>
</table>
```

Follow-up data (FU-rep-Lexis)

The Poisson likelihood for time-split data

One record per person-interval \((i, t)\):

\[
D\log(\lambda) - \lambda Y = \sum_{i, t}(d_{it}\log(\lambda) - \lambda y_{it})
\]

Assuming that the death indicator \((d_i \in \{0, 1\})\) is Poisson, with log-offset \(y_i\) will give the same result.

The model assume that rates are constant.

But the split data allows relaxing this to models that assume different rates for different \((d_{it}, y_{it})\).

Where are the \((d_{it}, y_{it})\) in the split data?
The Poisson likelihood for time-split data

If \( d \sim \text{Poisson}(\lambda y) \), i.e. with mean \((\lambda y)\) then the log-likelihood is

\[
d \log(\lambda y) - \lambda y
\]

If we assume a multiplicative model for the rates, i.e. an additive model for the log-rates, we can use a Poisson model which is multiplicative in the mean, \( \mu \), i.e. linear in \( \log(\mu) \):

\[
\log(\mu) = \log(\lambda y) = \log(\lambda) + \log(y)
\]

Regression model must include \( \log(y) \) as covariate with coefficient fixed to 1 — an offset-variable.
plot( spl2, c(1,3), col="black", lwd=2 )

Follow-up data (FU-rep-Lexis)
Where is \((d_{it}, y_{it})\) in the split data?

```r
> round( spl2, 2 )
  lex.id age per tfi lex.dur lex.Cst lex.Xst id sex birthdat
1   1  22.18 1938.79  0.00  1.00  0  0  1  2  1916.61
2   1  23.18 1939.79  1.00  4.00  0  0  1  2  1916.61
3   1  27.18 1943.79  5.00 12.82  0  0  1  2  1916.61
4   1  40.00 1956.61 17.82  2.18  0  0  1  2  1916.61
5   1  42.18 1958.79 20.00 17.82  0  0  1  2  1916.61
6   1  60.00 1976.61 37.82  0.18  0  1  1  2  1916.61
7   2  49.55 1945.77  0.00  1.00  0  0  640  2  1896.23
8   2  50.55 1946.77  1.00  4.00  0  0  640  2  1896.23
9   2  54.55 1950.77  5.00  5.45  0  0  640  2  1896.23
10  2  60.00 1956.23 10.45  8.14  0  1  640  2  1896.23
11  3  68.21 1955.18  0.00  1.00  0  0  3425  1  1886.97
12  3  69.21 1956.18  1.00  0.40  0  1  3425  1  1886.97
13  4  20.80 1957.61  0.00  1.00  0  0  4017  2  1936.81
14  4  21.80 1958.61  1.00  4.00  0  0  4017  2  1936.81
15  4  25.80 1962.61  5.00 14.20  0  0  4017  2  1936.81
16  4  40.00 1976.81 19.20  0.80  0  0  4017  2  1936.81
```

Follow-up data (FU-rep-Lexis) 134/ 327
Analysis of results

- $d_i$ — events in the variable: lex.Xst.
- $y_i$ — risk time: lex.dur (duration). Enters in the model via $\log(y)$ as offset.
- 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.

Follow-up data (FU-rep-Lexis)
Poisson model for split data

- Each interval contribute $\lambda Y$ to the log-likelihood.
- All intervals with the same set of covariate values (age, exposure, ...) have the same $\lambda$.
- The log-likelihood contribution from these is $\lambda \sum Y$ — the same as from aggregated data.
- The event intervals contribute each $D \log \lambda$.
- The log-likelihood contribution from those with the same lambda is $\sum D \log \lambda$ — the same as from aggregated data.
- The log-likelihood is the same for split data and aggregated data — no need to tabulate first.
Models for tabulated data

Statistical Analysis in the
Lexis Diagram:
Age-Period-Cohort models
May 2016
Max Planck Institut for Demographic Research, Rostock
http://BendixCarstensen/APC/MPIDR-2016
Conceptual set-up

Follow-up of the entire (male) population from 1943–2006 w.r.t. occurrence of testis cancer:

- Split follow-up time for all about 4 mio. men in 1-year classes by age and calendar time ($y$).
- Allocate testis cancer event ($d = 0, 1$) to each.
- Analyse all 200,000,000 records by a Poisson model.
Realistic set-up

- Tabulate the follow-up time and events by age and period.
- 100 age-classes.
- 65 periods (single calendar years).
- 6500 aggregate records of \((D, Y)\).
- Analyze by a Poisson model.
Practical set-up

- Tabulate only events (as obtained from the cancer registry) by age and period.
- 100 age-classes.
- 65 periods (single calendar years).
- 6500 aggregate records of $D$.
- Estimate the population follow-up based on census data from Statistics Denmark.
  Or get it from the human mortality database.
- Analyse by Poisson model.
Lexis diagram

Disease registers record events.

Official statistics collect population data.

1 Named after the German statistician and economist William Lexis (1837–1914), who devised this diagram in the book “Einleitung in die Theorie der Bevölkerungsstatistik” (Karl J. Trübner, Strassburg, 1875).
Lexis diagram

Registration of:
cases ($D$)
risk time,
person-years ($Y$)
in subsets of the Lexis diagram.

Models for tabulated data (tab-mod)
Registration of:
cases \( (D) \)
risk time,
person-years \( (Y) \)
in subsets of the Lexis diagram.
Rates available in each subset.
Register data

Classification of cases \( (D_{ap}) \) by age at diagnosis and date of diagnosis, and population \( (Y_{ap}) \) by age at risk and date at risk, in compartments of the Lexis diagram, e.g.:

<table>
<thead>
<tr>
<th>Age</th>
<th>1943</th>
<th>1948</th>
<th>1953</th>
<th>1958</th>
<th>Person-years</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>1</td>
<td>773812 744217 794123 972853</td>
</tr>
<tr>
<td>20</td>
<td>7</td>
<td>7</td>
<td>17</td>
<td>8</td>
<td>813022 744706 721810 770859</td>
</tr>
<tr>
<td>25</td>
<td>28</td>
<td>23</td>
<td>26</td>
<td>35</td>
<td>790501 781827 722968 698612</td>
</tr>
<tr>
<td>30</td>
<td>28</td>
<td>43</td>
<td>49</td>
<td>51</td>
<td>799293 774542 769298 711596</td>
</tr>
<tr>
<td>35</td>
<td>36</td>
<td>42</td>
<td>39</td>
<td>44</td>
<td>769356 782893 760213 760452</td>
</tr>
<tr>
<td>40</td>
<td>24</td>
<td>32</td>
<td>46</td>
<td>53</td>
<td>694073 754322 768471 749912</td>
</tr>
</tbody>
</table>
Reshape data to analysis form:

<table>
<thead>
<tr>
<th>A</th>
<th>P</th>
<th>D</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>15</td>
<td>1943</td>
<td>2 773812</td>
</tr>
<tr>
<td>2</td>
<td>20</td>
<td>1943</td>
<td>7 813022</td>
</tr>
<tr>
<td>3</td>
<td>25</td>
<td>1943</td>
<td>28 790501</td>
</tr>
<tr>
<td>4</td>
<td>30</td>
<td>1943</td>
<td>28 799293</td>
</tr>
<tr>
<td>5</td>
<td>35</td>
<td>1943</td>
<td>36 769356</td>
</tr>
<tr>
<td>6</td>
<td>40</td>
<td>1943</td>
<td>24 694073</td>
</tr>
<tr>
<td>1</td>
<td>15</td>
<td>1948</td>
<td>3 744217</td>
</tr>
<tr>
<td>2</td>
<td>20</td>
<td>1948</td>
<td>7 744706</td>
</tr>
<tr>
<td>3</td>
<td>25</td>
<td>1948</td>
<td>23 781827</td>
</tr>
<tr>
<td>4</td>
<td>30</td>
<td>1948</td>
<td>43 774542</td>
</tr>
<tr>
<td>5</td>
<td>35</td>
<td>1948</td>
<td>42 782893</td>
</tr>
<tr>
<td>6</td>
<td>40</td>
<td>1948</td>
<td>32 754322</td>
</tr>
<tr>
<td>1</td>
<td>15</td>
<td>1953</td>
<td>4 794123</td>
</tr>
<tr>
<td>2</td>
<td>20</td>
<td>1953</td>
<td>17 721810</td>
</tr>
<tr>
<td>3</td>
<td>25</td>
<td>1953</td>
<td>26 722968</td>
</tr>
<tr>
<td>4</td>
<td>30</td>
<td>1953</td>
<td>49 769298</td>
</tr>
<tr>
<td>5</td>
<td>35</td>
<td>1953</td>
<td>39 760213</td>
</tr>
<tr>
<td>6</td>
<td>40</td>
<td>1953</td>
<td>46 768471</td>
</tr>
<tr>
<td>1</td>
<td>15</td>
<td>1958</td>
<td>1 728553</td>
</tr>
<tr>
<td>2</td>
<td>20</td>
<td>1958</td>
<td>2 779852</td>
</tr>
</tbody>
</table>

Models for tabulated data (tab-mod) 144/327
Tabulated data

Once data are in tabular form, models are restricted:

- Rates must be assumed constant in each cell of the table / subset of the Lexis diagram.
- With large cells it is customary to put a separate parameter on each cell or on each level of classifying factors.
- Output from the model will be rates and rate-ratios.
- Since we use multiplicative Poisson, usually the log rates and the log-RR are reported.
Simple model for the testis cancer rates:

```r
> m0 <- glm( D ~ factor(A) + factor(P) + offset( log(Y/10^5) ),
+   family=poisson, data=ts )
> summary( m0 )

Call:
glm(formula = D ~ factor(A) + factor(P) + offset(log(Y/10^5)),
    family = poisson, data = ts)

Deviance Residuals:
     Min       1Q   Median       3Q      Max
-1.5991  -0.6974   0.1284   0.6671   1.8904

Coefficients:
                          Estimate Std. Error    z value  Pr(>|z|)
(Intercept)                -1.4758     0.3267     -4.517    6.26e-06
factor(A)20                1.4539     0.3545      4.101    4.11e-05
factor(A)25                2.5321     0.3301      7.671    1.71e-14
factor(A)30                2.9327     0.3254      9.013     < 2e-16
factor(A)35                2.8613     0.3259      8.779     < 2e-16
factor(A)40                2.8521     0.3263      8.741     < 2e-16
factor(P)1948              0.1753     0.1211      1.447     0.14778
factor(P)1953              0.3822     0.1163      3.286     0.00102
factor(P)1958              0.4659     0.1150      4.052     5.07e-05

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 333.866 on 23 degrees of freedom
    Residual deviance: 17.532 on 15 degrees of freedom
    AIC: 149.53

Models for tabulated data (tab-mod) 146/ 327
```
ci.lin() from the Epi package extracts coefficients and computes confidence intervals:

```r
> round( ci.lin( m0 ), 3 )

<table>
<thead>
<tr>
<th>Estimate</th>
<th>StdErr</th>
<th>z</th>
<th>P</th>
<th>2.5%</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-1.476</td>
<td>0.327</td>
<td>-4.517</td>
<td>0.000</td>
<td>-2.116</td>
</tr>
<tr>
<td>factor(A)20</td>
<td>1.454</td>
<td>0.354</td>
<td>4.101</td>
<td>0.000</td>
<td>0.759</td>
</tr>
<tr>
<td>factor(A)25</td>
<td>2.532</td>
<td>0.330</td>
<td>7.671</td>
<td>0.000</td>
<td>1.885</td>
</tr>
<tr>
<td>factor(A)30</td>
<td>2.933</td>
<td>0.325</td>
<td>9.013</td>
<td>0.000</td>
<td>2.295</td>
</tr>
<tr>
<td>factor(A)35</td>
<td>2.861</td>
<td>0.326</td>
<td>8.779</td>
<td>0.000</td>
<td>2.223</td>
</tr>
<tr>
<td>factor(A)40</td>
<td>2.852</td>
<td>0.326</td>
<td>8.741</td>
<td>0.000</td>
<td>2.213</td>
</tr>
<tr>
<td>factor(P)1948</td>
<td>0.175</td>
<td>0.121</td>
<td>1.447</td>
<td>0.148</td>
<td>-0.062</td>
</tr>
<tr>
<td>factor(P)1953</td>
<td>0.382</td>
<td>0.116</td>
<td>3.286</td>
<td>0.001</td>
<td>0.154</td>
</tr>
<tr>
<td>factor(P)1958</td>
<td>0.466</td>
<td>0.115</td>
<td>4.052</td>
<td>0.000</td>
<td>0.241</td>
</tr>
</tbody>
</table>
```
Subsets of parameter estimates accessed via a character string that is grep'd to the names.

```r
> round( ci.lin( m0, subset="P" ), 3 )

<table>
<thead>
<tr>
<th>Factor</th>
<th>Estimate</th>
<th>StdErr</th>
<th>z</th>
<th>P</th>
<th>2.5%</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>factor(P)1948</td>
<td>0.175</td>
<td>0.121</td>
<td>1.447</td>
<td>0.148</td>
<td>-0.062</td>
<td>0.413</td>
</tr>
<tr>
<td>factor(P)1953</td>
<td>0.382</td>
<td>0.116</td>
<td>3.286</td>
<td>0.001</td>
<td>0.154</td>
<td>0.610</td>
</tr>
<tr>
<td>factor(P)1958</td>
<td>0.466</td>
<td>0.115</td>
<td>4.052</td>
<td>0.000</td>
<td>0.241</td>
<td>0.691</td>
</tr>
</tbody>
</table>
```

Models for tabulated data (tab-mod)
Rates / rate-ratios are computed on the fly by \(\text{Exp=TRUE}\):

```r
> round( ci.lin( m0, subset="P", Exp=TRUE ), 3 )

<table>
<thead>
<tr>
<th>Estimate</th>
<th>StdErr</th>
<th>z</th>
<th>P</th>
<th>exp(Est.)</th>
<th>2.5%</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>factor(P)1948</td>
<td>0.175</td>
<td>0.121</td>
<td>1.447</td>
<td>0.148</td>
<td>1.192</td>
<td>0.940</td>
</tr>
<tr>
<td>factor(P)1953</td>
<td>0.382</td>
<td>0.116</td>
<td>3.286</td>
<td>0.001</td>
<td>1.466</td>
<td>1.167</td>
</tr>
<tr>
<td>factor(P)1958</td>
<td>0.466</td>
<td>0.115</td>
<td>4.052</td>
<td>0.000</td>
<td>1.593</td>
<td>1.272</td>
</tr>
</tbody>
</table>
```

Models for tabulated data (tab-mod)
Linear combinations of the parameters can be computed using the `ctr.mat` option:

```r
> CM <- rbind( c( 0,-1, 0),
+             c( 1,-1, 0),
+             c( 0, 0, 0),
+             c( 0,-1, 1) )
> round( ci.lin( m0, subset="P", ctr.mat=CM, Exp=TRUE ), 3 )
```

