

Representation of follow-up

Bendix Carstensen Steno Diabetes Center Copenhagen
Herlev, Denmark
`http://BendixCarstensen.com`

SPE, Tartu, Estonia,

June 2023

`http://BendixCarstensen.com/SPE`

From C:\Bendix\teach\SPE\git\lectures\time-rep\time-rep.tex

Tuesday 30 May, 2023, 08:40

Representation of follow-up

Bendix Carstensen

Representation of follow-up

SPE, Tartu, Estonia,

June 2023

<http://BendixCarstensen.com/SPE>

time-split

- ▶ In follow-up studies we estimate rates from:

- ▶ In follow-up studies we estimate rates from:
 - ▶ D — events, deaths

- ▶ In follow-up studies we estimate rates from:
 - ▶ D — events, deaths
 - ▶ Y — person-years

- ▶ In follow-up studies we estimate rates from:
 - ▶ D — events, deaths
 - ▶ Y — person-years
 - ▶ $\hat{\lambda} = D/Y$ rates

- ▶ In follow-up studies we estimate rates from:
 - ▶ D — events, deaths
 - ▶ Y — person-years
 - ▶ $\hat{\lambda} = D/Y$ rates
 - ▶ ... empirical counterpart of intensity — an **estimate**

- ▶ In follow-up studies we estimate rates from:
 - ▶ D — events, deaths
 - ▶ Y — person-years
 - ▶ $\hat{\lambda} = D/Y$ rates
 - ▶ ... empirical counterpart of intensity — an **estimate**
- ▶ Rates differ between persons.

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

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

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

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

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

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

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

Representation of follow-up data

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

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

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

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

These are specific for each **type** of outcome.

Stratification by age

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

Stratification by age

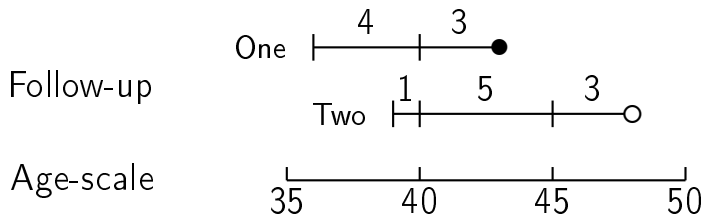
If follow-up is rather short, age at entry is OK for age-stratification.

If follow-up is long, stratification by categories of **current age** is preferable.

Stratification by age

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

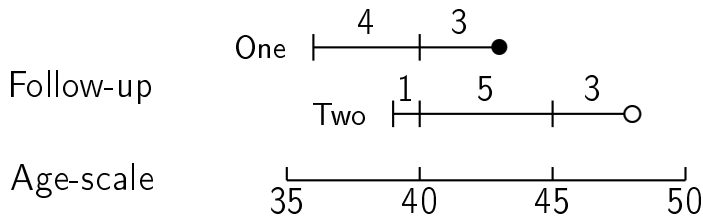
If follow-up is long, stratification by categories of **current age** is preferable.



Stratification by age

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

If follow-up is long, stratification by categories of **current age** is preferable.

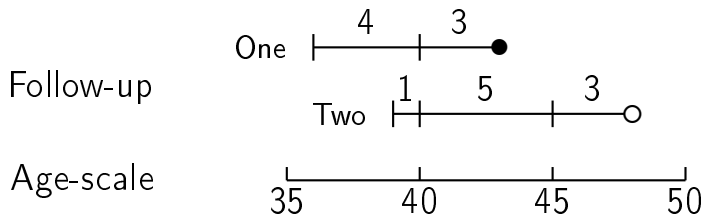


— allowing rates to vary across age-bands

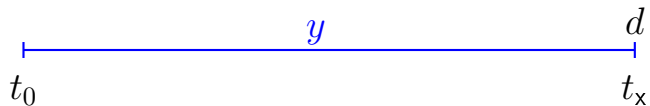
Stratification by age

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

If follow-up is long, stratification by categories of **current age** is preferable.



- allowing rates to vary across age-bands
- how do we do the split and why is it OK?

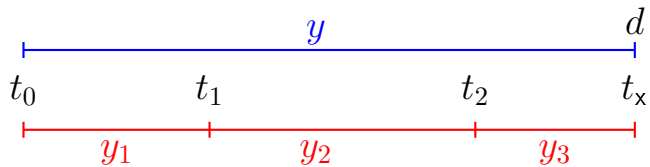


Probability

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

log-Likelihood

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

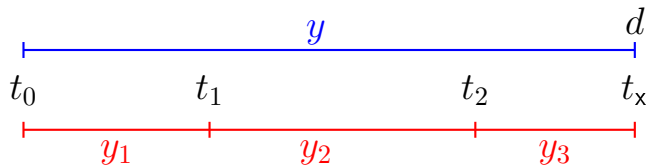


Probability

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

log-Likelihood

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



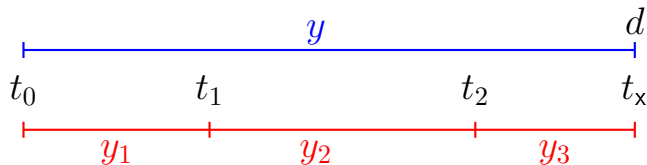
Probability

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

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

log-Likelihood

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



Probability

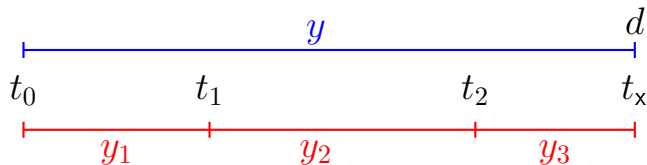
$$P(d \text{ at } t_x | \text{entry } t_0)$$

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

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

log-Likelihood

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



Probability

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

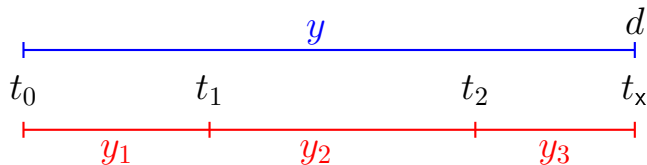
$$= P(\text{surv } t_0 \rightarrow t_1 | \text{entry } t_0)$$

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

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

log-Likelihood

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



Probability

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

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

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

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

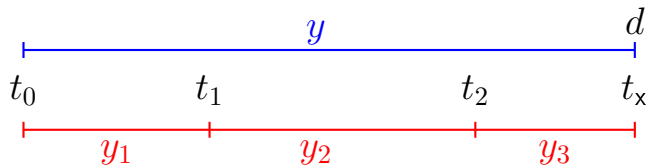
log-Likelihood

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

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

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

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



Probability

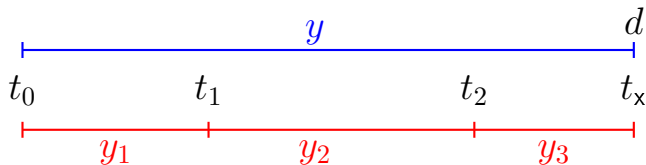
$$P(d \text{ at } t_x | \text{entry } t_0)$$

$$\begin{aligned}
 &= P(\text{surv } t_0 \rightarrow t_1 | \text{entry } t_0) \\
 &\times P(\text{surv } t_1 \rightarrow t_2 | \text{entry } t_1) \\
 &\times P(d \text{ at } t_x | \text{entry } t_2)
 \end{aligned}$$

log-Likelihood

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

$$\begin{aligned}
 &= 0 \log(\lambda_1) - \lambda_1 y_1 \\
 &+ 0 \log(\lambda_2) - \lambda_2 y_2 \\
 &+ d \log(\lambda_3) - \lambda_3 y_3
 \end{aligned}$$



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

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

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

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

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

— allows different rates (λ_i) in each interval

Dividing time into bands requires:

Dividing time into bands requires:

Origin: The date where the time scale is 0:

Dividing time into bands requires:

Origin: The date where the time scale is 0:

- ▶ Age — 0 at date of birth

Dividing time into bands requires:

Origin: The date where the time scale is 0:

- ▶ Age — 0 at date of birth
- ▶ Disease duration — 0 at date of diagnosis

Dividing time into bands requires:

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

Dividing time into bands requires:

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

Dividing time into bands requires:

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?

Dividing time into bands requires:

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?

Dividing time into bands requires:

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?

Dividing time into bands requires:

Origin: The date where the time scale is 0:

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

Intervals: How should it be subdivided:

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

Aim: Separate rate in each interval, mimicking continuous time by using small intervals:

—time at the beginning of interval as quantitative variable.

Example: cohort with 3 persons:

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

Example: cohort with 3 persons:

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

- Age bands: 10-years intervals of current age.

