Estimation and prediction in multistate models

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

University of Aberdeen,
17 August 2017

http://BendixCarstensen/AdvCoh/Aberdeen-2017

Diabetologia
DOI 10.1007/s00125-016-4065-6

CrossMark

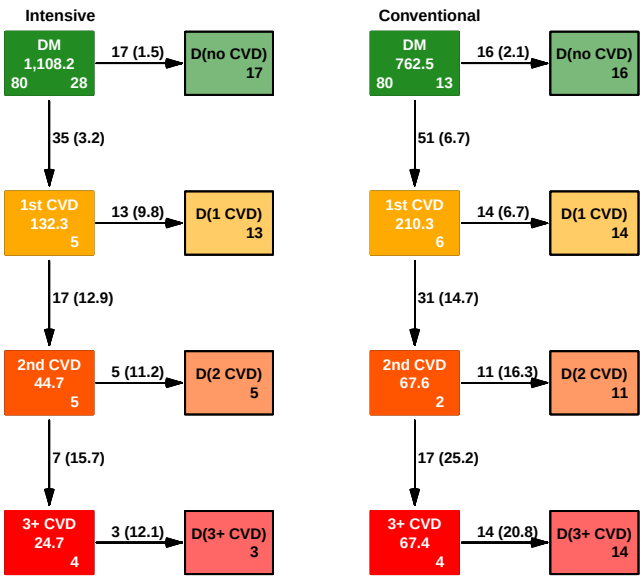
ARTICLE

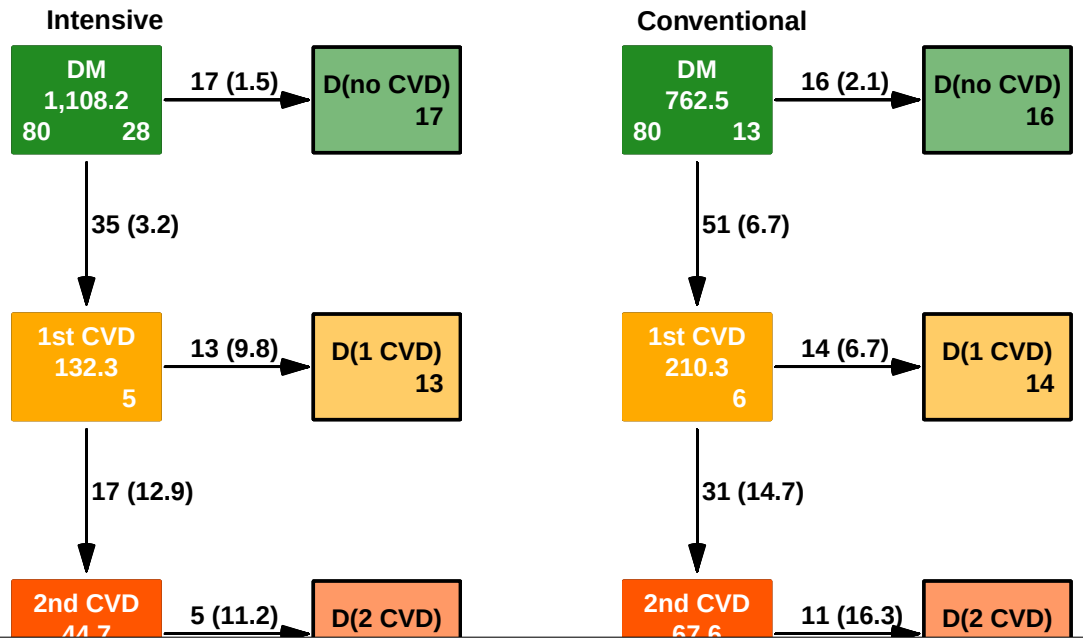
Years of life gained by multifactorial intervention in patients with type 2 diabetes mellitus and microalbuminuria: 21 years follow-up on the Steno-2 randomised trial