```
  Estimate StdErr  z   P  exp(Est.)  2.5%  97.5%
[1,]  -0.382  0.116 -3.286 0.001    0.682  0.543  0.857
[2,]  -0.207  0.110 -1.874 0.061    0.813  0.655  1.010
[3,]   0.000  0.000   NaN   NaN     1.000  1.000  1.000
[4,]   0.084  0.104  0.808 0.419    1.087  0.887  1.332
```
Age-Period and Age-Cohort models

Statistical Analysis in the Lexis Diagram:

Age-Period-Cohort models
May 2016
Max Planck Institut for Demographic Research, Rostock
http://BendixCarstensen/APC/MPIDR-2016
Register data - rates

Rates in “tiles” of the Lexis diagram:

$$\lambda(a, p) = \frac{D_{ap}}{Y_{ap}}$$

Descriptive epidemiology based on disease registers:
How do the rates vary across by age and time:

- Age-specific rates for a given period.
- Age-standardized rates as a function of calendar time. (Weighted averages of the age-specific rates).
Synthetic cohorts

Events and risk time in cells along the diagonals are among persons with roughly same date of birth.

Successively overlapping 10-year periods.
**Lexis diagram: data**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>471.0</td>
<td>512.8</td>
<td>571.1</td>
<td>622.5</td>
<td>680.8</td>
<td>748.2</td>
</tr>
<tr>
<td>25</td>
<td>569.4</td>
<td>600.3</td>
<td>639.3</td>
<td>715.4</td>
<td>732.7</td>
<td>718.3</td>
</tr>
<tr>
<td>35</td>
<td>755.3</td>
<td>731.5</td>
<td>753.5</td>
<td>738.1</td>
<td>746.4</td>
<td>698.2</td>
</tr>
<tr>
<td>45</td>
<td>758.7</td>
<td>754.3</td>
<td>768.5</td>
<td>749.9</td>
<td>756.5</td>
<td>709.8</td>
</tr>
<tr>
<td>55</td>
<td>769.4</td>
<td>768.6</td>
<td>790.5</td>
<td>791.5</td>
<td>811.6</td>
<td>702.3</td>
</tr>
<tr>
<td>65</td>
<td>799.3</td>
<td>774.6</td>
<td>769.3</td>
<td>711.6</td>
<td>700.1</td>
<td>679.9</td>
</tr>
<tr>
<td>75</td>
<td>790.5</td>
<td>761.8</td>
<td>723.0</td>
<td>698.6</td>
<td>764.8</td>
<td>696.2</td>
</tr>
<tr>
<td>85</td>
<td>813.0</td>
<td>744.7</td>
<td>721.8</td>
<td>770.9</td>
<td>760.3</td>
<td>803.8</td>
</tr>
<tr>
<td>95</td>
<td>773.8</td>
<td>744.2</td>
<td>794.1</td>
<td>972.9</td>
<td>1051.5</td>
<td>961.0</td>
</tr>
</tbody>
</table>

**Testis cancer cases in Denmark.**

**Male person-years in Denmark.**

Age-Period and Age-Cohort models (AP-AC)
### Data matrix: Testis cancer cases

#### Number of cases

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>15–19</td>
<td>7</td>
<td>13</td>
<td>13</td>
<td>15</td>
<td>33</td>
<td>35</td>
<td>37</td>
<td>49</td>
<td>51</td>
</tr>
<tr>
<td>20–24</td>
<td>31</td>
<td>46</td>
<td>49</td>
<td>55</td>
<td>85</td>
<td>110</td>
<td>140</td>
<td>151</td>
<td>150</td>
</tr>
<tr>
<td>25–29</td>
<td>62</td>
<td>63</td>
<td>82</td>
<td>87</td>
<td>103</td>
<td>153</td>
<td>201</td>
<td>214</td>
<td>268</td>
</tr>
<tr>
<td>30–34</td>
<td>66</td>
<td>82</td>
<td>88</td>
<td>103</td>
<td>124</td>
<td>164</td>
<td>207</td>
<td>209</td>
<td>258</td>
</tr>
<tr>
<td>35–39</td>
<td>56</td>
<td>56</td>
<td>67</td>
<td>99</td>
<td>124</td>
<td>142</td>
<td>152</td>
<td>188</td>
<td>209</td>
</tr>
<tr>
<td>40–44</td>
<td>47</td>
<td>65</td>
<td>64</td>
<td>67</td>
<td>85</td>
<td>103</td>
<td>119</td>
<td>121</td>
<td>155</td>
</tr>
<tr>
<td>45–49</td>
<td>30</td>
<td>37</td>
<td>54</td>
<td>45</td>
<td>64</td>
<td>63</td>
<td>66</td>
<td>92</td>
<td>86</td>
</tr>
<tr>
<td>50–54</td>
<td>28</td>
<td>22</td>
<td>27</td>
<td>46</td>
<td>36</td>
<td>50</td>
<td>49</td>
<td>61</td>
<td>64</td>
</tr>
<tr>
<td>55–59</td>
<td>14</td>
<td>16</td>
<td>25</td>
<td>26</td>
<td>29</td>
<td>28</td>
<td>43</td>
<td>42</td>
<td>34</td>
</tr>
</tbody>
</table>

#### Date of diagnosis (year – 1900)

Age-Period and Age-Cohort models (AP–AC)
### Data matrix: Male risk time

1000 person-years

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>15–19</td>
<td>744.2</td>
<td>794.1</td>
<td>972.9</td>
<td>1051.5</td>
<td>961.0</td>
<td>952.5</td>
<td>1011.1</td>
<td>1005.0</td>
<td>929.8</td>
</tr>
<tr>
<td>20–24</td>
<td>744.7</td>
<td>721.8</td>
<td>770.9</td>
<td>960.3</td>
<td>1053.8</td>
<td>967.5</td>
<td>953.0</td>
<td>1019.7</td>
<td>1017.3</td>
</tr>
<tr>
<td>25–29</td>
<td>781.8</td>
<td>723.0</td>
<td>698.6</td>
<td>764.8</td>
<td>962.7</td>
<td>1056.1</td>
<td>960.9</td>
<td>956.2</td>
<td>1031.6</td>
</tr>
<tr>
<td>30–34</td>
<td>774.5</td>
<td>769.3</td>
<td>711.6</td>
<td>700.1</td>
<td>769.9</td>
<td>960.4</td>
<td>1045.3</td>
<td>955.0</td>
<td>957.1</td>
</tr>
<tr>
<td>35–39</td>
<td>782.9</td>
<td>760.2</td>
<td>760.5</td>
<td>711.6</td>
<td>702.3</td>
<td>767.5</td>
<td>951.9</td>
<td>1035.7</td>
<td>948.6</td>
</tr>
<tr>
<td>40–44</td>
<td>754.3</td>
<td>768.5</td>
<td>749.9</td>
<td>756.5</td>
<td>709.8</td>
<td>696.5</td>
<td>757.8</td>
<td>940.3</td>
<td>1023.7</td>
</tr>
<tr>
<td>45–49</td>
<td>676.7</td>
<td>737.9</td>
<td>753.5</td>
<td>738.1</td>
<td>746.4</td>
<td>698.2</td>
<td>682.4</td>
<td>743.1</td>
<td>923.4</td>
</tr>
<tr>
<td>50–54</td>
<td>600.3</td>
<td>653.9</td>
<td>715.4</td>
<td>732.7</td>
<td>718.3</td>
<td>724.2</td>
<td>675.5</td>
<td>660.8</td>
<td>721.1</td>
</tr>
<tr>
<td>55–59</td>
<td>512.8</td>
<td>571.1</td>
<td>622.5</td>
<td>680.8</td>
<td>698.2</td>
<td>683.8</td>
<td>686.4</td>
<td>640.9</td>
<td>627.7</td>
</tr>
</tbody>
</table>

**Date of diagnosis (year − 1900)**

Age-Period and Age-Cohort models (AP-AC)
## Data matrix: Empirical rates

Rate per 1,000,000 person-years

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>15–19</td>
<td>9.4</td>
<td>16.4</td>
<td>13.4</td>
<td>14.3</td>
<td>34.3</td>
<td>36.7</td>
<td>36.6</td>
<td>48.8</td>
<td>54.8</td>
</tr>
<tr>
<td>20–24</td>
<td>41.6</td>
<td>63.7</td>
<td>63.6</td>
<td>57.3</td>
<td>80.7</td>
<td>113.7</td>
<td>146.9</td>
<td>148.1</td>
<td>147.4</td>
</tr>
<tr>
<td>25–29</td>
<td>79.3</td>
<td>87.1</td>
<td>117.4</td>
<td>113.8</td>
<td>107.0</td>
<td>144.9</td>
<td>209.2</td>
<td>223.8</td>
<td>259.8</td>
</tr>
<tr>
<td>30–34</td>
<td>85.2</td>
<td>106.6</td>
<td>123.7</td>
<td>147.1</td>
<td>161.1</td>
<td>170.8</td>
<td>198.0</td>
<td>218.8</td>
<td>269.6</td>
</tr>
<tr>
<td>35–39</td>
<td>71.5</td>
<td>73.7</td>
<td>88.1</td>
<td>139.1</td>
<td>176.6</td>
<td>185.0</td>
<td>159.7</td>
<td>181.5</td>
<td>220.3</td>
</tr>
<tr>
<td>40–44</td>
<td>62.3</td>
<td>84.6</td>
<td>85.3</td>
<td>88.6</td>
<td>119.8</td>
<td>147.9</td>
<td>157.0</td>
<td>128.7</td>
<td>151.4</td>
</tr>
<tr>
<td>45–49</td>
<td>44.3</td>
<td>50.1</td>
<td>71.7</td>
<td>61.0</td>
<td>85.7</td>
<td>90.2</td>
<td>96.7</td>
<td>123.8</td>
<td>93.1</td>
</tr>
<tr>
<td>50–54</td>
<td>46.6</td>
<td>33.6</td>
<td>37.7</td>
<td>62.8</td>
<td>50.1</td>
<td>69.0</td>
<td>72.5</td>
<td>92.3</td>
<td>88.7</td>
</tr>
<tr>
<td>55–59</td>
<td>27.3</td>
<td>28.0</td>
<td>40.2</td>
<td>38.2</td>
<td>41.5</td>
<td>40.9</td>
<td>62.6</td>
<td>65.5</td>
<td>54.2</td>
</tr>
</tbody>
</table>
The classical plots

Given a table of rates classified by age and period, we can do 4 “classical” plots:

- Rates versus age at diagnosis (period):
  — rates in the same ageclass connected.
- Rates versus age at diagnosis:
  — rates in the same birth-cohort connected.
- Rates versus date of diagnosis:
  — rates in the same ageclass connected.
- Rates versus date of date of birth:
  — rates in the same ageclass connected.

These plots can be produced by the R-function `rateplot`.
```r
> library( Epi )
> load( file="../data/testisDK.Rda" )
> head( testisDK )

    A   P   D   Y
1 17.5 1950.5 7 744.2172
2 22.5 1950.5 31 744.7055
3 27.5 1950.5 62 781.8272
4 32.5 1950.5 66 774.5415
5 37.5 1950.5 56 782.8932
6 42.5 1950.5 47 754.3220

> xtabs( D ~ A + P, data = testisDK )

    P
 A    1950.5 1955.5 1960.5 1965.5 1970.5 1975.5 1980.5 1985.5 1990.5
17.5    7     13    13    15     33     35     37     49     51
22.5    31    46     49    55     85    110    140    151    150
27.5    62    63     82    87    103    153    201    214    268
32.5    66    82     88   103    124    164    207    209    258
37.5    56    56     67    99    124    142    152    188    209
42.5    47    65     64    85   103    119    121    155
47.5    30    37     54    45     64     63     66     92     86
52.5    28    22     27    46    36    50     49     61     64
57.5    14    16     25    26     29     28     43     42     34

> trate <- xtabs( D ~ A + P, data = testisDK ) / 
       xtabs( Y ~ A + P, data = testisDK ) * 100
> par( mfrow=c(2,2) )
> rateplot( trate, col=rainbow(15), lwd=3, ann=TRUE )
```

Age-Period and Age-Cohort models (AP-AC)
wh = c("ap","ac","pa","ca")
for( i in 1:4 ) {
  pdf( paste("./graph/AP-AC-testisRate",i,".pdf",sep=""), height=6, width=6 )
  par( mar=c(3,3,1,1, mgp=c(3,1,0)/1.6, bty="n", las=1 ))
  rateplot( trate, wh[i], col=rainbow(15), lwd=3, ann=TRUE, a.lim=c(15,65) )
  dev.off()
}

Age-Period and Age-Cohort models (AP-AC)
Age-Period and Age-Cohort models (AP-AC)
Age-Period and Age-Cohort models (AP-AC)
Age-Period and Age-Cohort models (AP-AC)
Age-period model

Rates are proportional between periods:

\[ \lambda(a, p) = a_a \times b_p \quad \text{or} \quad \log[\lambda(a, p)] = \alpha_a + \beta_p \]

Choose \( p_0 \) as reference period, where \( \beta_{p_0} = 0 \)

\[ \log[\lambda(a, p_0)] = \alpha_a + \beta_{p_0} = \alpha_a \]
Fitting the model in R

Reference period is the 5th period (1970.5 ∼ 1968–72):

```r
> ap <- glm( D ~ factor(A) - 1 + relevel(factor(P), 5) +
+ offset(log(Y)),
+ family=poisson )
> summary( ap )

Call:
glm(formula = D ~ factor(A) - 1 + relevel(factor(P), 5) + offset(log(Y)), family = poisson)

Deviance Residuals:
    Min      1Q  Median      3Q     Max
-3.0925 -0.8784  0.1148  0.9790  2.7653

Coefficients:
                               Estimate  Std. Error   z value  Pr(>|z|)
factor(A)17.5                   -3.56605   0.07249   -49.194  < 2e-16
factor(A)22.5                   -2.38447   0.04992   -47.766  < 2e-16
factor(A)27.5                   -1.94496   0.04583   -42.442  < 2e-16
factor(A)32.5                   -1.85214   0.04597   -40.294  < 2e-16
factor(A)37.5                   -1.99308   0.04770   -41.787  < 2e-16
factor(A)42.5                   -2.23017   0.05057   -44.104  < 2e-16
factor(A)47.5                   -2.58125   0.05631   -45.839  < 2e-16
relevel(factor(P), 5)1950.5    -0.59619   0.06634   -8.988  < 2e-16
relevel(factor(P), 5)1955.5    -0.43458   0.06299   -6.899  5.25e-12
relevel(factor(P), 5)1960.5    -0.27874   0.06000   -4.646  3.39e-06
relevel(factor(P), 5)1965.5    -0.16978   0.05751   -2.952   0.00316
relevel(factor(P), 5)1975.5     0.16139   0.05143   -3.138   0.00170
relevel(factor(P), 5)1980.5     0.30238   0.04954   -6.104  1.04e-09
relevel(factor(P), 5)1985.5     0.37654   0.04853   -7.758  8.62e-15
relevel(factor(P), 5)1990.5     0.47123   0.04745    9.931  < 2e-16

Null deviance: 89358.53 on 81 degrees of freedom
Residual deviance: 117.70 on 64 degrees of freedom

Age-Period and Age-Cohort models (AP-AC)

164/327
Graph of estimates with confidence intervals

Age-Period and Age-Cohort models (AP-AC)
Age-cohort model

Rates are proportional between cohorts:

\[ \lambda(a, c) = a_a \times c_c \quad \text{or} \quad \log[\lambda(a, p)] = \alpha_a + \gamma_c \]

Choose \( c_0 \) as reference cohort, where \( \gamma_{c_0} = 0 \)

\[ \log[\lambda(a, c_0)] = \alpha_a + \gamma_{c_0} = \alpha_a \]
Fit the model in R

Reference period is the 9th cohort (1933 ∼ 1928–38):

```r
> ac <- glm( D ~ factor( A ) - 1 + relevel( factor( C ), 9 ) +
+         offset( log( Y ) ),
+         family=poisson )
> summary( ac )

Call:
glm(formula = D ~ factor(A) - 1 + relevel(factor(C), 9) + offset(log(Y)), family = poisson)

Deviance Residuals:
Min 1Q Median 3Q Max
-1.92700 -0.72364 -0.02422 0.59623 1.87770

Coefficients:
                                 Estimate  Std. Error  z value  Pr(>|z|)
factor(A)17.5                 -4.07597   0.08360  -48.753 < 2e-16
factor(A)22.5                 -2.72942   0.05683  -48.031 < 2e-16
factor(A)27.5                 -2.15347   0.05066  -42.505 < 2e-16
factor(A)32.5                 -1.90118   0.04878  -38.976 < 2e-16
factor(A)37.5                 -1.89404   0.04934  -38.387 < 2e-16
factor(A)42.5                 -1.98846   0.05178  -38.399 < 2e-16
factor(A)47.5                 -2.23047   0.05775  -38.626 < 2e-16
factor(A)52.5                 -2.40391   0.06463  -37.194 < 2e-16
factor(A)57.5                 -2.60253   0.07577  -34.346 < 2e-16
factor(A)62.5                 -2.84312   0.08767  -32.231 < 2e-16
factor(A)67.5                 -3.11375   0.09990  -31.142 < 2e-16
factor(A)72.5                 -3.41057   0.11442  -29.722 < 2e-16
factor(A)77.5                 -3.73581   0.13059  -28.912 < 2e-16
factor(A)82.5                 -4.08803   0.14796  -27.300 < 2e-16
factor(A)87.5                 -4.46840   0.16751  -26.320 < 2e-16
factor(A)92.5                 -4.86721   0.18853  -25.606 < 2e-16
factor(A)97.5                 -5.28556   0.21097  -24.989 < 2e-16
factor(A)102.5                -5.72387   0.23494  -24.382 < 2e-16
factor(A)107.5                -6.18334   0.25893  -23.565 < 2e-16
factor(A)112.5                -6.66409   0.28416  -22.905 < 2e-16
factor(A)117.5                -7.16612   0.31087  -22.344 < 2e-16
factor(A)122.5                -7.69048   0.33859  -21.662 < 2e-16
factor(A)127.5                -8.23809   0.36776  -20.734 < 2e-16
factor(A)132.5                -8.81006   0.39847  -21.219 < 2e-16
factor(A)137.5                -9.40751   0.43074  -21.873 < 2e-16
factor(A)142.5                -10.02265  0.46449  -21.978 < 2e-16
factor(A)147.5                -10.66631  0.49975  -21.220 < 2e-16
factor(A)152.5                -11.33027  0.53634  -20.910 < 2e-16
factor(A)157.5                -12.01626  0.57486  -20.800 < 2e-16
factor(A)162.5                -12.72552  0.61538  -20.633 < 2e-16
factor(A)167.5                -13.45937  0.65766  -20.538 < 2e-16
factor(A)172.5                -14.21870  0.70201  -20.292 < 2e-16
factor(A)177.5                -15.00612  0.74858  -19.883 < 2e-16
factor(A)182.5                -15.83396  0.79647  -19.801 < 2e-16
factor(A)187.5                -16.69302  0.84569  -19.297 < 2e-16
factor(A)192.5                -17.58549  0.89637  -18.834 < 2e-16
factor(A)197.5                -18.51318  0.94868  -18.377 < 2e-16
factor(A)202.5                -19.47779  1.00332  -18.332 < 2e-16
factor(A)207.5                -20.48169  1.06049  -18.164 < 2e-16
factor(A)212.5                -21.52715  1.11973  -17.960 < 2e-16
factor(A)217.5                -22.61615  1.18163  -17.779 < 2e-16
factor(A)222.5                -23.75179  1.24535  -17.567 < 2e-16
factor(A)227.5                -24.93564  1.31123  -17.490 < 2e-16
factor(A)232.5                -26.17106  1.37945  -17.683 < 2e-16
factor(A)237.5                -27.45918  1.45043  -17.761 < 2e-16
factor(A)242.5                -28.80153  1.52370  -17.668 < 2e-16
factor(A)247.5                -30.20258  1.60032  -17.605 < 2e-16
factor(A)252.5                -31.65597  1.68016  -17.480 < 2e-16
factor(A)257.5                -33.16407  1.76223  -17.796 < 2e-16
factor(A)262.5                -34.72959  1.84706  -17.889 < 2e-16
factor(A)267.5                -36.35594  1.93428  -17.659 < 2e-16
factor(A)272.5                -38.04932  2.02403  -18.827 < 2e-16
factor(A)277.5                -39.80451  2.11510  -18.813 < 2e-16
factor(A)282.5                -41.59862  2.20832  -19.274 < 2e-16
factor(A)287.5                -43.43787  2.30376  -19.262 < 2e-16
factor(A)292.5                -45.32536  2.40153  -19.714 < 2e-16
factor(A)297.5                -47.25515  2.50154  -19.254 < 2e-16
factor(A)302.5                -49.22912  2.60377  -18.898 < 2e-16
factor(A)307.5                -51.24775  2.70821  -19.592 < 2e-16
factor(A)312.5                -53.31320  2.81474  -19.073 < 2e-16
factor(A)317.5                -55.42782  2.92345  -18.861 < 2e-16
factor(A)322.5                -57.59464  3.03432  -18.824 < 2e-16
factor(A)327.5                -59.81664  3.14732  -19.112 < 2e-16
```

Age-Period and Age-Cohort models (AP-AC)
Graph of estimates with confidence intervals

Age-Period and Age-Cohort models (AP–AC)
Age-drift model

Statistical Analysis in the Lexis Diagram:
Age-Period-Cohort models
May 2016
Max Planck Institut for Demographic Research, Rostock
http://BendixCarstensen/APC/MPIDR-2016
Linear effect of period:

$$\log[\lambda(a, p)] = \alpha_a + \beta_p = \alpha_a + \beta(p - p_0)$$

that is, $\beta_p = \beta(p - p_0)$.