Example: cohort with 3 persons:

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

- ▶ Age bands: 10-years intervals of current age.
- ▶ Split Y for every subject accordingly

Example: cohort with 3 persons:

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

- ▶ Age bands: 10-years intervals of current age.
- ▶ Split Y for every subject accordingly
- ▶ Treat each segment as a separate unit of observation.

Example: cohort with 3 persons:

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

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

Splitting the follow-up

	subj. 1	subj. 2	subj. 3
Age at E ntry:	13.06	18.44	4.54
Age at e X it:	44.95	41.14	11.12
S tatus at exit:	Dead	Alive	Dead
<hr/>			
<i>Y</i>	31.89	22.70	6.58
<i>D</i>	1	0	1

	subj. 1		subj. 2		subj. 3		Σ	
Age	<i>Y</i>	<i>D</i>	<i>Y</i>	<i>D</i>	<i>Y</i>	<i>D</i>	<i>Y</i>	<i>D</i>
0–	0.00	0	0.00	0	5.46	0	5.46	0
10–	6.94	0	1.56	0	1.12	1	8.62	1
20–	10.00	0	10.00	0	0.00	0	20.00	0
30–	10.00	0	10.00	0	0.00	0	20.00	0
40–	4.95	1	1.14	0	0.00	0	6.09	1
Σ	31.89	1	22.70	0	6.58	1	60.17	2

Splitting the follow-up

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

Keeping track of calendar time too?

Follow-up intervals on several timescales

- ▶ The risk-time is the same on all timescales

Follow-up intervals on several timescales

- ▶ The risk-time is the same on all timescales
- ▶ Only need the entry point on each time scale:

Follow-up intervals on several timescales

- ▶ The risk-time is the same on all timescales
- ▶ Only need the entry point on each time scale:
 - ▶ Age at entry.

Follow-up intervals on several timescales

- ▶ The risk-time is the same on all timescales
- ▶ Only need the entry point on each time scale:
 - ▶ Age at entry.
 - ▶ Date of entry.

Follow-up intervals on several timescales

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

Follow-up intervals on several timescales

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

(d, y) (event, duration)

Follow-up intervals on several timescales

- ▶ The risk-time is the same on all timescales
- ▶ Only need the entry point on each time scale:
 - ▶ Age at entry.
 - ▶ Date of entry.
 - ▶ Time since treatment at entry.
 - if time of treatment is the entry, this is 0 for all.
- ▶ **Response variable** in analysis of rates:
 (d, y) (**event**, **duration**)
- ▶ **Covariates** in analysis of rates:

Follow-up intervals on several timescales

- ▶ The risk-time is the same on all timescales
- ▶ Only need the entry point on each time scale:
 - ▶ Age at entry.
 - ▶ Date of entry.
 - ▶ Time since treatment at entry.
 - if time of treatment is the entry, this is 0 for all.
- ▶ **Response variable** in analysis of rates:
 (d, y) (**event, duration**)
- ▶ **Covariates** in analysis of rates:
 - ▶ **timescales**

Follow-up intervals on several timescales

- ▶ The risk-time is the same on all timescales
- ▶ Only need the entry point on each time scale:
 - ▶ Age at entry.
 - ▶ Date of entry.
 - ▶ Time since treatment at entry.
 - if time of treatment is the entry, this is 0 for all.
- ▶ Response variable in analysis of rates:
 (d, y) (event, duration)
- ▶ Covariates in analysis of rates:
 - ▶ timescales
 - ▶ other (fixed) measurements

Follow-up intervals on several timescales

- ▶ The risk-time is the same on all timescales
- ▶ Only need the entry point on each time scale:
 - ▶ Age at entry.
 - ▶ Date of entry.
 - ▶ Time since treatment at entry.
 - if time of treatment is the entry, this is 0 for all.
- ▶ **Response variable** in analysis of rates:
 (d, y) (**event**, **duration**)
- ▶ **Covariates** in analysis of rates:
 - ▶ **timescales**
 - ▶ other (fixed) measurements
- ▶ ... do not confuse **duration** and **timescale** !

Follow-up data in Epi — Lexis objects I

```
> thoro[1:6,1:8]
```

	id	sex	birthdat	contrast	injecdat	volume	exitdat	exitstat
1	1	2	1916.609	1	1938.791	22	1976.787	1
2	2	2	1927.843	1	1943.906	80	1966.030	1
3	3	1	1902.778	1	1935.629	10	1959.719	1
4	4	1	1918.359	1	1936.396	10	1977.307	1
5	5	1	1902.931	1	1937.387	10	1945.387	1
6	6	2	1903.714	1	1937.316	20	1944.738	1

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

Follow-up data in Epi — Lexis objects II

NOTE: entry.status has been set to 0 for all.

NOTE: Dropping 2 rows with duration of follow up < tol

```
> summary(thL, timeScales = TRUE)
```

Transitions:

To

From	0	1	Records:	Events:	Risk time:	Persons:
	0 504 1964		2468	1964	51934.08	2468

Timescales:

age	dte	tfi
""	""	""

Definition of Lexis object

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

:

Definition of Lexis object

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

entry is defined on **three** timescales,

:

Definition of Lexis object

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

entry is defined on **three** timescales,
but **exit** is only needed on **one** timescale (or vice versa):

Definition of Lexis object

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

entry is defined on **three** timescales,
but exit is only needed on **one** timescale (or vice versa):
Follow-up time is the same on all timescales:

Definition of Lexis object

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

entry is defined on **three** timescales,
but exit is only needed on **one** timescale (or vice versa):
Follow-up time is the same on all timescales: $\text{exitdat} - \text{injecdat}$

Definition of Lexis object

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

entry is defined on **three** timescales,

but exit is only needed on **one** timescale (or vice versa):

Follow-up time is the same on all timescales: $\text{exitdat} - \text{injecdat}$

One element of entry and exit must have same name (**dte**).

The looks of a Lexis object

```
> thL[1:4,1:9]
```

	age	dte	tfi	lex.dur	lex.Cst	lex.Xst	lex.id
1	22.18	1938.79	0	37.99	0	1	1
2	49.54	1945.77	0	18.59	0	1	2
3	68.20	1955.18	0	1.40	0	1	3
4	20.80	1957.61	0	34.52	0	0	4
...							

The looks of a Lexis object

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

> summary(thL)
Transitions:
      To
From    0      1 Records:  Events:  Risk time:  Persons:
      0 504 1964      2468      1964      51934.08      2468
```

The looks of a Lexis object

```
> thL[1:4,1:9]
```

	age	dte	tfi	lex.dur	lex.Cst	lex.Xst	lex.id
1	22.18	1938.79	0	37.99	0	1	1
2	49.54	1945.77	0	18.59	0	1	2
3	68.20	1955.18	0	1.40	0	1	3
4	20.80	1957.61	0	34.52	0	0	4

...

```
> summary(thL)
```

Transitions:

	To					
From	0	1	Records:	Events:	Risk time:	Persons:
	0	504	1964	2468	51934.08	2468

The looks of a Lexis object

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

```
> summary(thL)
```

Transitions:

To

From	0	1	Records:	Events:	Risk time:	Persons:
	0	504 1964	2468	1964	51934.08	2468

The looks of a Lexis object

```
> thL[1:4,1:9]
```

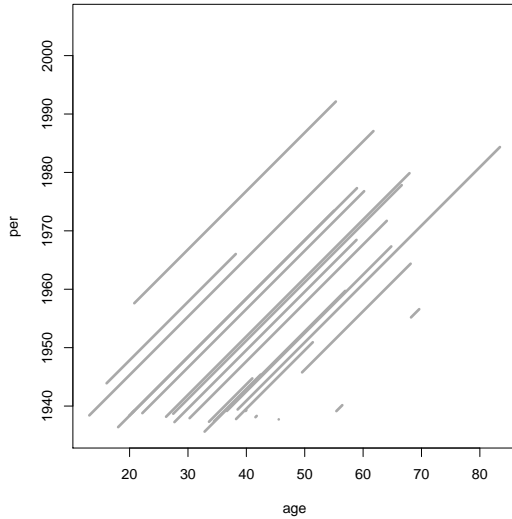
	age	dte	tfi	lex.dur	lex.Cst	lex.Xst	lex.id
1	22.18	1938.79	0	37.99	0	1	1
2	49.54	1945.77	0	18.59	0	1	2
3	68.20	1955.18	0	1.40	0	1	3
4	20.80	1957.61	0	34.52	0	0	4

...