Peter Gæde[1,2] · Jens Oellgaard[1,2,3] · Bendix Carstensen[3] · Peter Rossing[3,4,5] · Henrik Lund-Andersen[3,5,6] · Hans-Henrik Parving[5,7] · Oluf Pedersen[8]

Received: 7 April 2016 / Accepted: 1 July 2016

Abstract
Aims/hypothesis The aim of this work was to study the potential long-term impact of a 7.8 years intensified multifactorial pharmacological approaches. After 7.8 years the study continued as an observational follow-up with all patients receiving treatment as for the original intensive-therapy group. The pri-





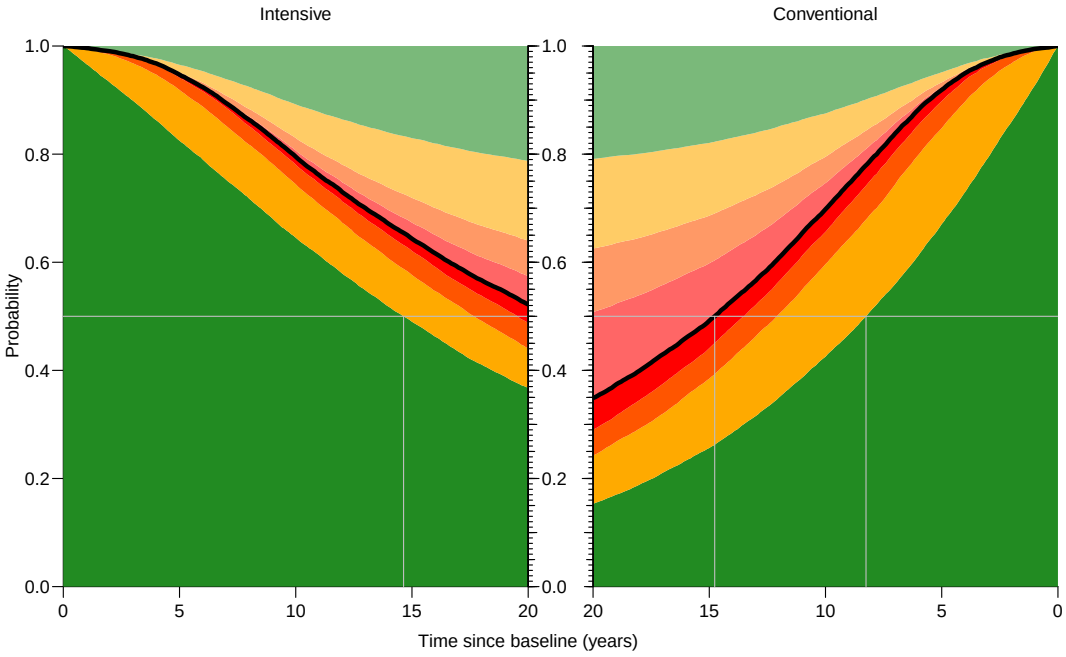
Models used

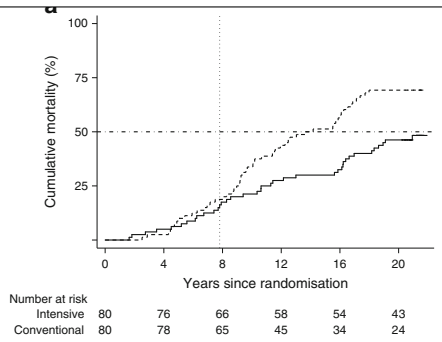
- One model for the 4 mortality rates
- One model for the 3 CVD rates
- . . . both models assume:
  - proportional hazards between CVD states $(0, 1, 2(, 3)$ CVD events)
  - proportional hazards between groups (conventional, intervention)
  - proportional hazards between levels of sex and age
- Which just means: multiplicative effects of the covariates: time since baseline, CVD state, group, sex and age
- **Proportional hazards** means: no interaction with the time scale