Linear effect of cohort:

$$\log[\lambda(a, p)] = \tilde{\alpha}_a + \gamma_c = \tilde{\alpha}_a + \gamma(c - c_0)$$

that is, $\gamma_c = \gamma(c - c_0)$
Age and linear effect of period:

```r
> apd <- glm( D ~ factor(A) - 1 + I(P-1970.5) +
+       offset( log(Y) ),
+       family=poisson )
> summary( apd )
```

Call:
glm(formula = D ~ factor(A) - 1 + I(P - 1970.5) + offset(log(Y)), family = poisson)

Deviance Residuals:

```
       Min          1Q      Median          3Q         Max
-2.97593  -0.77091   0.02809  0.95914  2.93076
```

Coefficients:

```
                              Estimate Std. Error     z value  Pr(>|z|)
factor(A)17.5               -3.58065   0.06306     -56.79  <2e-16 
...
factor(A)57.5               -3.17579   0.06256     -50.77  <2e-16 
I(P - 1970.5)               0.02653    0.00100      26.52  <2e-16 
```

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 89358.53 on 81 degrees of freedom
Residual deviance: 126.07 on 71 degrees of freedom
Age and linear effect of cohort:

> acd <- glm( D ~ factor( A ) - 1 + I(C-1933) +
+ offset( log( Y ) ),
+ family=poisson )
> summary( acd )

Call:
glm(formula = D ~ factor(A) - 1 + I(C - 1933) + offset(log(Y)), family = poisson)

Deviance Residuals:

Min        1Q  Median        3Q       Max
-2.97593  -0.77091   0.02809   0.95914   2.93076

Coefficients:

                         Estimate Std. Error   z value Pr(>|z|)
factor(A)17.5             -4.11117  0.06760  -60.82   <2e-16
...                        ...
factor(A)57.5             -2.64527  0.06423  -41.19   <2e-16
I(C - 1933)               0.02653  0.00100   26.52   <2e-16

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 89358.53 on 81 degrees of freedom
Residual deviance: 126.07 on 71 degrees of freedom
What goes on?

\[ \alpha_a + \beta (p - p_0) = \alpha_a + \beta (a + c - (a_0 + c_0)) \]

\[ = \alpha_a + \beta (a - a_0) + \beta (c - c_0) \]

cohort age-effect

The two models are the same. The parametrization is different.

The age-curve refers either
• to a period (cross-sectional rates) or
• to a cohort (longitudinal rates).
Which age-curve is period and which is cohort?

Age-drift model (Ad)
Age at entry

Statistical Analysis in the Lexis Diagram:
Age-Period-Cohort models
May 2016
Max Planck Institut for Demographic Research, Rostock
http://BendixCarstensen/APC/MPIDR-2016
Age at entry as covariate

\( t \): time since entry
\( e \): age at entry
\( a = e + t \): current age

\[
\log(\lambda(a, t)) = f(t) + \beta e = (f(t) - \beta t) + \beta a
\]

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).
Non-linear effects of time-scales

Arbitrary effects of the three variables $t$, $a$ and $e$: $\implies$ genuine extension of the model.

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

Three quantities can be arbitrarily moved between the three functions:

$$\tilde{f}(t) = f(a) - \mu_a - \mu_e + \gamma t$$
$$\tilde{g}(a) = g(p) + \mu_a - \gamma a$$
$$\tilde{h}(e) = h(c) + \mu_a + \gamma e$$

because $t - a + e = 0$.

This is the age-period-cohort modelling problem again.
“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. This would require modelling of multiple timescales. Which is best accomplished by splitting time.
Age-Period-Cohort model

Statistical Analysis in the
Lexis Diagram:
Age-Period-Cohort models
May 2016
Max Planck Institut for Demographic Research, Rostock
http://BendixCarstensen/APC/MPIDR-2016
The age-period-cohort model

\[ \log[\lambda(a, p)] = \alpha_a + \beta_p + \gamma_c \]

- Three effects:
  - Age (at diagnosis)
  - Period (of diagnosis)
  - Cohort (of birth)
- Modelled on the same scale.
- No assumptions about the shape of effects.
- Levels of A, P and C are assumed exchangeable
- no assumptions about the relationship of parameter estimates and the scaled values of A, P and C
> library( Epi )
> load( file="../data/testisDK.Rda" )
> head( testisDK )

A  P  D  Y
1 17.5 1950.5 7 744.2172
2 22.5 1950.5 31 744.7055
3 27.5 1950.5 62 781.8272
4 32.5 1950.5 66 774.5415
5 37.5 1950.5 56 782.8932
6 42.5 1950.5 47 754.3220

> m.apc <- glm( D ~ factor( A ) + factor( P ) + factor( P-A ),
+ offset = log(Y), family = poisson, data = testisDK )
> summary( m.apc )
### Fitting the model in R II

**Call:**
```r
glm(formula = D ~ factor(A) + factor(P) + factor(P - A), family = poisson,
data = testisDK, offset = log(Y))
```

**Deviance Residuals:**

<table>
<thead>
<tr>
<th></th>
<th>Min</th>
<th>1Q</th>
<th>Median</th>
<th>3Q</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-1.55709</td>
<td>-0.56174</td>
<td>0.01096</td>
<td>0.51221</td>
<td>1.32770</td>
</tr>
</tbody>
</table>

**Coefficients:** (1 not defined because of singularities)

| Estimator          | Estimate | Std. Error | z value | Pr(>|z|) |
|--------------------|----------|------------|---------|---------|
| (Intercept)        | -4.01129 | 0.16094    | -24.925 | < 2e-16 |
| factor(A)22.5      | 1.23961  | 0.07745    | 16.005  | < 2e-16 |
| factor(A)27.5      | 1.70594  | 0.08049    | 21.194  | < 2e-16 |
| factor(A)32.5      | 1.83935  | 0.08946    | 20.561  | < 2e-16 |
| factor(A)37.5      | 1.71786  | 0.10217    | 16.813  | < 2e-16 |
| factor(A)42.5      | 1.48259  | 0.11708    | 12.663  | < 2e-16 |
| factor(A)47.5      | 1.09057  | 0.13447    | 8.110   | 5.07e-16 |
| factor(A)52.5      | 0.76631  | 0.15271    | 5.018   | 5.22e-07 |
| factor(A)57.5      | 0.41050  | 0.16094    | 2.551   | 0.010751 |

**Age-Period-Cohort model (APC-cat)**

- 179/327
### Fitting the model in R III

| Age Group | Estimate | Std. Error | t value | Pr(>|t|) |
|-----------|----------|------------|---------|----------|
| 1955.5    | 0.18645  | 0.07514    | 2.482   | 0.013082 |
| 1960.5    | 0.37398  | 0.07949    | 4.705   | 2.54e-06 |
| 1965.5    | 0.52062  | 0.08858    | 5.877   | 4.17e-09 |
| 1970.5    | 0.72806  | 0.10013    | 7.271   | 3.56e-13 |
| 1975.5    | 0.90736  | 0.11422    | 7.944   | 1.96e-15 |
| 1980.5    | 1.02698  | 0.12978    | 7.913   | 2.51e-15 |
| 1985.5    | 1.06237  | 0.14641    | 7.256   | 3.98e-13 |
| 1990.5    | 1.10813  | 0.16094    | 6.885   | 5.76e-12 |
| 1995.5    | 1.1464  | 0.17621    | 6.519   | 7.95e-12 |
| 2000.5    | 1.1841  | 0.19366    | 6.002   | 1.50e-11 |
| 2005.5    | 1.2218  | 0.21212    | 5.349   | 2.25e-08 |

**Age-Period-Cohort model (APC-cat)**
Fitting the model in R IV

<table>
<thead>
<tr>
<th>Factor</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>z value</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>factor(P - A)1953</td>
<td>-0.32857</td>
<td>0.14601</td>
<td>-2.250</td>
<td>0.02443</td>
</tr>
<tr>
<td>factor(P - A)1958</td>
<td>-0.23140</td>
<td>0.14615</td>
<td>-1.583</td>
<td>0.11335</td>
</tr>
<tr>
<td>factor(P - A)1963</td>
<td>-0.18244</td>
<td>0.14978</td>
<td>-1.218</td>
<td>0.22320</td>
</tr>
<tr>
<td>factor(P - A)1968</td>
<td>-0.20961</td>
<td>0.16143</td>
<td>-1.298</td>
<td>0.19414</td>
</tr>
<tr>
<td>factor(P - A)1973</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 2463.197 on 80 degrees of freedom
Residual deviance: 35.459 on 49 degrees of freedom
AIC: 584.5

Number of Fisher Scoring iterations: 4
No. of parameters

$A$ has 9 levels
$P$ has 9 levels
$C = P - A$ has 17 levels

Age-drift model has $A + 1 = 10$ parameters
Age-period model has $A + P - 1 = 17$ parameters
Age-cohort model has $A + C - 1 = 25$ parameters
Age-period-cohort model has $A + P + C - 3 = 32$ parameters:

> \text{length( coef(m.apc) )}

[1] 33

> \text{sum( !is.na(coef(m.apc)) )}

[1] 32
Relationship of models

Testis cancer, Denmark

Age
865.08 / 72

739.01 / 1
p=0.0000

Age−drift
126.07 / 71

8.37 / 7
p=0.3010

Age−Period
117.7 / 64

60.6 / 15
p=0.0000

Age−Cohort
65.47 / 56

82.24 / 15
p=0.0000

Age−Period−Cohort
35.46 / 49

8.37 / 7
p=0.3010

739.01 / 1
p=0.0000

Testis cancer, Denmark

Age−Period−Cohort model (APC-cat)
## Test for effects

<table>
<thead>
<tr>
<th>Model</th>
<th>Deviance</th>
<th>d.f.</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age - drift</td>
<td>126.07</td>
<td>71</td>
<td></td>
</tr>
<tr>
<td>∆</td>
<td>60.60</td>
<td>15</td>
<td>0.000</td>
</tr>
<tr>
<td>Age - cohort</td>
<td>65.47</td>
<td>56</td>
<td></td>
</tr>
<tr>
<td>∆</td>
<td>30.01</td>
<td>7</td>
<td>0.000</td>
</tr>
<tr>
<td>Age - period - cohort</td>
<td>35.46</td>
<td>49</td>
<td></td>
</tr>
<tr>
<td>∆</td>
<td>82.24</td>
<td>15</td>
<td>0.000</td>
</tr>
<tr>
<td>Age - period</td>
<td>117.70</td>
<td>64</td>
<td></td>
</tr>
<tr>
<td>∆</td>
<td>8.37</td>
<td>7</td>
<td>0.301</td>
</tr>
<tr>
<td>Age - drift</td>
<td>126.07</td>
<td>71</td>
<td></td>
</tr>
</tbody>
</table>
How to choose a parametrization

- Standard approach: Put extremes of periods or cohorts to 0, and choose a reference for the other.
- Clayton & Schifflers: only 2nd order differences are invariants:

\[ \alpha_{i-1} - 2\alpha_i + \alpha_{i+1} \]

Implemented in Epi via the contrast type contr.2nd (later).

- Holford: Extract linear effects by regression:

\[
\lambda(a, p) = \hat{\alpha}_a + \hat{\beta}_p + \tilde{\alpha}_a + \tilde{\beta}_p + \hat{\mu}_a + \hat{\mu}_p + \hat{\delta}_a a + \hat{\delta}_p p + \tilde{\gamma}_c + \tilde{\gamma}_c + \hat{\mu}_c + \hat{\delta}_c c
\]
Putting it together again

Assumptions are needed, e.g.:

- Age is the major time scale.
- Cohort is the secondary time scale (the major secular trend).
- \( c_0 \) is the reference cohort.
- Period is the residual time scale: 0 on average, 0 slope.
Period effect, on average 0, slope is 0:

\[ g(p) = \tilde{\beta}_p = \beta_p - \hat{\mu}_p - \hat{\delta}_p p \]

Cohort effect, absorbing all time-trend \((\delta_p p = \delta_p (a + c))\) and risk relative to \(c_0\):

\[ h(c) = \gamma_c - \gamma_{c_0} + \hat{\delta}_p (c - c_0) \]

The rest is the age-effect:

\[ f(a) = \alpha_a + \hat{\mu}_p + \hat{\delta}_p a + \hat{\delta}_p c_0 + \gamma_{c_0} \]
How it all adds up:

\[
\lambda(a, p) = \hat{\alpha}_a + \hat{\beta}_p + \hat{\gamma}_c
\]

\[
= \hat{\alpha}_a + \gamma_{c_0} + \hat{\mu}_p + \delta_p(a + c_0) + \hat{\beta}_p - \hat{\mu}_p - \delta_p(a + c) + \hat{\gamma}_c - \gamma_{c_0} + \delta_p(c - c_0)
\]

Only the regression on period is needed! (For this model...)
Age-Period-Cohort model (APC-cat)
A simple practical approach

- First fit the age-cohort model, with cohort $c_0$ as reference and get estimates $\hat{\alpha}_a$ and $\hat{\gamma}_c$:

  \[
  \log[\lambda(a, p)] = \hat{\alpha}_a + \hat{\gamma}_c
  \]

- Then consider the full APC-model with age and cohort effects as estimated:

  \[
  \log[\lambda(a, p)] = \hat{\alpha}_a + \hat{\gamma}_c + \beta_p
  \]
The residual period effect can be estimated if we note that for the number of cases we have:

$$\log(\text{expected cases}) = \log[\lambda(a, p)Y] = \hat{\alpha}_a + \hat{\gamma}_c + \log(Y) + \beta_p$$

This is analogous to the expression for a Poisson model in general, but now is the offset not just $\log(Y)$ but $\hat{\alpha}_a + \hat{\gamma}_c + \log(Y)$, the log of the fitted values from the age-cohort model.

$\beta_p$s are estimated in a Poisson model with this as offset.

Advantage: We get the standard errors for free.
Age-Period-Cohort model (APC-cat)
Age-Period-Cohort model (APC-cat)
Using `contr.2nd` I

```r
> attach( testisDK )
> ( nA <- nlevels(factor(A)) )
[1] 9

> ( nP <- nlevels(factor(P)) )
[1] 9

> ( nC <- nlevels(factor(P-A)) )
[1] 17
```
Using `contr.2nd II`

```r
mp <- glm( D ~ factor(A) - 1 + I(P-1970) +
          C( factor(P), contr.2nd, nP-2 ) +
          C( factor(P-A), contr.2nd, nC-2 ),
          offset = log(Y), family = poisson, data = testisDK )
mc <- glm( D ~ factor(A) - 1 + I(P-A-1940) +
          C( factor(P), contr.2nd, nP-2 ) +
          C( factor(P-A), contr.2nd, nC-2 ),
          offset = log(Y), family = poisson, data = testisDK )
c( m.apc$deviance,
    mp$deviance,
    mc$deviance )
```

```
[1] 35.4587 35.4587 35.4587
```

```r
round( cbind( ci.exp(mp,subset="P)"),
        ci.exp(mc,subset="P") ), 4 )
```
Using \textit{contr.2nd} III

<table>
<thead>
<tr>
<th></th>
<th>exp(Est.)</th>
<th>2.5%</th>
<th>97.5%</th>
<th>exp(Est.)</th>
<th>2.5%</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>C(factor(P), contr.2nd, nP - 2)1</td>
<td>1.0011</td>
<td>0.7860</td>
<td>1.2751</td>
<td>1.0011</td>
<td>0.7860</td>
<td>1.2751</td>
</tr>
<tr>
<td>C(factor(P), contr.2nd, nP - 2)2</td>
<td>0.9599</td>
<td>0.7680</td>
<td>1.1998</td>
<td>0.9599</td>
<td>0.7680</td>
<td>1.1998</td>
</tr>
<tr>
<td>C(factor(P), contr.2nd, nP - 2)3</td>
<td>1.0627</td>
<td>0.8651</td>
<td>1.3053</td>
<td>1.0627</td>
<td>0.8651</td>
<td>1.3053</td>
</tr>
<tr>
<td>C(factor(P), contr.2nd, nP - 2)4</td>
<td>0.9722</td>
<td>0.8080</td>
<td>1.1699</td>
<td>0.9722</td>
<td>0.8080</td>
<td>1.1699</td>
</tr>
<tr>
<td>C(factor(P), contr.2nd, nP - 2)5</td>
<td>0.9421</td>
<td>0.7977</td>
<td>1.1126</td>
<td>0.9421</td>
<td>0.7977</td>
<td>1.1126</td>
</tr>
<tr>
<td>C(factor(P), contr.2nd, nP - 2)6</td>
<td>0.9192</td>
<td>0.7893</td>
<td>1.0706</td>
<td>0.9192</td>
<td>0.7893</td>
<td>1.0706</td>
</tr>
<tr>
<td>C(factor(P), contr.2nd, nP - 2)7</td>
<td>1.0104</td>
<td>0.8750</td>
<td>1.1668</td>
<td>1.0104</td>
<td>0.8750</td>
<td>1.1668</td>
</tr>
</tbody>
</table>

\begin{verbatim}
> round( rbind( ci.exp(mp,subset="I"),
+         ci.exp(mc,subset="I") ), 4 )

   exp(Est.)   2.5%   97.5%
I(P - 1970)   1.0468  0.926  1.1833
I(P - A - 1940) 1.0468  0.926  1.1833
\end{verbatim}

Age-Period-Cohort model (APC-cat)
Using `contr.2nd IV`

```r
> matplot( sort(unique(testisDK$A)),
+         cbind(ci.exp(mp,subset="A"),
+               ci.exp(mc,subset="A"){*}100,
+         log="y", xlab="Age", ylab="Incidence rate per 100,000 PY",
+         type="l",lty=1,lwd=c(3,1,1),col=rep(c("red","blue"),each=2) )
```
Tabulation in the Lexis diagram

Statistical Analysis in the Lexis Diagram:
Age-Period-Cohort models
May 2016
Max Planck Institut for Demographic Research, Rostock
http://BendixCarstensen/APC/MPIDR-2016

Lexis-tab
Tabulation of register data

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>773.8</td>
<td>744.2</td>
<td>794.1</td>
<td>972.9</td>
<td>1051.5</td>
<td>952.5</td>
</tr>
<tr>
<td>25</td>
<td>813.0</td>
<td>744.7</td>
<td>721.8</td>
<td>770.9</td>
<td>960.3</td>
<td>1053.8</td>
</tr>
<tr>
<td>35</td>
<td>56</td>
<td>66</td>
<td>82</td>
<td>711.6</td>
<td>103</td>
<td>124</td>
</tr>
<tr>
<td>45</td>
<td>29</td>
<td>30</td>
<td>37</td>
<td>54</td>
<td>45</td>
<td>66</td>
</tr>
<tr>
<td>55</td>
<td>6</td>
<td>471.0</td>
<td>14</td>
<td>571.1</td>
<td>25</td>
<td>622.5</td>
</tr>
</tbody>
</table>

Testis cancer cases in Denmark.