```
> summary(thL)
```

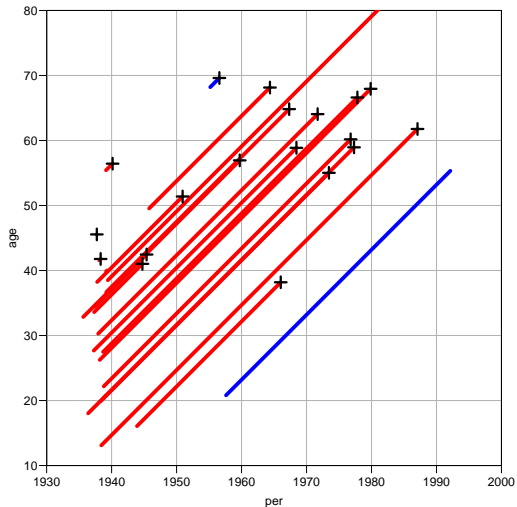
Transitions:

	To	From	0	1	Records:	Events:	Risk time:	Persons:
		0	504	1964	2468	1964	51934.08	2468



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

Representation of follow-up (time-split)



Lexis diagram

```
> 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 )
```


EINLEITUNG
IN DIE
THEORIE
DER
BEVÖLKERUNGSSTATISTIK

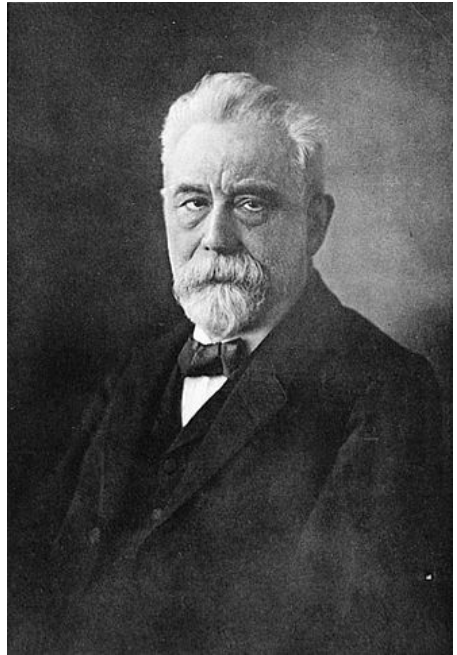
VON

W. LEXIS

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

STRASSBURG

KARL J. TRÜBNER



Splitting follow-up time

```
> spl1 <- splitLexis( thL, time.scale="age", breaks=seq(0,100,20) )
> round(spl1,1)
```

	age	dte	tfi	lex.dur	lex.Cst	lex.Xst	id	sex	birthdat	contrast	injecdat	vo
1	22.2	1938.8	0.0	17.8	0	0	1	2	1916.6	1	1938.8	
2	40.0	1956.6	17.8	20.0	0	0	1	2	1916.6	1	1938.8	
3	60.0	1976.6	37.8	0.2	0	1	1	2	1916.6	1	1938.8	
4	49.5	1945.8	0.0	10.5	0	0	640	2	1896.2	1	1945.8	
5	60.0	1956.2	10.5	8.1	0	1	640	2	1896.2	1	1945.8	
6	68.2	1955.2	0.0	1.4	0	1	3425	1	1887.0	2	1955.2	
7	20.8	1957.6	0.0	19.2	0	0	4017	2	1936.8	2	1957.6	
8	40.0	1976.8	19.2	15.3	0	0	4017	2	1936.8	2	1957.6	
...												

Splitting follow-up time

```
> spl1 <- splitLexis( thL, time.scale="age", breaks=seq(0,100,20) )  
> round(spl1,1)
```

	age	dte	tfi	lex.dur	lex.Cst	lex.Xst	id	sex	birthdat	contrast	injecdat	vo
1	22.2	1938.8	0.0	17.8	0	0	1	2	1916.6	1	1938.8	
2	40.0	1956.6	17.8	20.0	0	0	1	2	1916.6	1	1938.8	
3	60.0	1976.6	37.8	0.2	0	1	1	2	1916.6	1	1938.8	
4	49.5	1945.8	0.0	10.5	0	0	640	2	1896.2	1	1945.8	
5	60.0	1956.2	10.5	8.1	0	1	640	2	1896.2	1	1945.8	
6	68.2	1955.2	0.0	1.4	0	1	3425	1	1887.0	2	1955.2	
7	20.8	1957.6	0.0	19.2	0	0	4017	2	1936.8	2	1957.6	
8	40.0	1976.8	19.2	15.3	0	0	4017	2	1936.8	2	1957.6	
...												

Splitting follow-up time

```
> spl1 <- splitLexis( thL, time.scale="age", breaks=seq(0,100,20) )
> round(spl1,1)
```

	age	dte	tfi	lex.dur	lex.Cst	lex.Xst	id	sex	birthdat	contrast	injecdat	vo
1	22.2	1938.8	0.0	17.8	0	0	1	2	1916.6	1	1938.8	
2	40.0	1956.6	17.8	20.0	0	0	1	2	1916.6	1	1938.8	
3	60.0	1976.6	37.8	0.2	0	1	1	2	1916.6	1	1938.8	
4	49.5	1945.8	0.0	10.5	0	0	640	2	1896.2	1	1945.8	
5	60.0	1956.2	10.5	8.1	0	1	640	2	1896.2	1	1945.8	
6	68.2	1955.2	0.0	1.4	0	1	3425	1	1887.0	2	1955.2	
7	20.8	1957.6	0.0	19.2	0	0	4017	2	1936.8	2	1957.6	
8	40.0	1976.8	19.2	15.3	0	0	4017	2	1936.8	2	1957.6	
...												

Splitting follow-up time

```
> spl1 <- splitLexis( thL, time.scale="age", breaks=seq(0,100,20) )
> round(spl1,1)
```

	age	dte	tfi	lex.dur	lex.Cst	lex.Xst	id	sex	birthdat	contrast	injecdat	vo
1	22.2	1938.8	0.0	17.8	0	0	1	2	1916.6	1	1938.8	
2	40.0	1956.6	17.8	20.0	0	0	1	2	1916.6	1	1938.8	
3	60.0	1976.6	37.8	0.2	0	1	1	2	1916.6	1	1938.8	
4	49.5	1945.8	0.0	10.5	0	0	640	2	1896.2	1	1945.8	
5	60.0	1956.2	10.5	8.1	0	1	640	2	1896.2	1	1945.8	
6	68.2	1955.2	0.0	1.4	0	1	3425	1	1887.0	2	1955.2	
7	20.8	1957.6	0.0	19.2	0	0	4017	2	1936.8	2	1957.6	
8	40.0	1976.8	19.2	15.3	0	0	4017	2	1936.8	2	1957.6	
...												

Splitting follow-up time

```
> spl1 <- splitLexis( thL, time.scale="age", breaks=seq(0,100,20) )
> round(spl1,1)
```

	age	dte	tfi	lex.dur	lex.Cst	lex.Xst	id	sex	birthdat	contrast	injecdat	vo
1	22.2	1938.8	0.0	17.8	0	0	1	2	1916.6	1	1938.8	
2	40.0	1956.6	17.8	20.0	0	0	1	2	1916.6	1	1938.8	
3	60.0	1976.6	37.8	0.2	0	1	1	2	1916.6	1	1938.8	
4	49.5	1945.8	0.0	10.5	0	0	640	2	1896.2	1	1945.8	
5	60.0	1956.2	10.5	8.1	0	1	640	2	1896.2	1	1945.8	
6	68.2	1955.2	0.0	1.4	0	1	3425	1	1887.0	2	1955.2	
7	20.8	1957.6	0.0	19.2	0	0	4017	2	1936.8	2	1957.6	
8	40.0	1976.8	19.2	15.3	0	0	4017	2	1936.8	2	1957.6	
...												

Split on another timescale

```
> spl2 <- splitLexis( spl1, time.scale="tfi", breaks=c(0,1,5,20,100) )  
> round( spl2, 1 )
```

	lex.id	age	dte	tfi	lex.dur	lex.Cst	lex.Xst	id	sex	birthdat	contrast	inje
1	1	22.2	1938.8	0.0	1.0	0	0	1	2	1916.6	1	19
2	1	23.2	1939.8	1.0	4.0	0	0	1	2	1916.6	1	19
3	1	27.2	1943.8	5.0	12.8	0	0	1	2	1916.6	1	19
4	1	40.0	1956.6	17.8	2.2	0	0	1	2	1916.6	1	19
5	1	42.2	1958.8	20.0	17.8	0	0	1	2	1916.6	1	19
6	1	60.0	1976.6	37.8	0.2	0	1	1	2	1916.6	1	19
7	2	49.5	1945.8	0.0	1.0	0	0	640	2	1896.2	1	19
8	2	50.5	1946.8	1.0	4.0	0	0	640	2	1896.2	1	19
9	2	54.5	1950.8	5.0	5.5	0	0	640	2	1896.2	1	19
10	2	60.0	1956.2	10.5	8.1	0	1	640	2	1896.2	1	19
11	3	68.2	1955.2	0.0	1.0	0	0	3425	1	1887.0	2	19
12	3	69.2	1956.2	1.0	0.4	0	1	3425	1	1887.0	2	19
13	4	20.8	1957.6	0.0	1.0	0	0	4017	2	1936.8	2	19
14	4	21.8	1958.6	1.0	4.0	0	0	4017	2	1936.8	2	19
15	4	25.8	1962.6	5.0	14.2	0	0	4017	2	1936.8	2	19
16	4	40.0	1976.8	19.2	0.8	0	0	4017	2	1936.8	2	19
17	4	40.8	1977.6	20.0	14.5	0	0	4017	2	1936.8	2	19

Split on another timescale