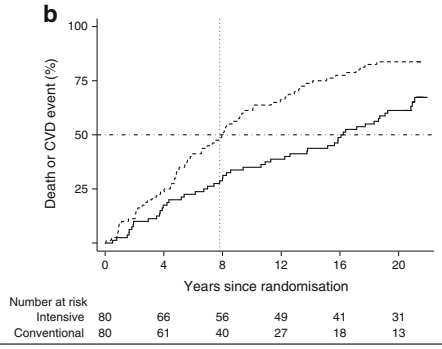
Hazard ratios

	Mortality	CVD event
HR, Int. vs. Conv.	0.83 (0.54; 1.30)	0.55 (0.39;0.77)
H_0: PH btw. CVD groups	p=0.438	p=0.261
H_0: HR = 1	p=0.425	p=0.001
HR vs. 0 CVD events:		
0 (ref.)	1.00	1.00
1	3.08 (1.82; 5.19)	2.43 (1.67;3.52)
2	4.42 (2.36; 8.29)	3.48 (2.15;5.64)
3+	7.76 (4.11;14.65)	

Then use fitted rates to estimate the probabilities of being in each state at all times. (This is immensely complicated).







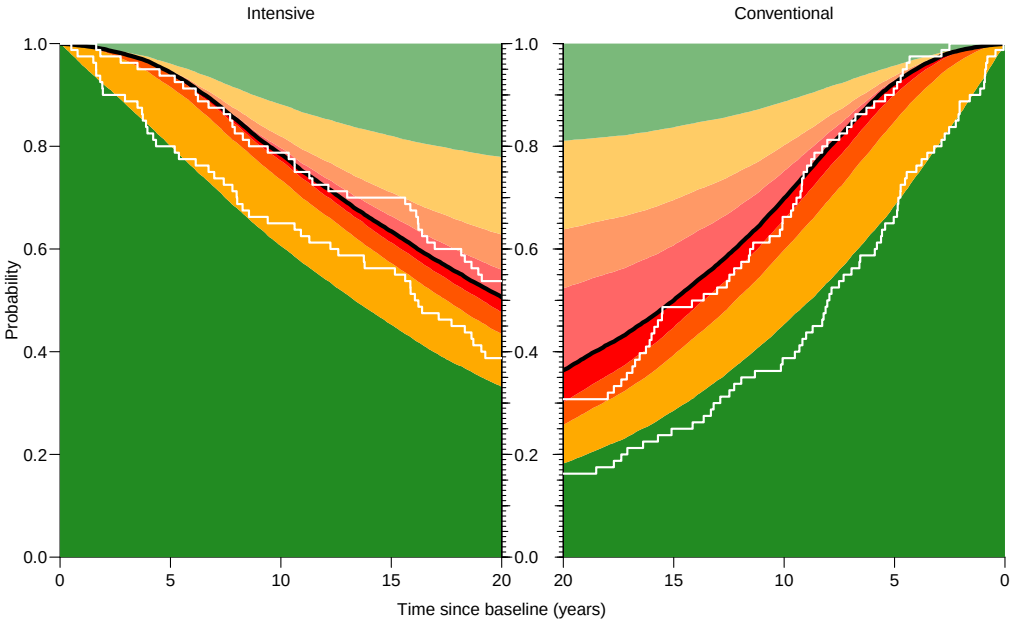
between groups (HR 0.83 [95% CI 0.54, 1.30], $p=0.43$). Thus, the reduced mortality was primarily due to reduced risk of CVD.

The patients in the intensive group experienced a total of 90 cardiovascular events vs 195 events in the conventional group. Nineteen intensive-group patients (24%) vs 34 conventional-group patients (43%) experienced more than one cardiovascular event. No significant between-group difference in the distribution of specific cardiovascular first-event types was observed (Table 2 and Fig. 4).