Male person-years in Denmark.
Tabulation of register data

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>773.8</td>
<td></td>
<td>744.2</td>
<td></td>
<td>972.9</td>
<td></td>
<td>1051.5</td>
<td></td>
<td>961.0</td>
<td></td>
<td>1011.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10</td>
<td></td>
<td>7</td>
<td></td>
<td>13</td>
<td></td>
<td>13</td>
<td></td>
<td>33</td>
<td></td>
<td>35</td>
</tr>
<tr>
<td></td>
<td></td>
<td>813.0</td>
<td></td>
<td>744.7</td>
<td></td>
<td>721.8</td>
<td></td>
<td>770.9</td>
<td></td>
<td>960.3</td>
<td></td>
<td>1053.8</td>
</tr>
<tr>
<td>25</td>
<td>30</td>
<td>813.0</td>
<td></td>
<td>744.7</td>
<td></td>
<td>46</td>
<td></td>
<td>721.8</td>
<td></td>
<td>960.3</td>
<td></td>
<td>1053.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>56</td>
<td>799.3</td>
<td></td>
<td>774.5</td>
<td></td>
<td>82</td>
<td></td>
<td>769.3</td>
<td></td>
<td>711.6</td>
<td></td>
<td>700.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>56</td>
<td></td>
<td>62</td>
<td></td>
<td>56</td>
<td></td>
<td>67</td>
<td></td>
<td>99</td>
<td></td>
<td>124</td>
</tr>
<tr>
<td></td>
<td></td>
<td>799.3</td>
<td></td>
<td>66</td>
<td></td>
<td>82</td>
<td></td>
<td>769.3</td>
<td></td>
<td>103</td>
<td></td>
<td>124</td>
</tr>
<tr>
<td></td>
<td></td>
<td>55</td>
<td></td>
<td>62</td>
<td></td>
<td>63</td>
<td></td>
<td>82</td>
<td></td>
<td>87</td>
<td></td>
<td>103</td>
</tr>
<tr>
<td></td>
<td></td>
<td>700.5</td>
<td></td>
<td>62</td>
<td></td>
<td>63</td>
<td></td>
<td>82</td>
<td></td>
<td>87</td>
<td></td>
<td>103</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>29</td>
<td></td>
<td>30</td>
<td></td>
<td>37</td>
<td></td>
<td>54</td>
<td></td>
<td>67</td>
<td></td>
<td>124</td>
</tr>
<tr>
<td></td>
<td></td>
<td>622.1</td>
<td></td>
<td>30</td>
<td></td>
<td>37</td>
<td></td>
<td>54</td>
<td></td>
<td>67</td>
<td></td>
<td>124</td>
</tr>
<tr>
<td>45</td>
<td>35</td>
<td>694.1</td>
<td></td>
<td>47</td>
<td></td>
<td>65</td>
<td></td>
<td>749.9</td>
<td></td>
<td>85</td>
<td></td>
<td>103</td>
</tr>
<tr>
<td></td>
<td></td>
<td>35</td>
<td></td>
<td>694.1</td>
<td></td>
<td>47</td>
<td></td>
<td>65</td>
<td></td>
<td>749.9</td>
<td></td>
<td>85</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Testis cancer cases in Denmark.

Male person-years in Denmark.

Tabulation in the Lexis diagram (Lexis-tab) 199/327
## Tabulation of register data

<table>
<thead>
<tr>
<th>Calendar time</th>
<th>Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>1983</td>
<td>30</td>
</tr>
<tr>
<td>1984</td>
<td>31</td>
</tr>
<tr>
<td>1985</td>
<td>32</td>
</tr>
<tr>
<td>1986</td>
<td>33</td>
</tr>
<tr>
<td>1987</td>
<td>34</td>
</tr>
<tr>
<td>1988</td>
<td>35</td>
</tr>
</tbody>
</table>

Testis cancer cases in Denmark.

Male person-years in Denmark.
Tabulation of register data

Testis cancer cases in Denmark.

Male person-years in Denmark.
Tabulation of register data

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>30</td>
<td>31</td>
<td>32</td>
<td>33</td>
<td>34</td>
<td>35</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>10</td>
<td>0</td>
<td>17</td>
</tr>
<tr>
<td>19.0</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>6</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>19.1</td>
<td>8</td>
<td>19.2</td>
<td>19.0</td>
<td>18.8</td>
<td>19.1</td>
<td>19.1</td>
</tr>
<tr>
<td>19.2</td>
<td>3</td>
<td>19.2</td>
<td>3</td>
<td>19.0</td>
<td>8</td>
<td>19.1</td>
</tr>
<tr>
<td>19.3</td>
<td>3</td>
<td>19.0</td>
<td>3</td>
<td>19.1</td>
<td>3</td>
<td>19.2</td>
</tr>
<tr>
<td>19.7</td>
<td>6</td>
<td>19.2</td>
<td>4</td>
<td>18.9</td>
<td>5</td>
<td>19.2</td>
</tr>
<tr>
<td>19.0</td>
<td>6</td>
<td>18.8</td>
<td>4</td>
<td>19.1</td>
<td>8</td>
<td>19.1</td>
</tr>
<tr>
<td>19.1</td>
<td>3</td>
<td>19.0</td>
<td>4</td>
<td>19.1</td>
<td>5</td>
<td>19.2</td>
</tr>
<tr>
<td>18.9</td>
<td>3</td>
<td>19.0</td>
<td>4</td>
<td>19.1</td>
<td>6</td>
<td>19.2</td>
</tr>
<tr>
<td>18.9</td>
<td>2</td>
<td>19.0</td>
<td>2</td>
<td>18.8</td>
<td>5</td>
<td>19.1</td>
</tr>
<tr>
<td>18.9</td>
<td>6</td>
<td>18.9</td>
<td>3</td>
<td>19.2</td>
<td>6</td>
<td>19.2</td>
</tr>
<tr>
<td>18.9</td>
<td>4</td>
<td>19.7</td>
<td>1</td>
<td>19.2</td>
<td>3</td>
<td>19.2</td>
</tr>
</tbody>
</table>

Testis cancer cases in Denmark.

Male person-years in Denmark.

Subdivision by year of birth (cohort).

Tabulation in the Lexis diagram (Lexis-tab)
Major sets in the Lexis diagram

A-sets: Classification by age and period. (□)
B-sets: Classification by age and cohort. (开设)
C-sets: Classification by cohort and period. (开设)

The mean age, period and cohort for these sets is just the mean of the tabulation interval.

The mean of the third variable is found by using $a = p - c$. 

Analysis of rates from a complete observation in a Lexis digram need not be restricted to these classical sets classified by two factors.

We may classify cases and risk time by all three factors:

**Upper triangles:** Classification by age and period, earliest born cohort. (⏀)

**Lower triangles:** Classification by age and cohort, last born cohort. (▽)
Mean time in triangles

Modelling requires that each set (=observation in the dataset) be assigned a value of age, period and cohort. So for each triangle we need:

- mean age at risk.
- mean date at risk.
- mean cohort at risk.
Means in upper (A) and lower (B) triangles:

Tabulation in the Lexis diagram (Lexis-tab)
Upper triangles (\(\triangleright\)), A:

\[
E_A(a) = \int_{p=0}^{p=1} \int_{a=0}^{a=p} a \times 2 \, da \, dp = \int_{p=0}^{p=1} 1 - p^2 \, dp = \frac{2}{3}
\]

\[
E_A(p) = \int_{a=0}^{a=1} \int_{p=0}^{p=a} p \times 2 \, dp \, da = \int_{a=0}^{a=1} a^2 \, dp = \frac{1}{3}
\]

\[
E_A(c) = \frac{1}{3} - \frac{2}{3} = -\frac{1}{3}
\]
Lower triangles (△), B:

\[ E_B(a) = \int_{p=0}^{p=1} \int_{a=0}^{a=p} a \times 2 \, da \, dp = \int_{p=0}^{p=1} p^2 \, dp = \frac{1}{3} \]

\[ E_B(p) = \int_{a=0}^{a=1} \int_{p=a}^{p=1} p \times 2 \, dp \, da = \int_{a=0}^{a=1} 1 - a^2 \, dp = \frac{2}{3} \]

\[ E_B(c) = \frac{2}{3} - \frac{1}{3} = \frac{1}{3} \]
Tabulation by age, period and cohort

Gives triangular sets with differing mean age, period and cohort:

These correct midpoints for age, period and cohort must be used in modelling.
Population figures

Population figures in the form of size of the population at certain date are available from most statistical bureaus.

This corresponds to population sizes along the vertical lines in the diagram.

We want risk time figures for the population in the squares and triangles in the diagram.
Prevalent population figures

\( \ell_{a,p} \) is the number of persons in age class \( a \) alive at the beginning of period (=year) \( p \).

The aim is to compute person-years for the triangles \( A \) and \( B \), respectively.
The area of the triangle is $1/2$, so the uniform measure over the triangle has density 2. Therefore a person dying in age $a$ at date $p$ in $A$ contributes $p$ risk time, so the average will be:

$$\int_{p=0}^{p=1} \int_{a=p}^{a=1} 2p \, da \, dp$$

$$= \int_{p=0}^{p=1} 2p - 2p^2 \, dp$$

$$= \left[ p^2 - \frac{2p^3}{3} \right]_{p=0}^{p=1} = \frac{1}{3}$$
A person dying in age $a$ at date $p$ in $B$ contributes $p - a$ risk time in $A$, so the average will be (aggregating using the density 2 of the uniform measure):

$$\int_{p=0}^{p=1} \int_{a=0}^{a=p} 2(p - a) \, da \, dp$$

$$= \int_{p=0}^{p=1} [2pa - a^2]_{a=0}^{a=p} \, dp$$

$$= \int_{p=0}^{p=1} p^2 \, dp = \frac{1}{3}$$
A person dying in age $a$ at date $p$ in $\mathbf{B}$ contributes a risk time in $\mathbf{B}$, so the average will be:

\[
\int_{p=0}^{p=1} \int_{a=0}^{a=p} 2a \, da \, dp
\]

\[
= \int_{p=0}^{p=1} p^2 \, dp = \frac{1}{3}
\]
Contributions to risk time in A and B:

<table>
<thead>
<tr>
<th></th>
<th>A:</th>
<th>B:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Survivors:</td>
<td>$\ell_{a+1,p+1} \times \frac{1}{2} y$</td>
<td>$\ell_{a+1,p+1} \times \frac{1}{2} y$</td>
</tr>
<tr>
<td>Dead in A:</td>
<td>$\frac{1}{2} (\ell_{a,p} - \ell_{a+1,p+1}) \times \frac{1}{3} y$</td>
<td>$\frac{1}{2} (\ell_{a,p} - \ell_{a+1,p+1}) \times \frac{1}{3} y$</td>
</tr>
<tr>
<td>Dead in B:</td>
<td>$\frac{1}{2} (\ell_{a,p} - \ell_{a+1,p+1}) \times \frac{1}{3} y$</td>
<td>$\frac{1}{2} (\ell_{a,p} - \ell_{a+1,p+1}) \times \frac{1}{3} y$</td>
</tr>
<tr>
<td>$\sum$</td>
<td>$(\frac{1}{3} \ell_{a,p} + \frac{1}{6} \ell_{a+1,p+1}) \times 1 y$</td>
<td>$(\frac{1}{6} \ell_{a,p} + \frac{1}{3} \ell_{a+1,p+1}) \times 1 y$</td>
</tr>
</tbody>
</table>

Tabulation in the Lexis diagram (Lexis-tab)
Population as of 1. January from Statistics Denmark:

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>22</td>
<td>33435</td>
<td>33540</td>
<td>32272</td>
<td>32637</td>
<td>32802</td>
<td>31709</td>
</tr>
<tr>
<td>23</td>
<td>35357</td>
<td>33579</td>
<td>33742</td>
<td>34163</td>
<td>32853</td>
<td>33156</td>
</tr>
<tr>
<td>24</td>
<td>38199</td>
<td>35400</td>
<td>33674</td>
<td>37803</td>
<td>34353</td>
<td>33070</td>
</tr>
<tr>
<td>25</td>
<td>37958</td>
<td>38257</td>
<td>35499</td>
<td>37318</td>
<td>37955</td>
<td>34526</td>
</tr>
<tr>
<td>26</td>
<td>38194</td>
<td>38048</td>
<td>38341</td>
<td>37292</td>
<td>37371</td>
<td>38119</td>
</tr>
<tr>
<td>27</td>
<td>39891</td>
<td>38221</td>
<td>38082</td>
<td>39273</td>
<td>37403</td>
<td>37525</td>
</tr>
</tbody>
</table>
Exercise:

Fill in the risk time figures in as many triangles as possible from the previous table for men and women, respectively.

Look at the \( N2Y \) function in Epi.
Summary:

Population risk time:

A: \( \left( \frac{1}{3} \ell_{a,p} + \frac{1}{6} \ell_{a+1,p+1} \right) \times 1\text{y} \)

B: \( \left( \frac{1}{6} \ell_{a-1,p} + \frac{1}{3} \ell_{a,p+1} \right) \times 1\text{y} \)

Mean age, period and cohort:
\( \frac{1}{3} \) into the interval.

Tabulation in the Lexis diagram (Lexis-tab)
APC-model for triangular data

Statistical Analysis in the
Lexis Diagram:
Age-Period-Cohort models
May 2016
Max Planck Institut for Demographic Research, Rostock
http://BendixCarstensen/APC/MPIDR-2016
Model for triangular data

- One parameter per distinct value on each timescale.
- Example: 3 age-classes and 3 periods:
  - 6 age parameters
  - 6 period parameters
  - 10 cohort parameters

- Model:

\[ \lambda_{ap} = \alpha_a + \beta_p + \gamma_c \]
Problem: Disconnected design!

Log-likelihood contribution from one triangle:

\[
D_{ap} \log(\lambda_{ap}) - \lambda_{ap} Y_{ap} = D_{ap} \log(\alpha_a + \beta_p + \gamma_c) - (\alpha_a + \beta_p + \gamma_c) Y_{ap}
\]

The log-likelihood can be separated:

\[
\sum_{a,p \in \triangle} D_{ap} \log(\lambda_{ap}) - \lambda_{ap} Y_{ap} + \sum_{a,p \in \triangledown} D_{ap} \log(\lambda_{ap}) - \lambda_{ap} Y_{ap}
\]

No common parameters between terms — we have two separate models:
One for upper triangles, one for lower.

APC-model for triangular data (APC-tri)
Illustration by lung cancer data

```r
> library( Epi )
> data( lungDK )
> lungDK[1:10,]

<table>
<thead>
<tr>
<th>A5</th>
<th>P5</th>
<th>C5</th>
<th>up</th>
<th>Ax</th>
<th>Px</th>
<th>Cx</th>
<th>D</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>40</td>
<td>1943</td>
<td>1898</td>
<td>1</td>
<td>43.3333</td>
<td>1944.667</td>
<td>1901.333</td>
<td>52</td>
</tr>
<tr>
<td>2</td>
<td>40</td>
<td>1943</td>
<td>1903</td>
<td>0</td>
<td>41.6667</td>
<td>1946.333</td>
<td>1904.667</td>
<td>28</td>
</tr>
<tr>
<td>3</td>
<td>40</td>
<td>1948</td>
<td>1903</td>
<td>1</td>
<td>43.3333</td>
<td>1949.667</td>
<td>1906.333</td>
<td>51</td>
</tr>
<tr>
<td>4</td>
<td>40</td>
<td>1948</td>
<td>1908</td>
<td>0</td>
<td>41.6667</td>
<td>1951.333</td>
<td>1909.667</td>
<td>30</td>
</tr>
<tr>
<td>5</td>
<td>40</td>
<td>1953</td>
<td>1908</td>
<td>1</td>
<td>43.3333</td>
<td>1954.667</td>
<td>1911.333</td>
<td>50</td>
</tr>
<tr>
<td>6</td>
<td>40</td>
<td>1953</td>
<td>1913</td>
<td>0</td>
<td>41.6667</td>
<td>1956.333</td>
<td>1914.667</td>
<td>23</td>
</tr>
<tr>
<td>7</td>
<td>40</td>
<td>1958</td>
<td>1913</td>
<td>1</td>
<td>43.3333</td>
<td>1959.667</td>
<td>1916.333</td>
<td>56</td>
</tr>
<tr>
<td>8</td>
<td>40</td>
<td>1958</td>
<td>1918</td>
<td>0</td>
<td>41.6667</td>
<td>1961.333</td>
<td>1919.667</td>
<td>43</td>
</tr>
<tr>
<td>9</td>
<td>40</td>
<td>1963</td>
<td>1918</td>
<td>1</td>
<td>43.3333</td>
<td>1964.667</td>
<td>1921.333</td>
<td>44</td>
</tr>
<tr>
<td>10</td>
<td>40</td>
<td>1963</td>
<td>1923</td>
<td>0</td>
<td>41.6667</td>
<td>1966.333</td>
<td>1924.667</td>
<td>38</td>
</tr>
</tbody>
</table>
```

APC-model for triangular data (APC-tri)
Fill in the number of cases ($D$) and person-years ($Y$) from previous slide.