```
> spl2 <- splitLexis( spl1, time.scale="tfi", breaks=c(0,1,5,20,100) )  
> round( spl2, 1 )
```

	lex.id	age	dte	tfi	lex.dur	lex.Cst	lex.Xst	id	sex	birthdat	contrast	inje
1	1	22.2	1938.8	0.0	1.0	0	0	1	2	1916.6	1	19
2	1	23.2	1939.8	1.0	4.0	0	0	1	2	1916.6	1	19
3	1	27.2	1943.8	5.0	12.8	0	0	1	2	1916.6	1	19
4	1	40.0	1956.6	17.8	2.2	0	0	1	2	1916.6	1	19
5	1	42.2	1958.8	20.0	17.8	0	0	1	2	1916.6	1	19
6	1	60.0	1976.6	37.8	0.2	0	1	1	2	1916.6	1	19
7	2	49.5	1945.8	0.0	1.0	0	0	640	2	1896.2	1	19
8	2	50.5	1946.8	1.0	4.0	0	0	640	2	1896.2	1	19
9	2	54.5	1950.8	5.0	5.5	0	0	640	2	1896.2	1	19
10	2	60.0	1956.2	10.5	8.1	0	1	640	2	1896.2	1	19
11	3	68.2	1955.2	0.0	1.0	0	0	3425	1	1887.0	2	19
12	3	69.2	1956.2	1.0	0.4	0	1	3425	1	1887.0	2	19
13	4	20.8	1957.6	0.0	1.0	0	0	4017	2	1936.8	2	19
14	4	21.8	1958.6	1.0	4.0	0	0	4017	2	1936.8	2	19
15	4	25.8	1962.6	5.0	14.2	0	0	4017	2	1936.8	2	19
16	4	40.0	1976.8	19.2	0.8	0	0	4017	2	1936.8	2	19
17	4	40.8	1977.6	20.0	14.5	0	0	4017	2	1936.8	2	19

Split on another timescale

```
> spl2 <- splitLexis( spl1, time.scale="tfi", breaks=c(0,1,5,20,100) )  
> round( spl2, 1 )
```

	lex.id	age	dte	tfi	lex.dur	lex.Cst	lex.Xst	id	sex	birthdat	contrast	inje
1	1	22.2	1938.8	0.0	1.0	0	0	1	2	1916.6	1	19
2	1	23.2	1939.8	1.0	4.0	0	0	1	2	1916.6	1	19
3	1	27.2	1943.8	5.0	12.8	0	0	1	2	1916.6	1	19
4	1	40.0	1956.6	17.8	2.2	0	0	1	2	1916.6	1	19
5	1	42.2	1958.8	20.0	17.8	0	0	1	2	1916.6	1	19
6	1	60.0	1976.6	37.8	0.2	0	1	1	2	1916.6	1	19
7	2	49.5	1945.8	0.0	1.0	0	0	640	2	1896.2	1	19
8	2	50.5	1946.8	1.0	4.0	0	0	640	2	1896.2	1	19
9	2	54.5	1950.8	5.0	5.5	0	0	640	2	1896.2	1	19
10	2	60.0	1956.2	10.5	8.1	0	1	640	2	1896.2	1	19
11	3	68.2	1955.2	0.0	1.0	0	0	3425	1	1887.0	2	19
12	3	69.2	1956.2	1.0	0.4	0	1	3425	1	1887.0	2	19
13	4	20.8	1957.6	0.0	1.0	0	0	4017	2	1936.8	2	19
14	4	21.8	1958.6	1.0	4.0	0	0	4017	2	1936.8	2	19
15	4	25.8	1962.6	5.0	14.2	0	0	4017	2	1936.8	2	19
16	4	40.0	1976.8	19.2	0.8	0	0	4017	2	1936.8	2	19
17	4	40.8	1977.6	20.0	14.5	0	0	4017	2	1936.8	2	19

Split on another timescale

```
> spl2 <- splitLexis( spl1, time.scale="tfi", breaks=c(0,1,5,20,100) )  
> round( spl2, 1 )
```

	lex.id	age	dte	tfi	lex.dur	lex.Cst	lex.Xst	id	sex	birthdat	contrast	inje
1	1	22.2	1938.8	0.0	1.0	0	0	1	2	1916.6	1	19
2	1	23.2	1939.8	1.0	4.0	0	0	1	2	1916.6	1	19
3	1	27.2	1943.8	5.0	12.8	0	0	1	2	1916.6	1	19
4	1	40.0	1956.6	17.8	2.2	0	0	1	2	1916.6	1	19
5	1	42.2	1958.8	20.0	17.8	0	0	1	2	1916.6	1	19
6	1	60.0	1976.6	37.8	0.2	0	1	1	2	1916.6	1	19
7	2	49.5	1945.8	0.0	1.0	0	0	640	2	1896.2	1	19
8	2	50.5	1946.8	1.0	4.0	0	0	640	2	1896.2	1	19
9	2	54.5	1950.8	5.0	5.5	0	0	640	2	1896.2	1	19
10	2	60.0	1956.2	10.5	8.1	0	1	640	2	1896.2	1	19
11	3	68.2	1955.2	0.0	1.0	0	0	3425	1	1887.0	2	19
12	3	69.2	1956.2	1.0	0.4	0	1	3425	1	1887.0	2	19
13	4	20.8	1957.6	0.0	1.0	0	0	4017	2	1936.8	2	19
14	4	21.8	1958.6	1.0	4.0	0	0	4017	2	1936.8	2	19
15	4	25.8	1962.6	5.0	14.2	0	0	4017	2	1936.8	2	19
16	4	40.0	1976.8	19.2	0.8	0	0	4017	2	1936.8	2	19
17	4	40.8	1977.6	20.0	14.5	0	0	4017	2	1936.8	2	19

Split on another timescale