Indicate birth cohorts on the axes for upper and lower triangles.

Mark mean date of birth for these.
Fill in the number of cases \((D)\) and person-years \((Y)\) from previous slide.

Indicate birth cohorts on the axes for upper and lower triangles.

Mark mean date of birth for these.

APC-model for triangular data (APC-tri)
APC-model with “synthetic” cohorts

```r
> mc <- glm( D ~ factor(A5) - 1 +
+          factor(P5-A5) +
+          factor(P5) + offset( log( Y ) ),
+          family=poisson )
> summary( mc )

... 

Null deviance: 1.0037e+08 on 220 degrees of freedom
Residual deviance: 8.8866e+02 on 182 degrees of freedom

No. parameters: 220 - 182 = 38.

A = 10, P = 11, C = 20  \Rightarrow  A + P + C - 3 = 38
```
APC-model with "correct" cohorts

\[
> \text{mx} <- \text{glm}( \ D \sim \text{factor(Ax)} - 1 + \\
+ \quad \text{factor(Cx)} + \\
+ \quad \text{factor(Px)} + \text{offset( log( Y ) )}, \\
+ \quad \text{family=poisson} ) \\
> \text{summary( mx )}
\]

... 

Null deviance: 1.0037e+08 on 220 degrees of freedom
Residual deviance: 2.8473e+02 on 144 degrees of freedom

No. parameters: \(220 - 144 = 76\) \((= 38 \times 2)\).

\[A = 20, \quad P = 22, \quad C = 40 \quad \Rightarrow \quad A + P + C - 3 = 79 \neq 76!\]

We have fitted two age-period-cohort models separately to upper and lower triangles.
APC-model for triangular data (APC-tri)
APC-model for triangular data (APC-tri)
Now, explicitly fit models for upper and lower triangles:

```r
> mx.u <- glm( D ~ factor(Ax) - 1 + 
> factor(Cx) + 
> factor(Px) + offset( log( Y/10^5 ) ), family=poisson, 
> data=lungDK[lungDK$up==1, ] 
> )
> mx.l <- glm( D ~ factor(Ax) - 1 + 
> factor(Cx) + 
> factor(Px) + offset( log( Y/10^5 ) ), family=poisson, 
> data=lungDK[lungDK$up==0, ] 
> )
> mx$deviance
[1] 284.7269
> mx.l$deviance
[1] 134.4566
> mx.u$deviance
[1] 150.2703
> mx.l$deviance+mx.u$deviance
[1] 284.7269

APC-model for triangular data (APC-tri)
```
APC-model for triangular data (APC-tri)
APC-model: Parametrization

Statistical Analysis in the Lexis Diagram:
Age-Period-Cohort models
May 2016
Max Planck Institut for Demographic Research, Rostock
http://BendixCarstensen/APC/MPIDR-2016
What’s the problem?

- One parameter is assigned to each distinct value of the timescales, the **scale** of the variables ia not used.
- The solution is to “tie together” the points on the scales together with smooth functions of the **mean** age, period and cohort with three functions:

\[
\lambda_{ap} = f(a) + g(p) + h(c)
\]

- The practical problem is how to choose a reasonable parametrization of these functions, and how to get estimates.
The identifiability problem still exists:

\[ c = p - a \iff p - a - c = 0 \]

\[
\lambda_{ap} = f(a) + g(p) + h(c)
= f(a) + g(p) + h(c) + \gamma(p - a - c)
= f(a) - \mu_a - \gamma a + \\
+ g(p) + \mu_a + \mu_c + \gamma p + \\
+ h(c) - \mu_c - \gamma c
\]

A decision on parametrization is needed. 
... it must be external to the model.
Smooth functions

\[ \log(\lambda(a, p)) = f(a) + g(p) + h(c) \]

Possible choices for parametric functions describing the effect of the three continuous variables:

- Polynomials / fractional polynomials.
- Linear / quadratic / cubic splines.
- Natural splines.

All of these contain the linear effect as special case.
**Parametrization of effects**

There are still three “free” parameters:

\[
\begin{align*}
\tilde{f}(a) &= f(a) - \mu_a - \gamma a \\
\tilde{g}(p) &= g(p) + \mu_a + \mu_c + \gamma p \\
\tilde{h}(c) &= h(c) - \mu_c - \gamma c
\end{align*}
\]

Any set of 3 numbers, \(\mu_a, \mu_c\) and \(\gamma\) will produce effects with the same sum. Choose \(\mu_a, \mu_c\) and \(\gamma\) according to some criterion for the functions.
Parametrization principle

1. The age-function should be interpretable as log age-specific rates in cohort $c_0$ after adjustment for the period effect.

2. The cohort function is 0 at a reference cohort $c_0$, interpretable as log-RR relative to cohort $c_0$.

3. The period function is 0 on average with 0 slope, interpretable as log-RR relative to the age-cohort prediction. (residual log-RR).

Longitudinal or cohort age-effects.

Biologically interpretable — what happens during the lifespan of a cohort?
Period-major parametrization

- Alternatively, the period function could be constrained to be 0 at a reference date, $p_0$.
- Then, age-effects at $a_0 = p_0 - c_0$ would equal the fitted rate for period $p_0$ (and cohort $c_0$), and the period effects would be residual log-RRs relative to $p_0$.
- Cross-sectional or period age-effects?
- Bureaucratically interpretable — what is seen at a particular date?

Might be wiser to look at predicted rates...
Implementation:

1. Obtain any set of parameters \( f(a) \), \( g(p) \), \( h(c) \).
2. Extract the trend from the period effect (find \( \mu \) and \( \beta \)):
   \[
   \tilde{g}(p) = \hat{g}(p) - (\mu + \beta p)
   \]
3. Decide on a reference cohort \( c_0 \).
4. Use the functions:
   \[
   \tilde{f}(a) = \hat{f}(a) + \mu + \beta a + \hat{h}(c_0) + \beta c_0
   
   \tilde{g}(p) = \hat{g}(p) - \mu - \beta p
   
   \tilde{h}(c) = \hat{h}(c) + \beta c - \hat{h}(c_0) - \beta c_0
   \]

These functions fulfill the criteria.
“Extract the trend”

- **Not** a well-defined concept:
  - Regress \( \hat{g}(p) \) on \( p \) for all units in the dataset.
  - Regress \( \hat{g}(p) \) on \( p \) for all different values of \( p \).
  - Weighted regression?

- How do we get the standard errors?
- Matrix-algebra!
- Projections!
Parametric function

Suppose that \( g(p) \) is parametrized using the design matrix \( M \), with the estimated parameters \( \pi \).

Example: 2nd order polynomial:

\[
M = \begin{bmatrix}
1 & p_1 & p_1^2 \\
1 & p_2 & p_2^2 \\
\vdots & \vdots & \vdots \\
1 & p_n & p_n^2
\end{bmatrix} \quad \pi = \begin{bmatrix}
\pi_0 \\
\pi_1 \\
\pi_2
\end{bmatrix} \quad g(p) = M\pi
\]

\( \text{nr}ow(M) \) is the no. of observations in the dataset,
\( \text{ncol}(M) \) is the no. of parameters
Extract the trend from $g$:

\[
\langle \tilde{g}(p)|1\rangle = 0, \quad \langle \tilde{g}(p)|p\rangle = 0
\]

i.e. $\tilde{g}$ is **orthogonal** to $[1|p]$.

Suppose $\tilde{g}(p) = \tilde{M}\pi$, then for any parameter vector $\pi$:

\[
\langle \tilde{M}\pi|1\rangle = 0, \quad \langle \tilde{M}\pi|p\rangle = 0 \implies \tilde{M} \perp [1|p]
\]

Thus we just need to be able to produce $\tilde{M}$ from $M$:
Projection on the orthogonal space of span$([1|p])$.

**NOTE:** Orthogonality requires an inner product!
Practical parametization

1. Set up model matrices for age, period and cohort, $M_a$, $M_p$ and $M_c$. Intercept in all three.

2. Extract the linear trend from $M_p$ and $M_c$, by projecting their columns onto the orthogonal complement of $[1|p]$ and $[1|c]$, respectively.

3. Center the cohort effect around $c_0$: Take a row from $\tilde{M}_c$ corresponding to $c_0$, replicate to dimension as $\tilde{M}_c$, and subtract it from $\tilde{M}_c$ to form $\tilde{M}_{c0}$. 
4. Use:
   \( M_a \) for the age-effects,
   \( \tilde{M}_p \) for the period effects and
   \( [c - c_0 | \tilde{M}_{c_0}] \) for the cohort effects.

5. Value of \( \hat{f}(a) \) is \( M_a \hat{\beta}_a \), similarly for the other two effects. Variance is found by \( M'_a \hat{\Sigma}_a M_a \), where \( \hat{\Sigma}_a \) is the variance-covariance matrix of \( \hat{\beta}_a \).
Information in the data and inner product

Log-lik for an observation \((D, Y)\), with log-rate \(\theta = \log(\lambda)\):

\[
l(\theta|D, Y) = D \theta - e^\theta Y, \quad l'_\theta = D - e^\theta Y, \quad l''_\theta = -e^\theta Y
\]

so \(I(\hat{\theta}) = e^{\hat{\theta}} Y = \hat{\lambda} Y = D\).

Log-lik for an observation \((D, Y)\), with rate \(\lambda\):

\[
l(\lambda|D, Y) = D \log(\lambda) - \lambda Y, \quad l'_\lambda = D/\lambda - Y, \quad l''_\lambda = -D/\lambda^2,
\]

so \(I(\hat{\lambda}) = D/\hat{\lambda}^2 = Y^2/D(= Y/\lambda)\)
Information in the data and inner product

- Two inner products:

\[ \langle m_j | m_k \rangle = \sum_i m_{ij} m_{ik} \quad \langle m_j | m_k \rangle = \sum_i m_{ij} w_i m_{ik} \]

- Weights could be chosen as:
  - \( w_i = D_i \), i.e. proportional to the information content for \( \theta \)
  - \( w_i = Y_i^2 / D_i \), i.e. proportional to the information content for \( \lambda \)
How to?

Implemented in `apc.fit` in the `Epi` package:

```r
> library( Epi )
> sessionInfo()

R version 3.2.5 (2016-04-14)
Platform: x86_64-pc-linux-gnu (64-bit)
Running under: Ubuntu 14.04.4 LTS

locale:
[1] LC_CTYPE=en_US.UTF-8 LC_NUMERIC=C LC_TIME=en_US.UTF-8
[7] LC_PAPER=en_US.UTF-8 LC_NAME=C LC_ADDRESS=C
[10] LC_TELEPHONE=C LC_MEASUREMENT=en_US.UTF-8 LC_IDENTIFICATION=C

attached base packages:
[1] utils     datasets    graphics   grDevices stats     methods   base
```

APC-model: Parametrization (APC-par)
How to? II

other attached packages:
[1] Epi_2.3

loaded via a namespace (and not attached):
[1] cmprsk_2.2-7 MASS_7.3-44 Matrix_1.2-1 plyr_1.8.3 parallel_3.2.5
[6] survival_2.39-2 etm_0.6-2 Rcpp_0.11.6 splines_3.2.5 grid_3.2.5
[11] numDeriv_2014.2-1 lattice_0.20-31

> library( splines )
> data( lungDK )
> mw <- apc.fit( A=lungDK$Ax,
+ P=lungDK$Px,
+ D=lungDK$D,
+ Y=lungDK$Y/10^5, dr.extr="w", npar=8,
+ ref.c=1900 )
NOTE: npar is specified as: A P C
8 8 8

[1] "ML of APC-model Poisson with log(Y) offset : ( ACP ):
"

Analysis of deviance for Age-Period-Cohort model

<table>
<thead>
<tr>
<th>Resid. Df</th>
<th>Resid. Dev</th>
<th>Df</th>
<th>Deviance</th>
<th>Pr(&gt;Chi)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>212</td>
<td>15468.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age-drift</td>
<td>211</td>
<td>6858.9</td>
<td>1</td>
<td>8609.7</td>
</tr>
<tr>
<td>Age-Cohort</td>
<td>205</td>
<td>1034.7</td>
<td>6</td>
<td>5824.1</td>
</tr>
<tr>
<td>Age-Period-Cohort</td>
<td>199</td>
<td>423.2</td>
<td>6</td>
<td>611.6</td>
</tr>
<tr>
<td>Age-Period</td>
<td>205</td>
<td>3082.6</td>
<td>-6</td>
<td>-2659.4</td>
</tr>
<tr>
<td>Age-drift</td>
<td>211</td>
<td>6858.9</td>
<td>-6</td>
<td>-3776.3</td>
</tr>
</tbody>
</table>

> plot(mw)

cp.offset  RR.fac
1765        100
How to? IV
Consult the help page for: `apc.fit` to see options for weights in inner product, type of function, variants of parametrization etc.
`apc.plot`, `apc.lines` and `apc.frame` to see how to plot the results.
Other models I

APC-model: Parametrization (APC-par)
> ml <- apc.fit( A=lungDK$Ax,
+   P=lungDK$Px,
+   D=lungDK$D,
+   Y=lungDK$Y/10^5, dr.extr="l", npar=8,
+   ref.c=1900 )

NOTE: npar is specified as: A P C
8 8 8
[1] "ML of APC-model Poisson with log(Y) offset : ( ACP ):

Analysis of deviance for Age-Period-Cohort model

<table>
<thead>
<tr>
<th>Resid. Df</th>
<th>Resid. Dev</th>
<th>Df</th>
<th>Deviance</th>
<th>Pr(&gt;Chi)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>212</td>
<td></td>
<td>15468.6</td>
<td></td>
</tr>
<tr>
<td>Age-drift</td>
<td>211</td>
<td>1</td>
<td>6858.9</td>
<td>&lt; 2.2e-16</td>
</tr>
<tr>
<td>Age-Cohort</td>
<td>205</td>
<td>6</td>
<td>1034.7</td>
<td>&lt; 2.2e-16</td>
</tr>
<tr>
<td>Age-Period-Cohort</td>
<td>199</td>
<td>6</td>
<td>423.2</td>
<td>&lt; 2.2e-16</td>
</tr>
<tr>
<td>Age-Period</td>
<td>205</td>
<td>-6</td>
<td>-3082.6</td>
<td>&lt; 2.2e-16</td>
</tr>
<tr>
<td>Age-drift</td>
<td>211</td>
<td>-6</td>
<td>6858.9</td>
<td>&lt; 2.2e-16</td>
</tr>
</tbody>
</table>

> m1 <- apc.fit( A=lungDK$Ax,
+   P=lungDK$Px,
+   D=lungDK$D,
+   APC-model: Parametrization (APC-par)
Y=lungDK$Y/10^5, dr.extr="1", npar=8, ref.c=1900 

NOTE: npar is specified as: A P C 
8 8 8 
[1] "ML of APC-model Poisson with log(Y) offset : ( ACP ):\n"

Analysis of deviance for Age-Period-Cohort model

<table>
<thead>
<tr>
<th>Resid. Df</th>
<th>Resid. Dev</th>
<th>Df</th>
<th>Deviance</th>
<th>Pr(&gt;Chi)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>212</td>
<td>1</td>
<td>8609.7</td>
<td>&lt; 2.2e-16</td>
</tr>
<tr>
<td>Age-drift</td>
<td>211</td>
<td>1</td>
<td>5824.1</td>
<td>&lt; 2.2e-16</td>
</tr>
<tr>
<td>Age-Cohort</td>
<td>205</td>
<td>6</td>
<td>611.6</td>
<td>&lt; 2.2e-16</td>
</tr>
<tr>
<td>Age-Period-Cohort</td>
<td>199</td>
<td>6</td>
<td>2659.4</td>
<td>&lt; 2.2e-16</td>
</tr>
<tr>
<td>Age-Period</td>
<td>205</td>
<td>-6</td>
<td>3776.3</td>
<td>&lt; 2.2e-16</td>
</tr>
<tr>
<td>Age-drift</td>
<td>211</td>
<td>-6</td>
<td>3776.3</td>
<td>&lt; 2.2e-16</td>
</tr>
</tbody>
</table>

> mw$Drift

exp(Est.)  2.5%  97.5%
APC (D-weights) 1.019662 1.019062 1.020263
A-d 1.023487 1.022971 1.024003

> ml$Drift

APC-model: Parametrization (APC-par)
exp(Est.)  2.5%   97.5%
APC (Y2/D-weights)  1.014869  1.013687  1.016053
A-d                  1.023487  1.022971  1.024003

> m1$Drift

exp(Est.)  2.5%   97.5%
APC (1-weights)  1.033027  1.032174  1.033879
A-d                  1.023487  1.022971  1.024003

> cnr <-
  + function( xf, yf )
  + {
    + cn <- par()$usr
    + xf <- ifelse( xf>1, xf/100, xf )
    + yf <- ifelse( yf>1, yf/100, yf )
    + if ( par()$xlog ) xx <- 10^xx
    + if ( par()$ylog ) yy <- 10^yy
    + list( x=xx, y=yy )
    + }

APC-model: Parametrization (APC-par)
APC-model: Parametrization (APC-par)

Weighted drift: 2.58 (2.42 - 2.74) %/year
APC-model: Parametrization (APC-par)
Lee-Carter model

Statistical Analysis in the Lexis Diagram:
Age-Period-Cohort models
May 2016
Max Planck Institut for Demographic Research, Rostock
http://BendixCarstensen/APC/MPIDR-2016
Lee-Carter model for (mortality) rates

\[ \log(\lambda_{x,t}) = a_x + b_x \times k_t \]

\( x \) is age; \( t \) is calendar time

- Formulated originally using as step-functions with one parameter per age/period.
- Implicitly assumes a data lay out by age and period: A, B or C-sets, but not Lexis triangles.
- Relative scaling of \( b_x \) and \( k_t \) cannot be determined
- \( k_t \) only determined up to an affine transformation:

\[ a_x + b_x(k_t + m) = (a_x + b_xm) = \tilde{a}_x + b_xk_t \]
Lee-Carter model in continuous time

\[
\log(\lambda(a, t)) = f(a) + b(a) \times k(t)
\]

- \( f(a), b(a) \) smooth functions of age:
  - \( a \) is a **scaled** variable.
- \( k(t) \) smooth function of period:
  - \( t \) is a **scaled** variable.
- Relative **scaling** of \( b(a) \) and \( k(t) \) cannot be determined
- \( k(t) \) only determined up to **affine** transformation:

\[
f(a) + b(a)(k(t) + m) = (f(a) + b(a)m) = \tilde{f}(a) + b(a)k(t)
\]
Lee-Carter model in continuous time

\[ \log(\lambda(a, t)) = f(a) + b(a) \times k(t) \]

- Lee-Carter model is an extension of the age-period model; if \( b(a) == 1 \) it is the age-period model.
- The extension is an age×period interaction, but not a traditional one:
  \[ \log(\lambda(a, t)) = f(a) + b(a) \times k(t) = f(a) + k(t) + (b(a) - 1) \times k(t) \]
- Main effect and interaction component of \( t \) are constrained to be identical.
- **NOTE**: the time variable, \( t \) could be either period, \( p \) or cohort, \( c = p - a \).
Main effect and interaction the same

Main effect and interaction component of $t$ are constrained to be identical.

None of these are Lee-Carter models:

```r
> glm( D ~ Ns(A, kn=a1.kn) + Ns(A, kn=a2.kn, i=T):Ns(P, kn=p.kn), ... )
> glm( D ~ Ns(A, kn=a1.kn) + Ns(A, kn=a2.kn, i=T)*Ns(P, kn=p.kn), ... )
> glm( D ~ Ns(A, kn=a1.kn) + Ns(P, kn=p.kn) + Ns(A, kn=a2.kn, i=T):Ns(P, kn=p.kn), ... 
```
Main effect and interaction the same

Main effect and interaction component of \( t \) are constrained to, \( i = T \) be identical.

An interaction between two spline terms is not the same as the product of two terms:

```r
> library( Epi )
> dfr <- data.frame( A=30:92, P=rep(1990:2010,3) )
> ( a.kn <- 4:8*10 ) ; ( p.kn <- c(1992+0:2*5) )
[1] 40 50 60 70 80

> mA <- with( dfr, model.matrix( ~ Ns(A,k=a.kn,i=T) -1 ) )
> mP <- with( dfr, model.matrix( ~ Ns(P,k=p.kn) ) )
> mAP <- with( dfr, model.matrix( ~ Ns(A,k=a.kn,i=T):Ns(P,k=p.kn) -1 ) )
> map <- with( dfr, model.matrix( ~ Ns(A,k=a.kn,i=T)*Ns(P,k=p.kn) -1 ) )
> cbind( colnames(mA) )
[1,] "Ns(A, k = a.kn, i = T)1"           "Ns(A, k = a.kn, i = T)2"           "Ns(A, k = a.kn, i = T)3"           "Ns(A, k = a.kn, i = T)4"           "Ns(A, k = a.kn, i = T)5"           "Ns(P, k = p.kn)1"           "Ns(P, k = p.kn)2"           "Ns(A, k = a.kn, i = T)1:Ns(P, k = p.kn)1"           "Ns(A, k = a.kn, i = T)2:Ns(P, k = p.kn)1"           "Ns(A, k = a.kn, i = T)3:Ns(P, k = p.kn)1"           "Ns(A, k = a.kn, i = T)4:Ns(P, k = p.kn)1"           "Ns(A, k = a.kn, i = T)5:Ns(P, k = p.kn)1"           "Ns(A, k = a.kn, i = T)1:Ns(P, k = p.kn)2"           "Ns(A, k = a.kn, i = T)2:Ns(P, k = p.kn)2"           "Ns(A, k = a.kn, i = T)3:Ns(P, k = p.kn)2"           "Ns(A, k = a.kn, i = T)4:Ns(P, k = p.kn)2"           "Ns(A, k = a.kn, i = T)5:Ns(P, k = p.kn)2"

> cbind( colnames(mP) )
[1,] "(Intercept)"           "Ns(P, k = p.kn)1"           "Ns(P, k = p.kn)2"

> cbind( colnames(mAP) )
[1,] "Ns(A, k = a.kn, i = T)1:Ns(P, k = p.kn)1"           "Ns(A, k = a.kn, i = T)2:Ns(P, k = p.kn)1"           "Ns(A, k = a.kn, i = T)3:Ns(P, k = p.kn)1"           "Ns(A, k = a.kn, i = T)4:Ns(P, k = p.kn)1"           "Ns(A, k = a.kn, i = T)5:Ns(P, k = p.kn)1"           "Ns(A, k = a.kn, i = T)1:Ns(P, k = p.kn)2"           "Ns(A, k = a.kn, i = T)2:Ns(P, k = p.kn)2"           "Ns(A, k = a.kn, i = T)3:Ns(P, k = p.kn)2"           "Ns(A, k = a.kn, i = T)4:Ns(P, k = p.kn)2"           "Ns(A, k = a.kn, i = T)5:Ns(P, k = p.kn)2"

> cbind( colnames(map) )
[1,] "Ns(A, k = a.kn, i = T)1"           "Ns(A, k = a.kn, i = T)2"           "Ns(A, k = a.kn, i = T)3"           "Ns(A, k = a.kn, i = T)4"           "Ns(A, k = a.kn, i = T)5"           "Ns(P, k = p.kn)1"           "Ns(P, k = p.kn)2"           "Ns(A, k = a.kn, i = T)1:Ns(P, k = p.kn)1"           "Ns(A, k = a.kn, i = T)2:Ns(P, k = p.kn)1"           "Ns(A, k = a.kn, i = T)3:Ns(P, k = p.kn)1"           "Ns(A, k = a.kn, i = T)4:Ns(P, k = p.kn)1"           "Ns(A, k = a.kn, i = T)5:Ns(P, k = p.kn)1"           "Ns(A, k = a.kn, i = T)1:Ns(P, k = p.kn)2"           "Ns(A, k = a.kn, i = T)2:Ns(P, k = p.kn)2"           "Ns(A, k = a.kn, i = T)3:Ns(P, k = p.kn)2"           "Ns(A, k = a.kn, i = T)4:Ns(P, k = p.kn)2"           "Ns(A, k = a.kn, i = T)5:Ns(P, k = p.kn)2"           "Ns(A, k = a.kn, i = T)1:Ns(P, k = p.kn)3"           "Ns(A, k = a.kn, i = T)2:Ns(P, k = p.kn)3"           "Ns(A, k = a.kn, i = T)3:Ns(P, k = p.kn)3"           "Ns(A, k = a.kn, i = T)4:Ns(P, k = p.kn)3"           "Ns(A, k = a.kn, i = T)5:Ns(P, k = p.kn)3"

Lee-Carter model (LeeCarter)
Lee-Carter model interpretation

\[
\log(\lambda(a, p)) = f(a) + b(a) \times k(p)
\]

- **Constraints:**
  - \(f(a)\) is the basic age-specific mortality
  - \(k(p)\) is the rate-ratio (RR) as a function of \(p\):
    - relative to \(p_{\text{ref}}\) where \(k(p_{\text{ref}}) = 1\)
    - for persons aged \(a_{\text{ref}}\) where \(b(a_{\text{ref}}) = 0\)
  - \(b(a)\) is an age-specific multiplier for the RR
  - Choose \(p_{\text{ref}}\) and \(a_{\text{ref}}\) *a priori*. 

Lee-Carter model (LeeCarter)
Danish lung cancer data I

```r
> lung <- read.table( "../data/apc-Lung.txt", header=T )
> head( lung )

    sex A  P  C  D  Y
 1    1 1 0 1943 1942 0 19546.2
 2    1 1 0 1943 1943 0 20796.5
 3    1 1 0 1944 1943 0 20681.3
 4    1 1 0 1944 1944 0 22478.5
 5    1 1 0 1945 1944 0 22369.2
 6    1 1 0 1945 1945 0 23885.0

> # Only A by P classification - and only ages over 40
> ltab <- xtabs( cbind(D,Y) ~ A + P, data=subset(lung,sex==1) )
> str( ltab )
```

Lee-Carter model (LeeCarter)
Danish lung cancer data II

xtabs [1:90, 1:61, 1:2] 0 0 0 0 0 0 0 0 0 0 ...
- attr(*, "dimnames")=List of 3
  ..$ A: chr [1:90] "0" "1" "2" "3" ...
  ..$ P: chr [1:61] "1943" "1944" "1945" "1946" ...
  ..$ : chr [1:2] "D" "Y"
- attr(*, "class")= chr [1:2] "xtabs" "table"
- attr(*, "call")= language xtabs(formula = cbind(D, Y) ~ A + P, data = subset(lung, sex == 1))
Lee-Carter with demography

```r
> library(demography)
> lcM <- demogdata( data = as.matrix(ltab[40:90,,"D"]/ltab[40:90,,"Y"]),
+                  pop = as.matrix(ltab[40:90,,"Y"]),
+                  ages = as.numeric(dimnames(ltab)[[1]][40:90]),
+                  years = as.numeric(dimnames(ltab)[[2]]),
+                  type = "Lung cancer incidence",
+                  label = "Denmark",
+                  name = "Male"
> str(lcM)
```

Lee-Carter model (LeeCarter)
Lee-Carter with demography II

List of 7
$ year : num [1:61] 1943 1944 1945 1946 1947 ...
$ age : num [1:51] 39 40 41 42 43 44 45 46 47 48 ...
$ rate :List of 1
  ..$ Male: num [1:51, 1:61] 1.05e-04 7.10e-05 7.31e-05 3.73e-05 2.30e-04 ...
  ...- attr(*, "dimnames")=List of 2
  ... ..$ : chr [1:51] "39" "40" "41" "42" ...
  ... ..$ : chr [1:61] "1943" "1944" "1945" "1946" ...
$ pop :List of 1
  ..$ Male: num [1:51, 1:61] 28488 28152 27363 26791 26092 ...
  ...- attr(*, "dimnames")=List of 2
  ... ..$ : chr [1:51] "39" "40" "41" "42" ...
  ... ..$ : chr [1:61] "1943" "1944" "1945" "1946" ...
$ type : chr "Lung cancer incidence"
$ label : chr "Denmark"
$ lambda: num 1
- attr(*, "class")= chr "demogdata"
Lee-Carter with demography III
lca estimation function checks the type argument, so we make a workaround:

```r
> mrt <- function(x) { x$type <- "mortality" ; x }
> dmg.lcM <- lca(mrt(lcM), interpolate=TRUE )
> par( mfcol=c(2,2) )
> matplot( dmg.lcM$age, exp(dmg.lcM$ax)*1000,
+        log="y", ylab="Lung cancer incidence rates per 1000 PY",
+        xlab="Age", type="l", lty=1, lwd=4 )
> plot( NA, xlim=0:1, ylim=0:1, axes=FALSE, xlab="", ylab="" )
> matplot( dmg.lcM$age, dmg.lcM$bx,
+        ylab="Age effect",
+        xlab="Age", type="l", lty=1, lwd=4 )
> matplot( dmg.lcM$year, dmg.lcM$kt,
+        ylab="Time effect",
+        xlab="Date", type="l", lty=1, lwd=4 )
> abline(h=0)
```
Lee-Carter with demography
Lee-Carter re-scaled I

> par( mfcol=c(2,2) )
> matplot( dmg.lcM$age, exp(dmg.lcM$ax+dmg.lcM$bx*20)*1000, 
+       log="y", ylab="Lung cancer incidence rates per 1000 PY", 
+       xlab="Age", type="l", lty=1, lwd=4 )
> plot( NA, xlim=0:1, ylim=0:1, axes=FALSE, xlab="", ylab="" )
> matplot( dmg.lcM$age, dmg.lcM$bx*50, 
+       ylab="Age effect", 
+       xlab="Age", type="l", lty=1, lwd=4 )
> abline(h=1)
> matplot( dmg.lcM$year, (dmg.lcM$kt-20)/50, 
+       ylab="Time effect", 
+       xlab="Date", type="l", lty=1, lwd=4 )
> abline(h=0)
Lee-Carter with demography rescaled
Lee-Carter with \texttt{ilc}

- The \texttt{lca.rh} function fits the model using maximum likelihood (proportional scaling)
- Fits the more general model and submodels of it:

\[
\log(\lambda(a, p)) = f(a) + b(a) \times k(p) + c(a)m(p - a)
\]

- Age interaction with betewwn age and both period and/or cohort (=period-age)
- Extension of APC-model:
  \[b(a) = 1 \text{ and } a(a) = 1 \Leftrightarrow \text{APC model.}\]
Lee-Carter with \texttt{ilc}

> library( ilc )
> ilc.lcM <- lca.rh( mrt(lcM), model="lc", interpolate=TRUE )

Original sample: Mortality data for Denmark
   Series: Male
   Years: 1943 - 2003
   Ages: 39 - 89

Applied sample: Mortality data for Denmark (Corrected: interpolate)
   Series: Male
   Years: 1943 - 2003
   Ages: 39 - 89

Fitting model: \[ LC = a(x)+b1(x)*k(t) \] - with poisson error structure and with deaths as weights -
Note: 0 cells have 0/NA deaths and 0 have 0/NA exposure out of a total of 3111 data cells.
## Lee-Carter with \textit{ilc II}

Starting values are:

<table>
<thead>
<tr>
<th>per</th>
<th>per.c</th>
<th>age</th>
<th>age.c</th>
<th>bx1.c</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1943</td>
<td>39</td>
<td>-9.687</td>
<td>0.02</td>
</tr>
<tr>
<td>2</td>
<td>1944</td>
<td>40</td>
<td>-9.487</td>
<td>0.02</td>
</tr>
<tr>
<td>3</td>
<td>1945</td>
<td>41</td>
<td>-9.408</td>
<td>0.02</td>
</tr>
<tr>
<td>4</td>
<td>1946</td>
<td>42</td>
<td>-9.151</td>
<td>0.02</td>
</tr>
<tr>
<td>5</td>
<td>1947</td>
<td>43</td>
<td>-8.929</td>
<td>0.02</td>
</tr>
<tr>
<td>6</td>
<td>1948</td>
<td>44</td>
<td>-8.73</td>
<td>0.02</td>
</tr>
<tr>
<td>7</td>
<td>1949</td>
<td>45</td>
<td>-8.475</td>
<td>0.02</td>
</tr>
<tr>
<td>8</td>
<td>1950</td>
<td>46</td>
<td>-8.426</td>
<td>0.02</td>
</tr>
<tr>
<td>9</td>
<td>1951</td>
<td>47</td>
<td>-8.145</td>
<td>0.02</td>
</tr>
<tr>
<td>10</td>
<td>1952</td>
<td>48</td>
<td>-7.991</td>
<td>0.02</td>
</tr>
<tr>
<td>11</td>
<td>1953</td>
<td>49</td>
<td>-7.808</td>
<td>0.02</td>
</tr>
<tr>
<td>12</td>
<td>1954</td>
<td>50</td>
<td>-7.549</td>
<td>0.02</td>
</tr>
<tr>
<td>13</td>
<td>1955</td>
<td>51</td>
<td>-7.473</td>
<td>0.02</td>
</tr>
<tr>
<td>14</td>
<td>1956</td>
<td>52</td>
<td>-7.376</td>
<td>0.02</td>
</tr>
<tr>
<td>15</td>
<td>1957</td>
<td>53</td>
<td>-7.199</td>
<td>0.02</td>
</tr>
<tr>
<td>16</td>
<td>1958</td>
<td>54</td>
<td>-7.032</td>
<td>0.02</td>
</tr>
<tr>
<td>17</td>
<td>1959</td>
<td>55</td>
<td>-6.893</td>
<td>0.02</td>
</tr>
</tbody>
</table>
Lee-Carter with \textit{ilc} III

<table>
<thead>
<tr>
<th>Year</th>
<th>Age</th>
<th>Value</th>
<th>Value 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>18</td>
<td>1960</td>
<td>56</td>
<td>-6.798</td>
</tr>
<tr>
<td>19</td>
<td>1961</td>
<td>57</td>
<td>-6.698</td>
</tr>
<tr>
<td>20</td>
<td>1962</td>
<td>58</td>
<td>-6.596</td>
</tr>
<tr>
<td>21</td>
<td>1963</td>
<td>59</td>
<td>-6.524</td>
</tr>
<tr>
<td>22</td>
<td>1964</td>
<td>60</td>
<td>-6.463</td>
</tr>
<tr>
<td>23</td>
<td>1965</td>
<td>61</td>
<td>-6.325</td>
</tr>
<tr>
<td>24</td>
<td>1966</td>
<td>62</td>
<td>-6.271</td>
</tr>
<tr>
<td>25</td>
<td>1967</td>
<td>63</td>
<td>-6.25</td>
</tr>
<tr>
<td>26</td>
<td>1968</td>
<td>64</td>
<td>-6.194</td>
</tr>
<tr>
<td>27</td>
<td>1969</td>
<td>65</td>
<td>-6.171</td>
</tr>
<tr>
<td>28</td>
<td>1970</td>
<td>66</td>
<td>-6.056</td>
</tr>
<tr>
<td>29</td>
<td>1971</td>
<td>67</td>
<td>-6.113</td>
</tr>
<tr>
<td>30</td>
<td>1972</td>
<td>68</td>
<td>-6.021</td>
</tr>
<tr>
<td>31</td>
<td>1973</td>
<td>69</td>
<td>-6.039</td>
</tr>
<tr>
<td>32</td>
<td>1974</td>
<td>70</td>
<td>-5.993</td>
</tr>
<tr>
<td>33</td>
<td>1975</td>
<td>71</td>
<td>-5.98</td>
</tr>
<tr>
<td>34</td>
<td>1976</td>
<td>72</td>
<td>-5.951</td>
</tr>
<tr>
<td>35</td>
<td>1977</td>
<td>73</td>
<td>-5.905</td>
</tr>
<tr>
<td>36</td>
<td>1978</td>
<td>74</td>
<td>-5.969</td>
</tr>
</tbody>
</table>

Lee-Carter model (LeeCarter)
### Lee-Carter with ilc IV

<table>
<thead>
<tr>
<th>Year</th>
<th>Age</th>
<th>Sex</th>
<th>Mortality Rate</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>37</td>
<td>1979</td>
<td>0</td>
<td>-6.008</td>
<td>0.02</td>
</tr>
<tr>
<td>38</td>
<td>1980</td>
<td>0</td>
<td>-6.044</td>
<td>0.02</td>
</tr>
<tr>
<td>39</td>
<td>1981</td>
<td>0</td>
<td>-5.998</td>
<td>0.02</td>
</tr>
<tr>
<td>40</td>
<td>1982</td>
<td>0</td>
<td>-6.029</td>
<td>0.02</td>
</tr>
<tr>
<td>41</td>
<td>1983</td>
<td>0</td>
<td>-6.146</td>
<td>0.02</td>
</tr>
<tr>
<td>42</td>
<td>1984</td>
<td>0</td>
<td>-6.1</td>
<td>0.02</td>
</tr>
<tr>
<td>43</td>
<td>1985</td>
<td>0</td>
<td>-6.118</td>
<td>0.02</td>
</tr>
<tr>
<td>44</td>
<td>1986</td>
<td>0</td>
<td>-6.025</td>
<td>0.02</td>
</tr>
<tr>
<td>45</td>
<td>1987</td>
<td>0</td>
<td>-6.247</td>
<td>0.02</td>
</tr>
<tr>
<td>46</td>
<td>1988</td>
<td>0</td>
<td>-6.111</td>
<td>0.02</td>
</tr>
<tr>
<td>47</td>
<td>1989</td>
<td>0</td>
<td>-6.177</td>
<td>0.02</td>
</tr>
<tr>
<td>48</td>
<td>1990</td>
<td>0</td>
<td>-6.281</td>
<td>0.02</td>
</tr>
<tr>
<td>49</td>
<td>1991</td>
<td>0</td>
<td>-6.305</td>
<td>0.02</td>
</tr>
<tr>
<td>50</td>
<td>1992</td>
<td>0</td>
<td>-6.213</td>
<td>0.02</td>
</tr>
<tr>
<td>51</td>
<td>1993</td>
<td>0</td>
<td>-6.638</td>
<td>0.02</td>
</tr>
<tr>
<td>52</td>
<td>1994</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>53</td>
<td>1995</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>54</td>
<td>1996</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>55</td>
<td>1997</td>
<td>0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Lee-Carter with ilc V

<table>
<thead>
<tr>
<th>Year</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>56</td>
<td>1998</td>
</tr>
<tr>
<td>57</td>
<td>1999</td>
</tr>
<tr>
<td>58</td>
<td>2000</td>
</tr>
<tr>
<td>59</td>
<td>2001</td>
</tr>
<tr>
<td>60</td>
<td>2002</td>
</tr>
<tr>
<td>61</td>
<td>2003</td>
</tr>
</tbody>
</table>

Iterative fit:

<table>
<thead>
<tr>
<th>#iter</th>
<th>Dev</th>
<th>non-conv</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>26123.55</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>9403.337</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>5219.715</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>4269.859</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>3982.804</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>3878.544</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>3836.703</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>3818.834</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>3810.839</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>3807.133</td>
<td>0</td>
</tr>
</tbody>
</table>
## Lee-Carter with ilc VI

<table>
<thead>
<tr>
<th>Year</th>
<th>Value</th>
<th>Shift</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>3805.368</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>3804.512</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>3804.089</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>3803.878</td>
<td>0</td>
</tr>
<tr>
<td>15</td>
<td>3803.772</td>
<td>0</td>
</tr>
<tr>
<td>16</td>
<td>3803.718</td>
<td>0</td>
</tr>
<tr>
<td>17</td>
<td>3803.69</td>
<td>0</td>
</tr>
<tr>
<td>18</td>
<td>3803.676</td>
<td>0</td>
</tr>
<tr>
<td>19</td>
<td>3803.669</td>
<td>0</td>
</tr>
<tr>
<td>20</td>
<td>3803.665</td>
<td>0</td>
</tr>
<tr>
<td>21</td>
<td>3803.663</td>
<td>0</td>
</tr>
<tr>
<td>22</td>
<td>3803.662</td>
<td>0</td>
</tr>
<tr>
<td>23</td>
<td>3803.661</td>
<td>0</td>
</tr>
<tr>
<td>24</td>
<td>3803.661</td>
<td>0</td>
</tr>
<tr>
<td>25</td>
<td>3803.661</td>
<td>0</td>
</tr>
<tr>
<td>26</td>
<td>3803.661</td>
<td>0</td>
</tr>
<tr>
<td>27</td>
<td>3803.66</td>
<td>0</td>
</tr>
<tr>
<td>28</td>
<td>3803.66</td>
<td>0</td>
</tr>
<tr>
<td>29</td>
<td>3803.66</td>
<td>0</td>
</tr>
</tbody>
</table>

**Lee-Carter model (LeeCarter)**
Lee-Carter with ilc VII

30  3803.66  0
31  3803.66  0
32  3803.66  0
33  3803.66  0
34  3803.66  0

Iterations finished in: 34 steps

Updated values are:

<table>
<thead>
<tr>
<th>per</th>
<th>per.c</th>
<th>age</th>
<th>age.c</th>
<th>bx1.c</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1943</td>
<td>39</td>
<td>-9.54531</td>
<td>0.0019</td>
</tr>
<tr>
<td>2</td>
<td>1944</td>
<td>40</td>
<td>-9.34555</td>
<td>0.00613</td>
</tr>
<tr>
<td>3</td>
<td>1945</td>
<td>41</td>
<td>-9.27014</td>
<td>0.00171</td>
</tr>
<tr>
<td>4</td>
<td>1946</td>
<td>42</td>
<td>-9.03109</td>
<td>0.00174</td>
</tr>
<tr>
<td>5</td>
<td>1947</td>
<td>43</td>
<td>-8.79572</td>
<td>0.00036</td>
</tr>
<tr>
<td>6</td>
<td>1948</td>
<td>44</td>
<td>-8.64242</td>
<td>0.00348</td>
</tr>
<tr>
<td>7</td>
<td>1949</td>
<td>45</td>
<td>-8.4011</td>
<td>0.00422</td>
</tr>
<tr>
<td>8</td>
<td>1950</td>
<td>46</td>
<td>-8.35569</td>
<td>0.00618</td>
</tr>
<tr>
<td>9</td>
<td>1951</td>
<td>47</td>
<td>-8.08493</td>
<td>0.00431</td>
</tr>
</tbody>
</table>

Lee-Carter model (LeeCarter)
## Lee-Carter with ilc VIII

<table>
<thead>
<tr>
<th>Year</th>
<th>Age</th>
<th>Value 1</th>
<th>Age</th>
<th>Value 2</th>
<th>Age</th>
<th>Value 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>1952</td>
<td>-28.25000</td>
<td>48</td>
<td>-7.95317</td>
<td>0.00269</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>1953</td>
<td>-33.56753</td>
<td>49</td>
<td>-7.75764</td>
<td>0.00692</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>1954</td>
<td>-28.16299</td>
<td>50</td>
<td>-7.52418</td>
<td>0.00338</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>1955</td>
<td>-25.93964</td>
<td>51</td>
<td>-7.44269</td>
<td>0.00752</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>1956</td>
<td>-21.26733</td>
<td>52</td>
<td>-7.33407</td>
<td>0.01031</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>1957</td>
<td>-17.95370</td>
<td>53</td>
<td>-7.16891</td>
<td>0.00774</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>1958</td>
<td>-16.32569</td>
<td>54</td>
<td>-7.00417</td>
<td>0.00789</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>1959</td>
<td>-7.92142</td>
<td>55</td>
<td>-6.87498</td>
<td>0.00862</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>1960</td>
<td>-9.67085</td>
<td>56</td>
<td>-6.76735</td>
<td>0.01002</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>1961</td>
<td>-5.13527</td>
<td>57</td>
<td>-6.67977</td>
<td>0.01128</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>1962</td>
<td>-4.23977</td>
<td>58</td>
<td>-6.57225</td>
<td>0.01469</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>1963</td>
<td>-1.90709</td>
<td>59</td>
<td>-6.49916</td>
<td>0.013</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>1964</td>
<td>-0.65036</td>
<td>60</td>
<td>-6.4307</td>
<td>0.0152</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>1965</td>
<td>3.31265</td>
<td>61</td>
<td>-6.30139</td>
<td>0.0168</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>1966</td>
<td>4.51564</td>
<td>62</td>
<td>-6.247</td>
<td>0.0184</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>1967</td>
<td>7.16008</td>
<td>63</td>
<td>-6.20883</td>
<td>0.01935</td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>1968</td>
<td>10.36382</td>
<td>64</td>
<td>-6.16206</td>
<td>0.02197</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>1969</td>
<td>10.60063</td>
<td>65</td>
<td>-6.11728</td>
<td>0.02439</td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>1970</td>
<td>12.25461</td>
<td>66</td>
<td>-6.03717</td>
<td>0.02497</td>
<td></td>
</tr>
</tbody>
</table>

Lee-Carter model (LeeCarter)
<table>
<thead>
<tr>
<th>Year</th>
<th>Year</th>
<th>Observation</th>
<th>t</th>
<th>c 1</th>
<th>c 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>29</td>
<td>1971</td>
<td>14.63642</td>
<td>67</td>
<td>-6.08387</td>
<td>0.028</td>
</tr>
<tr>
<td>30</td>
<td>1972</td>
<td>16.05776</td>
<td>68</td>
<td>-5.99082</td>
<td>0.02718</td>
</tr>
<tr>
<td>31</td>
<td>1973</td>
<td>15.53593</td>
<td>69</td>
<td>-6.00028</td>
<td>0.02854</td>
</tr>
<tr>
<td>32</td>
<td>1974</td>
<td>17.21334</td>
<td>70</td>
<td>-5.96719</td>
<td>0.02994</td>
</tr>
<tr>
<td>33</td>
<td>1975</td>
<td>17.80268</td>
<td>71</td>
<td>-5.95329</td>
<td>0.03323</td>
</tr>
<tr>
<td>34</td>
<td>1976</td>
<td>18.44457</td>
<td>72</td>
<td>-5.9555</td>
<td>0.03255</td>
</tr>
<tr>
<td>35</td>
<td>1977</td>
<td>18.71973</td>
<td>73</td>
<td>-5.9058</td>
<td>0.03205</td>
</tr>
<tr>
<td>36</td>
<td>1978</td>
<td>20.06082</td>
<td>74</td>
<td>-5.97665</td>
<td>0.03762</td>
</tr>
<tr>
<td>37</td>
<td>1979</td>
<td>20.31816</td>
<td>75</td>
<td>-6.01915</td>
<td>0.03916</td>
</tr>
<tr>
<td>38</td>
<td>1980</td>
<td>20.87884</td>
<td>76</td>
<td>-6.02213</td>
<td>0.03915</td>
</tr>
<tr>
<td>39</td>
<td>1981</td>
<td>21.61232</td>
<td>77</td>
<td>-5.99743</td>
<td>0.03704</td>
</tr>
<tr>
<td>40</td>
<td>1982</td>
<td>21.85089</td>
<td>78</td>
<td>-6.03741</td>
<td>0.03809</td>
</tr>
<tr>
<td>41</td>
<td>1983</td>
<td>22.96473</td>
<td>79</td>
<td>-6.12152</td>
<td>0.0405</td>
</tr>
<tr>
<td>42</td>
<td>1984</td>
<td>21.50736</td>
<td>80</td>
<td>-6.08339</td>
<td>0.03654</td>
</tr>
<tr>
<td>43</td>
<td>1985</td>
<td>23.22937</td>
<td>81</td>
<td>-6.12649</td>
<td>0.035</td>
</tr>
<tr>
<td>44</td>
<td>1986</td>
<td>20.20563</td>
<td>82</td>
<td>-6.00846</td>
<td>0.02978</td>
</tr>
<tr>
<td>45</td>
<td>1987</td>
<td>21.53699</td>
<td>83</td>
<td>-6.2544</td>
<td>0.04013</td>
</tr>
<tr>
<td>46</td>
<td>1988</td>
<td>20.54046</td>
<td>84</td>
<td>-6.08511</td>
<td>0.03306</td>
</tr>
<tr>
<td>47</td>
<td>1989</td>
<td>19.63340</td>
<td>85</td>
<td>-6.11129</td>
<td>0.02548</td>
</tr>
</tbody>
</table>
### Lee-Carter with ilc X

<table>
<thead>
<tr>
<th>Year</th>
<th>Year</th>
<th>Value 1</th>
<th>Value 2</th>
<th>Value 3</th>
<th>Value 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>48</td>
<td>1990</td>
<td>17.48203</td>
<td>86</td>
<td>-6.24171</td>
<td>0.02873</td>
</tr>
<tr>
<td>49</td>
<td>1991</td>
<td>17.31414</td>
<td>87</td>
<td>-6.24948</td>
<td>0.02586</td>
</tr>
<tr>
<td>50</td>
<td>1992</td>
<td>18.04416</td>
<td>88</td>
<td>-6.11791</td>
<td>0.01529</td>
</tr>
<tr>
<td>51</td>
<td>1993</td>
<td>17.91747</td>
<td>89</td>
<td>-6.51232</td>
<td>0.01146</td>
</tr>
<tr>
<td>52</td>
<td>1994</td>
<td>18.39041</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>53</td>
<td>1995</td>
<td>17.32639</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>54</td>
<td>1996</td>
<td>15.72621</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>55</td>
<td>1997</td>
<td>16.81425</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>56</td>
<td>1998</td>
<td>15.71813</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>57</td>
<td>1999</td>
<td>15.95432</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>58</td>
<td>2000</td>
<td>17.93764</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>59</td>
<td>2001</td>
<td>16.86795</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>2002</td>
<td>15.63661</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>61</td>
<td>2003</td>
<td>11.11935</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Total sums are:**

<table>
<thead>
<tr>
<th>b0</th>
<th>b1</th>
<th>itx</th>
<th>kt</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

```r
> plot(ilc.lcM)
```
Lee-Carter with \texttt{ilc}

Age-Period LC Regression for Denmark [Male]

Main age effects

Period Interaction effects

Cohort Interaction effects

Period effects

Cohort effects

Lee-Carter model (\texttt{LeeCarter})
Lee-Carter with Epi

- LCa.fit fits the Lee-Carter model using natural splines for the **quantitative** effects of age and time.
- Normalizes effects to a reference age and period.
- The algorithm alternately fits a main age and period effects and the age-interaction effect.
Lee-Carter model using natural splines:
\[
\log(\text{Rate}) = a(\text{Age}) + b(\text{Age})k(\text{Period})
\]
with 6, 5 and 6 parameters respectively (1 aliased).
Deviance: 8566.554 on 6084 d.f.

\[
> \text{plot( LCa.Mlc, rnam="Lung cancer incidence per 1000 PY" )}
\]
Lee-Carter model (LeeCarter)
Lee-Carter and the APC-model

- Lee-Carter model is an interaction extension of the Age-Period model
- ...or an interaction extension of the Age-Cohort model
- Age-Period-Cohort model is:
  - interaction extension
  - the smallest union of Age-Period and Age-Cohort
- Extended Lee-Carter (from the ilc package)
  \[
  \log(\lambda(a, p)) = f(a) + b(a) \times k(p) + c(a)m(p - a)
  \]
is the union of all of these.
Lee-Carter and the APC-model
Fit L-Ca models in Epi I

> LCa.P <- LCa.fit( Mlc, ref.b=60, ref.t=1980 )

LCa.fit convergence in 11 iterations, deviance: 8566.554 on 6084 d.f.

> LCa.C <- LCa.fit( Mlc, ref.b=60, ref.t=1980, model="C", maxit=200, eps=10e-4 )

LCa.fit convergence in 95 iterations, deviance: 8125.318 on 6084 d.f.

> ( a.kn <- LCa.P$a.kn )

8.333333% 25% 41.66667% 58.33333% 75% 91.66667%
53 60 65 69 74 80

> LCa.C$a.kn

8.333333% 25% 41.66667% 58.33333% 75% 91.66667%
53 60 65 69 74 80

Lee-Carter model (LeeCarter)
Fit L-Ca models in Epi II

> ( p.kn <- LCa.P$t.kn )

8.333333%  25%  41.66667%  58.33333%  75%  91.66667%

> ( c.kn <- LCa.C$t.kn )

8.333333%  25%  41.66667%  58.33333%  75%  91.66667%
1893  1904  1911  1918  1925  1935

> AP <- glm( D ~ Ns(A,knots=a.kn)+Ns(P,knots=p.kn),
> offset=log(Y), family=poisson, data=Mlc )
> AC <- glm( D ~ Ns(A,knots=a.kn)+ Ns(P-A,knots=c.kn),
> offset=log(Y), family=poisson, data=Mlc )
> APC <- glm( D ~ Ns(A,knots=a.kn)+Ns(P,knots=p.kn)+Ns(P-A,knots=c.kn),
> offset=log(Y), family=poisson, data=Mlc )
> c( AP$deviance, AP$df.res )

[1] 11010.88 6089.00
Fit L-Ca models in Epi III

```r
> c( AC$deviance, AC$df.res )
[1] 8583.249 6089.000

> c( APC$deviance, APC$df.res )
[1] 7790.446 6085.000

> c( LCa.P$dev, LCa.P$df )
[1] 8566.554 6084.000

> c( LCa.C$dev, LCa.C$df )
[1] 8125.318 6084.000
```
Fit L-Ca models in Epi IV

A–P
11010.9 / 6089

A–C
8583.2 / 6089

APC
7790.4 / 6085

LCa–P
8566.6 / 6084

LCa–C
8125.3 / 6084

Lee-Carter model (LeeCarter)
APC-models for several datasets

Statistical Analysis in the Lexis Diagram:
Age-Period-Cohort models
May 2016
Max Planck Institut for Demographic Research, Rostock
http://BendixCarstensen/APC/MPIDR-2016
Two APC-models

- APC-models for two sets of rates (men/women, types of events):

\[
\log(\lambda_i(a, p)) = f_i(a) + g_i(p) + h_i(p - a), \quad i = 1, 2
\]

- Rate-ratio also an APC-model:

\[
\log(\text{RR}(a, p)) = \log(\lambda_1(a, p)) - \log(\lambda_2(a, p))
\]
\[
= (f_1(a) - f_2(a)) + (g_1(p) - g_2(p))
\]
\[
+ (h_1(p - a) - h_2(p - a))
\]
\[
= f_{\text{RR}}(a) + g_{\text{RR}}(p) + h_{\text{RR}}(p - a)
\]

- Modeled separately and the ratio effects reported as any other APC-model.
Two sets of data I

Example: Testis cancer in Denmark, Seminoma and non-Seminoma cases.