```
> spl2 <- splitLexis( spl1, time.scale="tfi", breaks=c(0,1,5,20,100) )  
> round( spl2, 1 )
```

	lex.id	age	dte	tfi	lex.dur	lex.Cst	lex.Xst	id	sex	birthdat	contrast	inje
1	1	22.2	1938.8	0.0	1.0	0	0	1	2	1916.6	1	19
2	1	23.2	1939.8	1.0	4.0	0	0	1	2	1916.6	1	19
3	1	27.2	1943.8	5.0	12.8	0	0	1	2	1916.6	1	19
4	1	40.0	1956.6	17.8	2.2	0	0	1	2	1916.6	1	19
5	1	42.2	1958.8	20.0	17.8	0	0	1	2	1916.6	1	19
6	1	60.0	1976.6	37.8	0.2	0	1	1	2	1916.6	1	19
7	2	49.5	1945.8	0.0	1.0	0	0	640	2	1896.2	1	19
8	2	50.5	1946.8	1.0	4.0	0	0	640	2	1896.2	1	19
9	2	54.5	1950.8	5.0	5.5	0	0	640	2	1896.2	1	19
10	2	60.0	1956.2	10.5	8.1	0	1	640	2	1896.2	1	19
11	3	68.2	1955.2	0.0	1.0	0	0	3425	1	1887.0	2	19
12	3	69.2	1956.2	1.0	0.4	0	1	3425	1	1887.0	2	19
13	4	20.8	1957.6	0.0	1.0	0	0	4017	2	1936.8	2	19
14	4	21.8	1958.6	1.0	4.0	0	0	4017	2	1936.8	2	19
15	4	25.8	1962.6	5.0	14.2	0	0	4017	2	1936.8	2	19
16	4	40.0	1976.8	19.2	0.8	0	0	4017	2	1936.8	2	19
17	4	40.8	1977.6	20.0	14.5	0	0	4017	2	1936.8	2	19

Split on another timescale

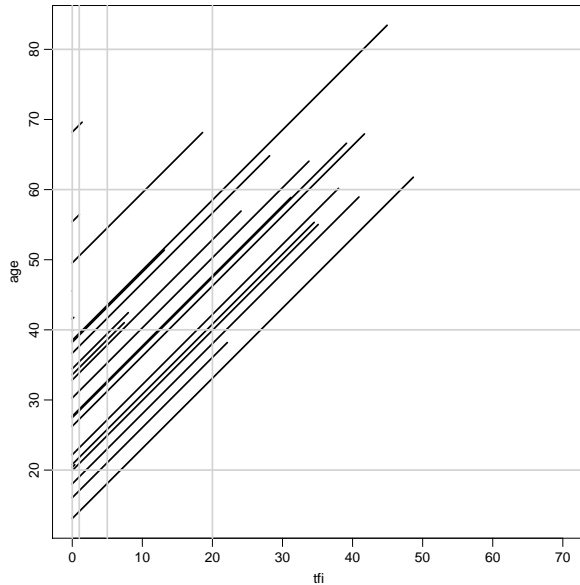
```
> spl2 <- splitLexis( spl1, time.scale="tfi", breaks=c(0,1,5,20,100) )  
> round( spl2, 1 )
```

	lex.id	age	dte	tfi	lex.dur	lex.Cst	lex.Xst	id	sex	birthdat	contrast	inje
1	1	22.2	1938.8	0.0	1.0	0	0	1	2	1916.6	1	19
2	1	23.2	1939.8	1.0	4.0	0	0	1	2	1916.6	1	19
3	1	27.2	1943.8	5.0	12.8	0	0	1	2	1916.6	1	19
4	1	40.0	1956.6	17.8	2.2	0	0	1	2	1916.6	1	19
5	1	42.2	1958.8	20.0	17.8	0	0	1	2	1916.6	1	19
6	1	60.0	1976.6	37.8	0.2	0	1	1	2	1916.6	1	19
7	2	49.5	1945.8	0.0	1.0	0	0	640	2	1896.2	1	19
8	2	50.5	1946.8	1.0	4.0	0	0	640	2	1896.2	1	19
9	2	54.5	1950.8	5.0	5.5	0	0	640	2	1896.2	1	19
10	2	60.0	1956.2	10.5	8.1	0	1	640	2	1896.2	1	19
11	3	68.2	1955.2	0.0	1.0	0	0	3425	1	1887.0	2	19
12	3	69.2	1956.2	1.0	0.4	0	1	3425	1	1887.0	2	19
13	4	20.8	1957.6	0.0	1.0	0	0	4017	2	1936.8	2	19
14	4	21.8	1958.6	1.0	4.0	0	0	4017	2	1936.8	2	19
15	4	25.8	1962.6	5.0	14.2	0	0	4017	2	1936.8	2	19
16	4	40.0	1976.8	19.2	0.8	0	0	4017	2	1936.8	2	19
17	4	40.8	1977.6	20.0	14.5	0	0	4017	2	1936.8	2	19

Split on another timescale

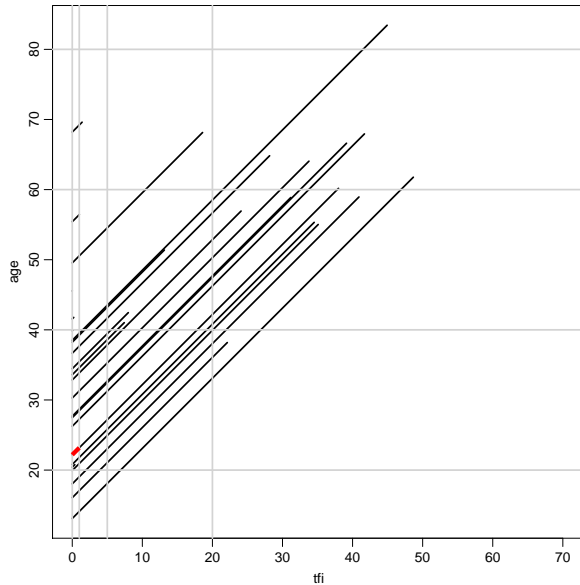
```
> spl2 <- splitLexis( spl1, time.scale="tfi", breaks=c(0,1,5,20,100) )  
> round( spl2, 1 )
```

	lex.id	age	dte	tfi	lex.dur	lex.Cst	lex.Xst	id	sex	birthdat	contrast	inje
1	1	22.2	1938.8	0.0	1.0	0	0	1	2	1916.6	1	19
2	1	23.2	1939.8	1.0	4.0	0	0	1	2	1916.6	1	19
3	1	27.2	1943.8	5.0	12.8	0	0	1	2	1916.6	1	19
4	1	40.0	1956.6	17.8	2.2	0	0	1	2	1916.6	1	19
5	1	42.2	1958.8	20.0	17.8	0	0	1	2	1916.6	1	19
6	1	60.0	1976.6	37.8	0.2	0	1	1	2	1916.6	1	19
7	2	49.5	1945.8	0.0	1.0	0	0	640	2	1896.2	1	19
8	2	50.5	1946.8	1.0	4.0	0	0	640	2	1896.2	1	19
9	2	54.5	1950.8	5.0	5.5	0	0	640	2	1896.2	1	19
10	2	60.0	1956.2	10.5	8.1	0	1	640	2	1896.2	1	19
11	3	68.2	1955.2	0.0	1.0	0	0	3425	1	1887.0	2	19
12	3	69.2	1956.2	1.0	0.4	0	1	3425	1	1887.0	2	19
13	4	20.8	1957.6	0.0	1.0	0	0	4017	2	1936.8	2	19
14	4	21.8	1958.6	1.0	4.0	0	0	4017	2	1936.8	2	19
15	4	25.8	1962.6	5.0	14.2	0	0	4017	2	1936.8	2	19
16	4	40.0	1976.8	19.2	0.8	0	0	4017	2	1936.8	2	19
17	4	40.8	1977.6	20.0	14.5	0	0	4017	2	1936.8	2	19



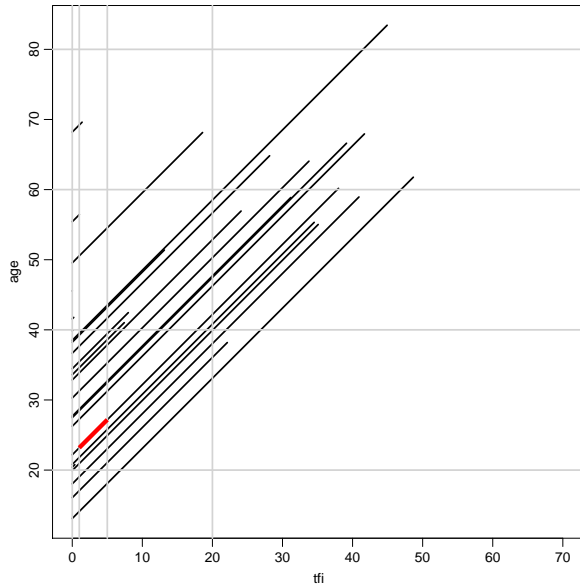
age	tfi	lex.dur	lex.Cst	lex.Xst
22.2	0.0	1.0	0	0
23.2	1.0	4.0	0	0
27.2	5.0	12.8	0	0
40.0	17.8	2.2	0	0
42.2	20.0	17.8	0	0
60.0	37.8	0.2	0	1

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



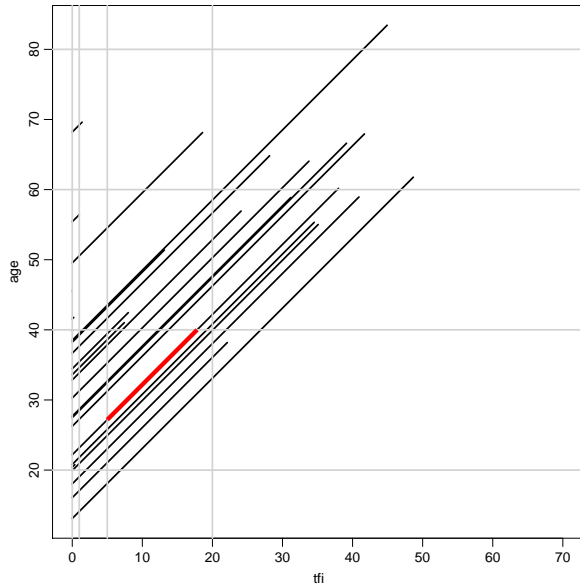
age	tfi	lex.dur	lex.Cst	lex.Xst
22.2	0.0	1.0	0	0
23.2	1.0	4.0	0	0
27.2	5.0	12.8	0	0
40.0	17.8	2.2	0	0
42.2	20.0	17.8	0	0
60.0	37.8	0.2	0	1

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



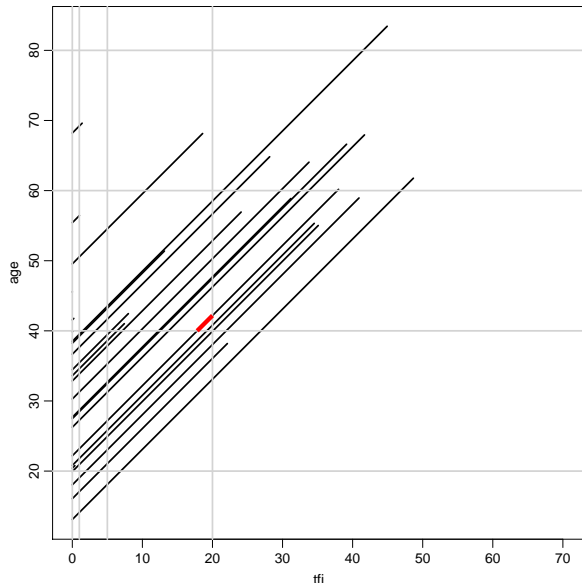
age	tfi	lex.dur	lex.Cst	lex.Xst
22.2	0.0	1.0	0	0
23.2	1.0	4.0	0	0
27.2	5.0	12.8	0	0
40.0	17.8	2.2	0	0
42.2	20.0	17.8	0	0
60.0	37.8	0.2	0	1

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

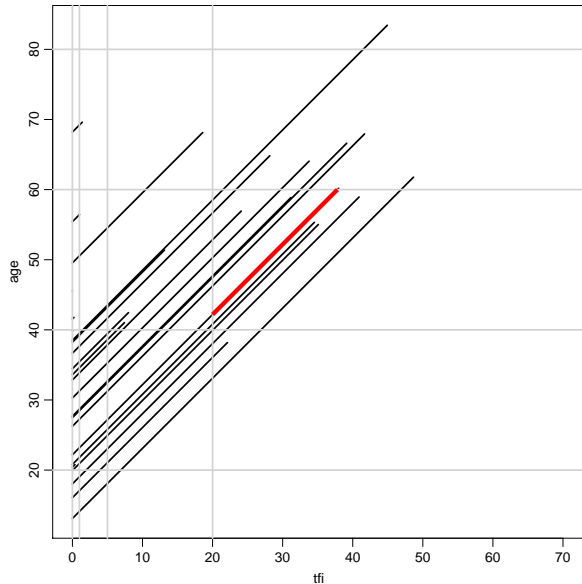
age	tfi	lex.dur	lex.Cst	lex.Xst
22.2	0.0	1.0	0	0
23.2	1.0	4.0	0	0
27.2	5.0	12.8	0	0
40.0	17.8	2.2	0	0
42.2	20.0	17.8	0	0
60.0	37.8	0.2	0	1

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



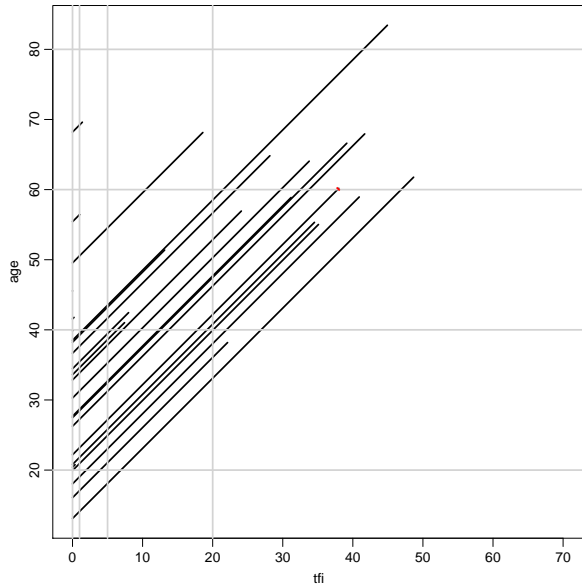
age	tfi	lex.dur	lex.Cst	lex.Xst
22.2	0.0	1.0	0	0
23.2	1.0	4.0	0	0
27.2	5.0	12.8	0	0
40.0	17.8	2.2	0	0
42.2	20.0	17.8	0	0
60.0	37.8	0.2	0	1

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



age	tfi	lex.dur	lex.Cst	lex.Xst
22.2	0.0	1.0	0	0
23.2	1.0	4.0	0	0
27.2	5.0	12.8	0	0
40.0	17.8	2.2	0	0
42.2	20.0	17.8	0	0
60.0	37.8	0.2	0	1

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



age	tfi	lex.dur	lex.Cst	lex.Xst
22.2	0.0	1.0	0	0
23.2	1.0	4.0	0	0
27.2	5.0	12.8	0	0
40.0	17.8	2.2	0	0
42.2	20.0	17.8	0	0
60.0	37.8	0.2	0	1

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

Splitting on several timescales

```
> spl1 <- splitLexis(thL , time.scale = "age", breaks = seq(0, 100, 20))  
> spl2 <- splitLexis(spl1, time.scale = "tfi", breaks = c(0, 1, 5, 20, 100))  
> summary(spl2)
```

Transitions:

To

From	0	1	Records:	Events:	Risk time:	Persons:	
	0	8250	1964	10214	1964	51934.08	2468

```
> library(popEpi)  
> splx <- splitMulti(thL, age = seq(0, 100, 20), tfi = c(0, 1, 5, 20, 100))  
> summary(splx)
```

Transitions:

To

From	0	1	Records:	Events:	Risk time:	Persons:	
	0	8248	1964	10212	1964	51916.98	2468

```
> # NOTE: splitMulti excludes follow-up outside range of breaks
```

Likelihood for time-split data

- ▶ We assume that rates are constant in each (small) intervals

Likelihood for time-split data

- ▶ We assume that rates are constant in each (small) intervals
- ▶ Each observation in the dataset represents an interval, contributing a term to the (log-)likelihood for the rate

Likelihood for time-split data

- ▶ We assume that rates are constant in each (small) intervals
- ▶ Each observation in the dataset represents an interval, contributing a term to the (log-)likelihood for the rate
- ▶ Each **term** looks like a contribution from a Poisson variate (albeit with values only 0 or 1)

Likelihood for time-split data

- ▶ We assume that rates are constant in each (small) intervals
- ▶ Each observation in the dataset represents an interval, contributing a term to the (log-)likelihood for the rate
- ▶ Each **term** looks like a contribution from a Poisson variate (albeit with values only 0 or 1)
- ▶ So the likelihood from a single **person** looks like the likelihood from several independent Poisson variates

Likelihood for time-split data

- ▶ We assume that rates are constant in each (small) intervals
- ▶ Each observation in the dataset represents an interval, contributing a term to the (log-)likelihood for the rate
- ▶ Each **term** looks like a contribution from a Poisson variate (albeit with values only 0 or 1)
- ▶ So the likelihood from a single **person** looks like the likelihood from several independent Poisson variates
- ▶ ...but the data are neither independent nor Poisson

Analysis of time-split data

Observations (records) classified by p —person and i —interval

► d_{pi} — events in the variable: `lex.Xst & lex.Xst!=lex.Cst`

Analysis of time-split data

Observations (records) classified by p —person and i —interval

- ▶ d_{pi} — events in the variable: `lex.Xst & lex.Xst!=lex.Cst`
- ▶ y_{pi} — risk time: `lex.dur` (duration)

Analysis of time-split data

Observations (records) classified by p —person and i —interval

- ▶ d_{pi} — events in the variable: `lex.Xst & lex.Xst!=lex.Cst`
- ▶ y_{pi} — risk time: `lex.dur` (duration)
- ▶ Covariates are:

Analysis of time-split data

Observations (records) classified by p —person and i —interval

- ▶ d_{pi} — events in the variable: `lex.Xst & lex.Xst!=lex.Cst`
- ▶ y_{pi} — risk time: `lex.dur` (duration)
- ▶ Covariates are:
 - ▶ timescales (age, period, time in study)

Analysis of time-split data

Observations (records) classified by p —person and i —interval

- ▶ d_{pi} — events in the variable: `lex.Xst & lex.Xst!=lex.Cst`
- ▶ y_{pi} — risk time: `lex.dur` (duration)
- ▶ Covariates are:
 - ▶ timescales (age, period, time in study)
 - ▶ other variables for this person (constant in each interval).

Analysis of time-split data

Observations (records) classified by p —person and i —interval

- ▶ d_{pi} — events in the variable: `lex.Xst & lex.Xst!=lex.Cst`
- ▶ y_{pi} — risk time: `lex.dur` (duration)
- ▶ Covariates are:
 - ▶ timescales (age, period, time in study)
 - ▶ other variables for this person (constant in each interval).
- ▶ Likelihood for rates for one person is identical to a Poisson likelihood for many independent Poisson variates

Analysis of time-split data

Observations (records) classified by p —person and i —interval

- ▶ d_{pi} — events in the variable: `lex.Xst & lex.Xst!=lex.Cst`
- ▶ y_{pi} — risk time: `lex.dur` (duration)
- ▶ Covariates are:
 - ▶ timescales (age, period, time in study)
 - ▶ other variables for this person (constant in each interval).
- ▶ Likelihood for rates for one person is identical to a Poisson likelihood for many independent Poisson variates
- ▶ Modeling rates using `glm` or `gam`:
time-scales and other covariates are treated alike

Fitting a simple model—data:

```
> stat.table(contrast,  
+           list(D = sum(lex.Xst),  
+               Y = sum(lex.dur),  
+               Rate = ratio(lex.Xst, lex.dur, 100)),  
+           margin = TRUE,  
+           data = spl2)
```

contrast	D	Y	Rate
1	928.00	20094.74	4.62
2	1036.00	31839.35	3.25
Total	1964.00	51934.08	3.78

Fitting a simple model

contrast	D	Y	Rate
1	928.00	20094.74	4.62
2	1036.00	31839.35	3.25

```
> m0 <- glm((lex.Xst==1) ~ factor(contrast) - 1,  
+           offset = log(lex.dur / 100),  
+           family = poisson,  
+           data = spl2)  
> round(ci.exp(m0), 2)
```

	exp(Est.)	2.5%	97.5%
factor(contrast)1	4.62	4.33	4.93
factor(contrast)2	3.25	3.06	3.46

... a Poisson model for mortality using log-person-years as offset

Fitting a simple model

contrast	D	Y	Rate
1	928.00	20094.74	4.62
2	1036.00	31839.35	3.25

```
> m0 <- glm(cbind(lex.Xst, lex.dur / 100) ~ factor(contrast) - 1,  
+           family = poisreg,  
+           data = spl2)  
> round(ci.exp(m0), 2)
```

	exp(Est.)	2.5%	97.5%
factor(contrast)1	4.62	4.33	4.93
factor(contrast)2	3.25	3.06	3.46

... a Poisson model for mortality rates based on deaths and person-years

Fitting a simple model

The wrapper `glm.Lexis` requires that `lex.Cst` and `lex.Xst` are factors—use `factorize` to make them:

```
> splf <- factorize(spl2)
> m0 <- glm.Lexis(splf, ~ factor(contrast) - 1, scale = 100)

stats::glm Poisson analysis of Lexis object splf with log link:
Rates for the transition:
0->1
, lex.dur (person-time) scaled by 100
> round(ci.exp(m0), 2)
```

	exp(Est.)	2.5%	97.5%
factor(contrast)1	4.62	4.33	4.93
factor(contrast)2	3.25	3.06	3.46

... a Poisson model for mortality rates based on deaths and person-years in a `Lexis` object

Fitting a simple model — aggregate data

contrast	D	Y	Rate
1	928.00	20094.74	4.62
2	1036.00	31839.35	3.25

As long as we only use covariates that take only a few values, we can model the aggregate data directly:

```
> mx <- glm(cbind(c(928, 1036), c(20094.74, 31839.35) / 100) ~ factor(1:2) - 1,  
+           family = poisreg )  
> round(ci.exp(mx), 2)
```

	exp(Est.)	2.5%	97.5%
factor(1:2)1	4.62	4.33	4.93
factor(1:2)2	3.25	3.06	3.46

SMR

Bendix Carstensen

Representation of follow-up

SPE, Tartu, Estonia,

June 2023

<http://BendixCarstensen.com/SPE>

Cohorts where all are exposed

When there is no comparison group we may ask:

Do mortality rates in cohort differ from those of an **external** population, for example:

Cohorts where all are exposed

When there is no comparison group we may ask:

Do mortality rates in cohort differ from those of an **external** population, for example:

Rates from:

- ▶ Occupational cohorts

Cohorts where all are exposed

When there is no comparison group we may ask:

Do mortality rates in cohort differ from those of an **external** population, for example:

Rates from:

- ▶ Occupational cohorts
- ▶ Patient cohorts

Cohorts where all are exposed

When there is no comparison group we may ask:

Do mortality rates in cohort differ from those of an **external** population, for example:

Rates from:

- ▶ Occupational cohorts
- ▶ Patient cohorts

Cohorts where all are exposed

When there is no comparison group we may ask:

Do mortality rates in cohort differ from those of an **external** population, for example:

Rates from:

- ▶ Occupational cohorts
- ▶ Patient cohorts

compared with reference rates obtained from:

- ▶ Population statistics (mortality rates)

Cohorts where all are exposed

When there is no comparison group we may ask:

Do mortality rates in cohort differ from those of an **external** population, for example:

Rates from:

- ▶ Occupational cohorts
- ▶ Patient cohorts

compared with reference rates obtained from:

- ▶ Population statistics (mortality rates)
- ▶ Hospital registers (disease rates)

Cohort rates vs. population rates: RSR

► **Additive:** $\lambda(a) = \delta(a) + \lambda_p(a)$

Cohort rates vs. population rates: RSR

- ▶ **Additive:** $\lambda(a) = \delta(a) + \lambda_p(a)$
- ▶ Note that the survival (since $a = a_0$, say) is:

$$\begin{aligned} S(a) &= \exp\left(-\int_{a_0}^a \delta(a) + \lambda_p(a) \, da\right) \\ &= \exp\left(-\int_{a_0}^a \delta(a) \, da\right) \times S_p(a) \end{aligned}$$

$$\Rightarrow \quad r(a) = S(a)/S_p(a) = \exp\left(-\int_{a_0}^a \delta(a) \, da\right)$$

Cohort rates vs. population rates: RSR

- ▶ **Additive:** $\lambda(a) = \delta(a) + \lambda_p(a)$
- ▶ Note that the survival (since $a = a_0$, say) is:

$$S(a) = \exp\left(-\int_{a_0}^a \delta(a) + \lambda_p(a) da\right)$$

$$= \exp\left(-\int_{a_0}^a \delta(a) da\right) \times S_p(a)$$

$$\Rightarrow r(a) = S(a)/S_p(a) = \exp\left(-\int_{a_0}^a \delta(a) da\right)$$

- ▶ **Additive** model for **rates** \Leftrightarrow **Relative survival** model.

Cohort rates vs. population rates: SMR

► **Multiplicative:** $\lambda(a) = \theta \times \lambda_p(a)$

Cohort rates vs. population rates: SMR

- ▶ **Multiplicative:** $\lambda(a) = \theta \times \lambda_p(a)$
- ▶ Cohort rates proportional to reference rates, λ_p :
 $\lambda(a) = \theta \times \lambda_p(a)$ — θ the same in all age-bands.

Cohort rates vs. population rates: SMR

- ▶ **Multiplicative:** $\lambda(a) = \theta \times \lambda_p(a)$
- ▶ Cohort rates proportional to reference rates, λ_p :
 $\lambda(a) = \theta \times \lambda_p(a)$ — θ the same in all age-bands.
- ▶ D_a deaths during Y_a person-years an age-band a gives the likelihood:

$$\begin{aligned} D_a \log(\lambda(a)) - \lambda(a) Y_a &= D_a \log(\theta \lambda_p(a)) - \theta \lambda_p(a) Y_a \\ &= D_a \log(\theta) + D_a \log(\lambda_p(a)) - \theta (\lambda_p(a) Y_a) \end{aligned}$$

Cohort rates vs. population rates: SMR

- ▶ **Multiplicative:** $\lambda(a) = \theta \times \lambda_p(a)$
- ▶ Cohort rates proportional to reference rates, λ_p :
 $\lambda(a) = \theta \times \lambda_p(a)$ — θ the same in all age-bands.
- ▶ D_a deaths during Y_a person-years an age-band a gives the likelihood:

$$\begin{aligned} D_a \log(\lambda(a)) - \lambda(a) Y_a &= D_a \log(\theta \lambda_p(a)) - \theta \lambda_p(a) Y_a \\ &= D_a \log(\theta) + D_a \log(\lambda_p(a)) - \theta (\lambda_p(a) Y_a) \end{aligned}$$

- ▶ The constant $D_a \log(\lambda_p(a))$ does not involve θ , and so can be dropped.

- $\lambda_p(a)Y_a = E_a$ is the “expected” number of cases in age a , so the log-likelihood contribution from age a is:

$$D_a \log(\theta) - \theta(\lambda_p(a)Y_a) = D_a \log(\theta) - \theta(E_a)$$

- ▶ $\lambda_p(a)Y_a = E_a$ is the “expected” number of cases in age a , so the log-likelihood contribution from age a is:

$$D_a \log(\theta) - \theta(\lambda_p(a)Y_a) = D_a \log(\theta) - \theta(E_a)$$

- ▶ The log-likelihood is similar to the log-likelihood for a rate, so:

$$\hat{\theta} = \sum_a D_a / \sum_a E_a = \text{Observed/Expected} = \text{SMR}$$

Modeling the SMR in practice

- ▶ As for the rates, the SMR can be modelled using individual data.

Modeling the SMR in practice

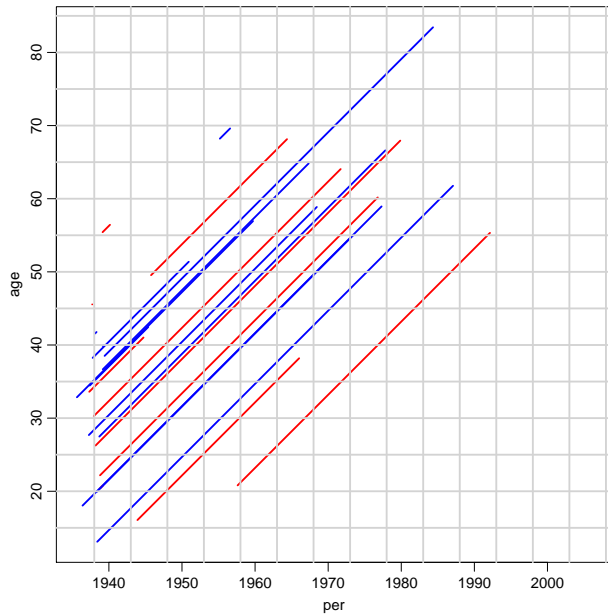
- ▶ As for the rates, the SMR can be modelled using individual data.
- ▶ Response is d_i , the event indicator (`lex.Xst`).

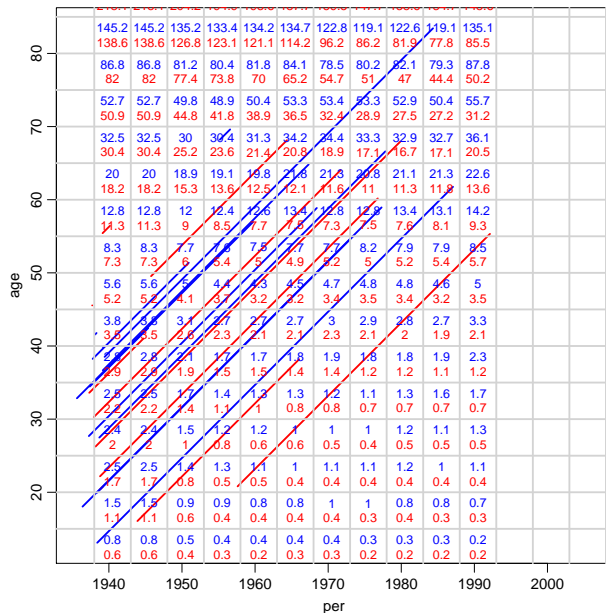
Modeling the SMR in practice

- ▶ As for the rates, the SMR can be modelled using individual data.
- ▶ Response is d_i , the event indicator (`lex.Xst`).
- ▶ log-offset is the expected value for each piece of follow-up,
$$e_i = y_i \times \lambda_p (\text{lex.dur} * \text{rate})$$

Modeling the SMR in practice

- ▶ As for the rates, the SMR can be modelled using individual data.
- ▶ Response is d_i , the event indicator (`lex.Xst`).
- ▶ log-offset is the expected value for each piece of follow-up,
 $e_i = y_i \times \lambda_p$ (`lex.dur * rate`)
- ▶ λ_p is the population rate corresponding to the age, period and sex of the follow-up period y_i .





Split the data to fit with population data