```r
> th <- read.table( "../data/testis-hist.txt", header=TRUE )
> str( th )

'data.frame': 29160 obs. of 9 variables:
$ a : int 0 0 0 0 0 0 1 1 1 1 ... 
$ p : int 1943 1943 1943 1943 1943 1943 1943 1943 1943 1943 ... 
$ c : int 1942 1942 1942 1943 1943 1943 1941 1941 1941 1942 ... 
$ y : num 18853 18853 18853 20796 20796 ... 
$ age : num 0.667 0.667 0.667 0.333 0.333 ... 
$ diag : num 1943 1943 1943 1944 1944 ... 
$ birth: num 1943 1943 1943 1944 1944 ... 
$ hist : int 1 2 3 1 2 3 1 2 3 1 ... 
$ d : int 0 1 0 0 0 0 0 0 0 0 ... 
```

APC-models for several datasets (APC2)
Two sets of data II

> head(th)

       a  p   c     y age     diag  birth   hist  d
  1 0 1943 1942 18853.0 1943.333 1942.667 1 0
  2 0 1943 1942 18853.0 1943.333 1942.667 2 1
  3 0 1943 1942 18853.0 1943.333 1942.667 3 0
  4 0 1943 1943 20796.5 1943.667 1943.333 1 0
  5 0 1943 1943 20796.5 1943.667 1943.333 2 0
  6 0 1943 1943 20796.5 1943.667 1943.333 3 0

> th <- transform(th,
+      hist = factor(hist, labels=c("Sem","nS","Oth") ),
+      A = age,
+      P = diag,
+      D = d,
+      Y = y/10^5 )[,c("A","P","D","Y","hist")]

APC-models for several datasets (APC2)
```r
> library( Epi )
> stat.table( list( Histology = hist ),
+     list( D = sum(D),
+           Y = sum(Y) ),
+     margins = TRUE,
+     data = th )
```

<table>
<thead>
<tr>
<th>Histology</th>
<th>D</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sem</td>
<td>4708.00</td>
<td>1275.25</td>
</tr>
<tr>
<td>nS</td>
<td>3632.00</td>
<td>1275.25</td>
</tr>
<tr>
<td>Oth</td>
<td>466.00</td>
<td>1275.25</td>
</tr>
<tr>
<td>Total</td>
<td>8806.00</td>
<td>3825.76</td>
</tr>
</tbody>
</table>

First step is separate analyses for each subtype (Sem,nS)
> apc.Sem <- apc.fit( subset( th, hist=="Sem" ),
+       parm = "ACP",
+       ref.c = 1970,
+       npar = c(A=8,P=8,C=8) )

[1] "ML of APC-model Poisson with log(Y) offset : ( ACP ):
"

Analysis of deviance for Age-Period-Cohort model

<table>
<thead>
<tr>
<th></th>
<th>Resid. Df</th>
<th>Resid. Dev</th>
<th>Df</th>
<th>Deviance</th>
<th>Pr(&gt;Chi)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>9712</td>
<td>6845.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age-drift</td>
<td>9711</td>
<td>6255.1</td>
<td>1</td>
<td>590.70</td>
<td>&lt; 2.2e-16</td>
</tr>
<tr>
<td>Age-Cohort</td>
<td>9705</td>
<td>6210.0</td>
<td>6</td>
<td>45.09</td>
<td>4.500e-08</td>
</tr>
<tr>
<td>Age-Period-Cohort</td>
<td>9699</td>
<td>6184.1</td>
<td>6</td>
<td>25.90</td>
<td>0.0002323</td>
</tr>
<tr>
<td>Age-Period</td>
<td>9705</td>
<td>6241.9</td>
<td>-6</td>
<td>-57.75</td>
<td>1.289e-10</td>
</tr>
<tr>
<td>Age-drift</td>
<td>9711</td>
<td>6255.1</td>
<td>-6</td>
<td>-13.24</td>
<td>0.0393950</td>
</tr>
</tbody>
</table>

APC-models for several datasets (APC2)
> apc.nS <- apc.fit( subset( th, hist=="nS" ),
+ parm = "ACP",
+ ref.c = 1970,
+ npar = c(A=8,P=8,C=8) )

[1] "ML of APC-model Poisson with log(Y) offset : ( ACP ):\n"

Analysis of deviance for Age-Period-Cohort model

<table>
<thead>
<tr>
<th></th>
<th>Resid. Df</th>
<th>Resid. Dev</th>
<th>Df</th>
<th>Deviance</th>
<th>Pr(&gt;Chi)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>9712</td>
<td>6316.4</td>
<td>6</td>
<td>697.29</td>
<td>&lt; 2.2e-16</td>
</tr>
<tr>
<td>Age-drift</td>
<td>9711</td>
<td>5619.1</td>
<td>1</td>
<td>697.29</td>
<td>&lt; 2.2e-16</td>
</tr>
<tr>
<td>Age-Cohort</td>
<td>9705</td>
<td>5575.6</td>
<td>6</td>
<td>43.51</td>
<td>9.243e-08</td>
</tr>
<tr>
<td>Age-Period-Cohort</td>
<td>9699</td>
<td>5502.9</td>
<td>6</td>
<td>72.75</td>
<td>1.117e-13</td>
</tr>
<tr>
<td>Age-Period</td>
<td>9705</td>
<td>5550.8</td>
<td>-6</td>
<td>-47.91</td>
<td>1.229e-08</td>
</tr>
<tr>
<td>Age-drift</td>
<td>9711</td>
<td>5619.1</td>
<td>-6</td>
<td>-68.34</td>
<td>8.945e-13</td>
</tr>
</tbody>
</table>

> apc.Sem$Drift

<table>
<thead>
<tr>
<th></th>
<th>exp(Est.)</th>
<th>2.5%</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>APC (D-weights)</td>
<td>1.023586</td>
<td>1.021563</td>
<td>1.025614</td>
</tr>
<tr>
<td>A-d</td>
<td>1.023765</td>
<td>1.021773</td>
<td>1.025761</td>
</tr>
</tbody>
</table>

APC-models for several datasets (APC2)
APC-models for several datasets (APC2)
APC-models for several datasets (APC2)

non-Seminoma
Seminoma
RR
Analysis of two rates: Formal tests I

```r
> Ma <- ns( A, df=15, intercept=TRUE )
> Mp <- ns( P, df=15 )
> Mc <- ns( P-A, df=20 )
> Mp <- detrend( Mp, P, weight=D )
> Mc <- detrend( Mc, P-A, weight=D )

> m.apc <- glm( D ~ -1 + Ma:type + Mp:type + Mc:type + offset( log(Y) ), family=poisson )
> m.ap <- update( m.apc, . ~ . - Mc:type + Mc )
> m.ac <- update( m.apc, . ~ . - Mp:type + Mp )
> m.a <- update( m.ap, . ~ . - Mp:type + Mp )

> anova( m.a, m.ac, m.apc, m.ap, m.a, test="Chisq"
```

Analysis of Deviance Table

Model 1: D ~ Mc + Mp + Ma:type + offset(log(Y)) - 1
Model 2: D ~ Mp + Ma:type + type:Mc + offset(log(Y)) - 1
Model 3: D ~ -1 + Ma:type + Mp:type + Mc:type + offset(log(Y))
Model 4: D ~ Mc + Ma:type + type:Mp + offset(log(Y)) - 1

APC-models for several datasets (APC2)
Analysis of two rates: Formal tests II

Model 5: $D \sim Mc + Mp + Ma:\text{type} + \text{offset}(\log(Y)) - 1$

| Resid. Df | Resid. Dev  | Df  | Deviance | P(>|Chi|)   |
|-----------|-------------|-----|----------|------------|
| 1         | 10737       |     | 10553.7  |            |
| 2         | 10718       | 19  | 10367.9  | 185.7      | 2.278e-29 |
| 3         | 10704       | 14  | 10199.6  | 168.3      | 1.513e-28 |
| 4         | 10723       | -19 | 10508.6  | -309.0     | 2.832e-54 |
| 5         | 10737       | -14 | 10553.7  | -45.0      | 4.042e-05 |
APC-model: Interactions

Statistical Analysis in the Lexis Diagram:

Age-Period-Cohort models
May 2016
Max Planck Institut for Demographic Research, Rostock
http://BendixCarstensen/APC/MPIDR-2016

APC-int
Analysis of DM-rates: Age $\times$ sex interaction I

- 10 centres
- 2 sexes
- Age: 0-15
- Period 1989–1999

- Is the sex-effect the same between all centres?
- How are the timetrends.
library( Epi )
library( splines )
load( file="c:/Bendix/Artikler/A_P_C/IDDM/Eurodiab/data/tri.Rdata" )

dm <- dm[dm$cen=="D1: Denmark",]

# Define knots and points of prediction
n.A <- 5
n.C <- 8
n.P <- 5
pA <- seq(1/(3*n.A),1-1/(3*n.A),,n.A )
pC <- seq(1/(3*n.C),1-1/(3*n.C),,n.P )
pP <- seq(1/(3*n.P),1-1/(3*n.P),,n.C )
c0 <- 1985
attach( dm, warn.conflicts=FALSE )
A.kn <- quantile( rep( A, D ), probs=pA[-c(1,n.A)] )
A.ok <- quantile( rep( A, D ), probs=pA[ c(1,n.A)] )
A.pt <- sort( A[match( unique(A), A )] )
C.kn <- quantile( rep( C, D ), probs=pC[-c(1,n.C)] )
C.ok <- quantile( rep( C, D ), probs=pC[ c(1,n.C)] )
C.pt <- sort( C[match( unique(C), C )] )

APC-model: Interactions (APC-int)
299/ 327
Analysis of DM-rates: Age × sex interaction III

P.kn <- quantile( rep( P, D ), probs=pP[-c(1,n.P)] )
P.ok <- quantile( rep( P, D ), probs=pP[ c(1,n.P)] )
P.pt <- sort( P[match( unique(P), P )] )

# Age-cohort model with age-sex interaction
# The model matrices for the ML fit
Ma <- ns( A, kn=A.kn, Bo=A.ok, intercept=T )
Mc <- cbind( C-c0, detrend( ns( C, kn=C.kn, Bo=C.ok ), C, weight=D ) )
Mp <- detrend( ns( P, kn=P.kn, Bo=P.ok ), P, weight=D )

# The prediction matrices
Pa <- Ma[match(A.pt,A),,drop=F]
Pc <- Mc[match(C.pt,C),,drop=F]
Pp <- Mp[match(P.pt,P),,drop=F]

# Fit the apc model by ML
apcs <- glm( D ~ Ma:sex - 1 + Mc + Mp +
            offset( log (Y/10^5) ),
            family=poisson,
data=dm )

summary( apcs )
Analysis of DM-rates: Age × sex interaction IV

ci.lin( apcs )
ci.lin( apcs, subset="sexF", Exp=T)
ci.lin( apcs, subset="sexF", ctr.mat=Pa, Exp=T)

# Extract the effects
F.inc <- ci.lin( apcs, subset="sexF", ctr.mat=Pa, Exp=T)[,5:7]
M.inc <- ci.lin( apcs, subset="sexM", ctr.mat=Pa, Exp=T)[,5:7]
MF.RR <- ci.lin( apcs, subset=c("sexM","sexF"), ctr.mat=cbind(Pa,-Pa), Exp=T)[,5:7]
c.RR <- ci.lin( apcs, subset="Mc", ctr.mat=Pc, Exp=T)[,5:7]

# plt( paste( "DM-DK" ), width=11 )
par( mar=c(4,4,1,4), mgp=c(3,1,0)/1.6, las=1 )
# The the frame for the effects
fr <- apc.frame( a.lab=c(0,5,10,15),
    a.tic=c(0,5,10,15),
    r.lab=c(c(1,1.5,3,5),c(1,1.5,3,5)*10),
    r.tic=c(c(1,1.5,2,5),c(1,1.5,2,5)*10),
    cp.lab=seq(1980,2000,10),
    cp.tic=seq(1975,2000,5),
)
Analysis of DM-rates: Age×sex interaction V

```r
rr.ref=5,
gap=1,
col.grid=gray(0.9),
a.txt="",
cp.txt="",
r.txt="",
rr.txt=""
)

# Draw the estimates
matlines( A.pt, M.inc, lwd=c(3,1,1), lty=1, col="blue" )
matlines( A.pt, F.inc, lwd=c(3,1,1), lty=1, col="red" )
matlines( C.pt - fr[1], c.RR * fr[2],
          lwd=c(3,1,1), lty=1, col="black" )
matlines( P.pt - fr[1], p.RR * fr[2],
          lwd=c(3,1,1), lty=1, col="black" )
matlines( A.pt, MF.RR * fr[2],
          lwd=c(3,1,1), lty=1, col=gray(0.6) )
abline(h=fr[2])
```

APC-model: Interactions (APC-int)
APC-model: Interactions (APC-int)
Predicting future rates

Statistical Analysis in the
Lexis Diagram:
Age-Period-Cohort models
May 2016
Max Planck Institut for Demographic Research, Rostock
http://BendixCarstensen/APC(MPIDR-2016)
Prediction of future rates

Model:
\[ \log(\lambda(a, p)) = f(a) + g(p) + h(c) \]

- Why not just extend the estimated functions into the future?
- The parametrization curse — the model as stated is not uniquely parametrized.
- Predictions from the model must be invariant under reparametrization.
Identifiability

Predictions based in the three functions \( f(a) \), \( g(p) \) and \( h(c) \) must give the same prediction also for the reparametrized version:

\[
\begin{align*}
\log(\lambda(a, p)) &= \tilde{f}(a) + \tilde{g}(p) + \tilde{h}(c) \\
&= (f(a) - \gamma a) + \\
&\quad (g(p) + \gamma p) + \\
&\quad (h(c) - \gamma c)
\end{align*}
\]
**Parametrization invariance**

- Prediction of the future course of $g$ and $h$ must preserve addition of a linear term in the argument:

  \[
  \text{pred}(g(p) + \gamma p) = \text{pred}(g(p)) + \gamma p \\
  \text{pred}(h(c) - \gamma c) = \text{pred}(h(c)) - \gamma c
  \]

- If this is met, the predictions made will not depend on the parametrization chosen.
- If one of the conditions does not hold, the prediction will depend on the parametrization chosen.
- Any linear combination of (known) function values of $g(p)$ and $h(c)$ will work.
Identifiability

- Any linear combination of function values of $g(p)$ and $h(c)$ will work.
- Coefficients in the linear combinations used for $g$ and $h$ must be the same; otherwise the prediction will depend on the specific parametrization.
- What works best in reality is difficult to say: depends on the subject matter.
Example: Breast cancer in Denmark

Predicting future rates (predict)
Practicalities

- Long term predictions notoriously unstable.
- Decreasing slopes are possible, the requirement is that at any future point changes in the parametrization should cancel out in the predictions.
Predicted age-specific breast cancer rates at 2020 (black),
in the 1950 cohort (blue),
and the estimated age-effects (red).
Continuous outcomes

Statistical Analysis in the
Lexis Diagram:

Age-Period-Cohort models
May 2016
Max Planck Institut for Demographic Research, Rostock
http://BendixCarstensen/APC/MPIDR-2016

cont
APC-model for quantitative outcomes

- The classical model is:
  \[ \log(\lambda(a, p)) = f(a) + g(p) + h(p - a) \]

- In principle it would be possible to use an identity-link model:
  \[ \lambda(a, p) = f(a) + g(p) + h(p - a) \]

- ... or use APC-modelling for **measurement** data such as BMI, measured at different times and ages:
  \[ \text{BMI}_{ap} = f(a) + g(p) + h(p - a) + e_{ap}, \quad e_i \sim \mathcal{N}(0, \sigma^2) \]

- ... or more precisely:
  \[ \text{BMI}_i = f(a(i)) + g(p(i)) + h(p(i) - a(i)) + e_i, \quad e_i \sim \mathcal{N}(0, \sigma^2) \]
APC-model for quantitative outcomes

- Model:

\[
\text{BMI}_i = f(a(i)) + g(p(i)) + h(p(i) - a(i)) + e_i, \quad e_i \sim \mathcal{N}(0, \sigma^2)
\]

- But the identification problem is still the same:

\[
c(i) = p(i) - a(i), \quad \forall i
\]

- But the same machinery applies with extraction of the effects — and plotting of predictions of

  - \( E(\text{BMI}) \)
  - quantiles of BMI
APC-model for quantitative outcomes

- Australian surveys
- 40,000+ person surveyed at different times
- Date of birth, data of survey, sex and BMI known.
- How does BMI evolve in the population?
- Linear model \( \mathbb{E}(\text{BMI}) \)
- Quantile regression (median, quantile)
- — the latter is not a model
Continuous outcomes (cont)
Continuous outcomes (cont)
Continuous outcomes (cont)
Continuous outcomes (cont)
Continuous outcomes (cont)

BMI percentiles in 1950 cohort (kg/m²)

BMI differences

Age Date of birth Date of survey

Continuous outcomes (cont)

322/327
Continuous outcomes (cont)