```
> thad <- splitMulti(thL, age=seq(0,90,5), dte=seq(1938,2038,5) )  
> summary( thad )
```

Transitions:

To

From	0	1	Records:	Events:	Risk time:	Persons:
0	21059	1939	22998	1939	51787.96	2463

Split the data to fit with population data

```
> thad <- splitMulti(thL, age=seq(0,90,5), dte=seq(1938,2038,5) )  
> summary( thad )
```

Transitions:

To

From	0	1	Records:	Events:	Risk time:	Persons:	
	0	21059	1939	22998	1939	51787.96	2463

Create variables to fit with the population data

```
> thad$agr <- timeBand( thad, "age", "left" )  
> thad$per <- timeBand( thad, "dte", "left" )  
> round( thad[1:5,c("lex.id","age","agr","dte","per","lex.dur","lex.Xst","sex")],
```

lex.id	age	dte	lex.dur	lex.Xst	agr	per	sex
1	22.18	1938.79	2.82	0	20	1938	2
1	25.00	1941.61	1.39	0	25	1938	2
1	26.39	1943.00	3.61	0	25	1943	2
1	30.00	1946.61	1.39	0	30	1943	2
1	31.39	1948.00	3.61	0	30	1948	2

```

> data( gmortDK )
> dim( gmortDK )

[1] 418  21

> gmortDK[1:6,1:6]

  agr per sex  risk    dt    rt
1   0  38   1 996019 14079 14.135
2   5  38   1 802334   726  0.905
3  10  38   1 753017   600  0.797
4  15  38   1 773393  1167  1.509
5  20  38   1 813882  2031  2.495
6  25  38   1 789990  1862  2.357

> gmortDK$per <- gmortDK$per+1900
> #
> thadx <- merge( thad, gmortDK[,c("agr","per","sex","rt")] )
> #
> thadx$E <- thadx$lex.dur * thadx$rt / 1000

```

```

> stat.table(contrast,
+             list( D = sum(lex.Xst),
+                   Y = sum(lex.dur),
+                   E = sum(E),
+                   SMR = ratio(lex.Xst, E)),
+             margin = TRUE,
+             data = thadx)

```

contrast	D	Y	E	SMR
1	917.00	20045.46	214.66	4.27
2	1022.00	31742.51	447.21	2.29
Total	1939.00	51787.96	661.87	2.93

contrast	D	Y	E	SMR
1	917.00	20045.46	214.66	4.27
2	1022.00	31742.51	447.21	2.29

```
> m.SMR <- glm(cbind(lex.Xst, E) ~ factor(contrast) - 1,
+             family = poisreg,
+             data = thadx)
> round(ci.exp(m.SMR), 2)
```

	exp(Est.)	2.5%	97.5%
factor(contrast)1	4.27	4.00	4.56
factor(contrast)2	2.29	2.15	2.43

contrast	D	Y	E	SMR
1	917.00	20045.46	214.66	4.27
2	1022.00	31742.51	447.21	2.29

```
> m.SMR <- glm(cbind(lex.Xst, E) ~ factor(contrast) - 1,
+             family = poisreg,
+             data = thadx)
> round(ci.exp(m.SMR), 2)
```

	exp(Est.)	2.5%	97.5%
factor(contrast)1	4.27	4.00	4.56
factor(contrast)2	2.29	2.15	2.43

► Analysis of SMR is like analysis of rates:

contrast	D	Y	E	SMR
1	917.00	20045.46	214.66	4.27
2	1022.00	31742.51	447.21	2.29

```
> m.SMR <- glm(cbind(lex.Xst, E) ~ factor(contrast) - 1,
+             family = poisreg,
+             data = thadx)
> round(ci.exp(m.SMR), 2)
```

```
              exp(Est.) 2.5% 97.5%
factor(contrast)1      4.27 4.00  4.56
factor(contrast)2      2.29 2.15  2.43
```

- Analysis of SMR is like analysis of rates:
- Replace Y with E — that's all! (`glm.Lexis` not usable)

contrast	D	Y	E	SMR
1	917.00	20045.46	214.66	4.27
2	1022.00	31742.51	447.21	2.29

```
> m.SMR <- glm(cbind(lex.Xst, E) ~ factor(contrast) - 1,
+             family = poisreg,
+             data = thadx)
> round(ci.exp(m.SMR), 2)
```

```
              exp(Est.) 2.5% 97.5%
factor(contrast)1      4.27 4.00  4.56
factor(contrast)2      2.29 2.15  2.43
```

- Analysis of SMR is like analysis of rates:
- Replace Y with E — that's all! (`glm.Lexis` not usable)
- ...it's the calculation of E that is difficult