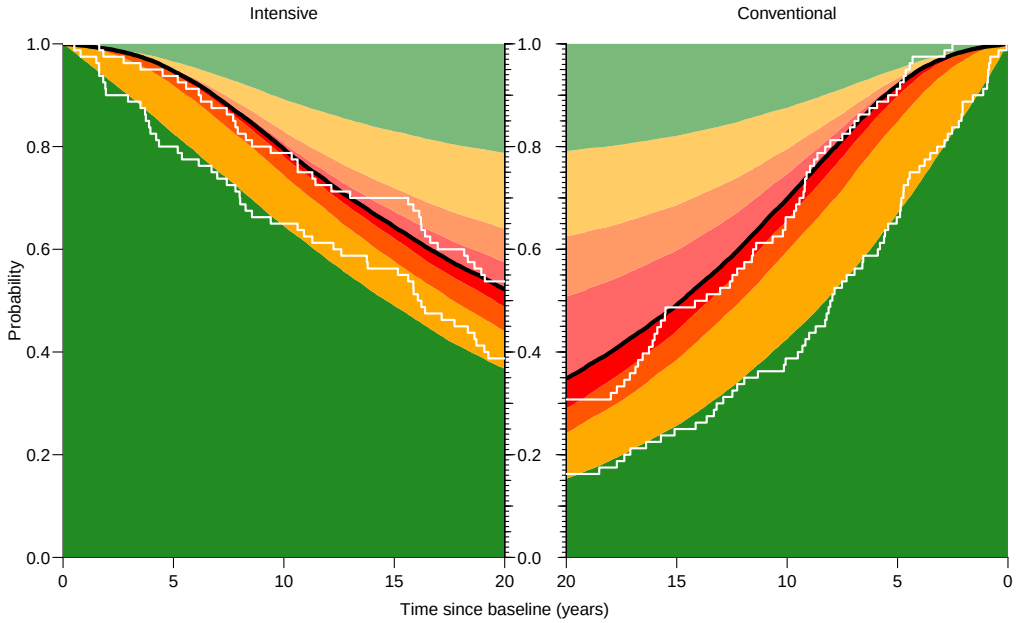
Microvascular complications Hazard rates of progression rates in microvascular complications compared with baseline status are shown Fig. 3. Sensitivity analyses showed a negligible effect of the random dates imputation.

Progression of retinopathy was decreased by 33% in the intensive-therapy group (Fig. 5). Blindness in at least one eye was reduced in the intensive-therapy group with an HR of 0.47 (95% CI 0.23, 0.98, $p=0.044$). Autonomic neuropathy was decreased by 41% in the intensive-therapy group (Fig. 5). We observed no difference between groups in the progression of peripheral neuropathy (Fig. 5). Progression to diabetic nephropathy (macroalbuminuria) was reduced by 48% in the intensive-therapy group (Fig. 5). Ten patients in the conventional-therapy groups vs five patients in the intensive-therapy group progressed to end-stage renal disease ($p=0.061$).

Discussion



Same allocation HR for Death and CVD between CVD levels



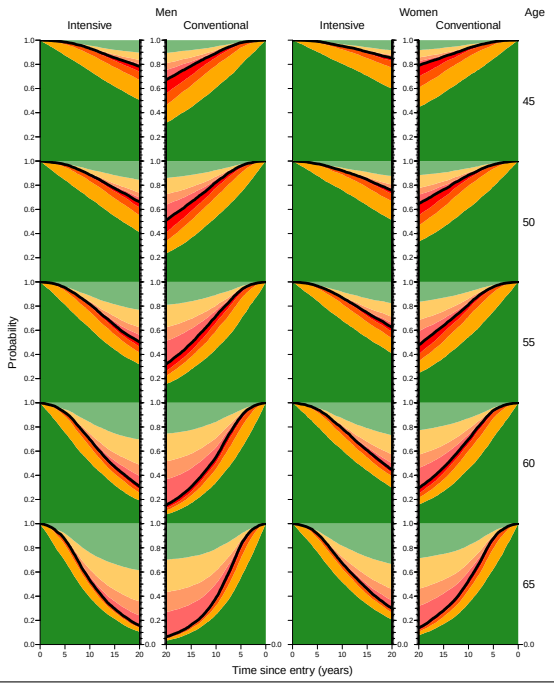
Different allocation HRs for Death and CVD between CVD levels

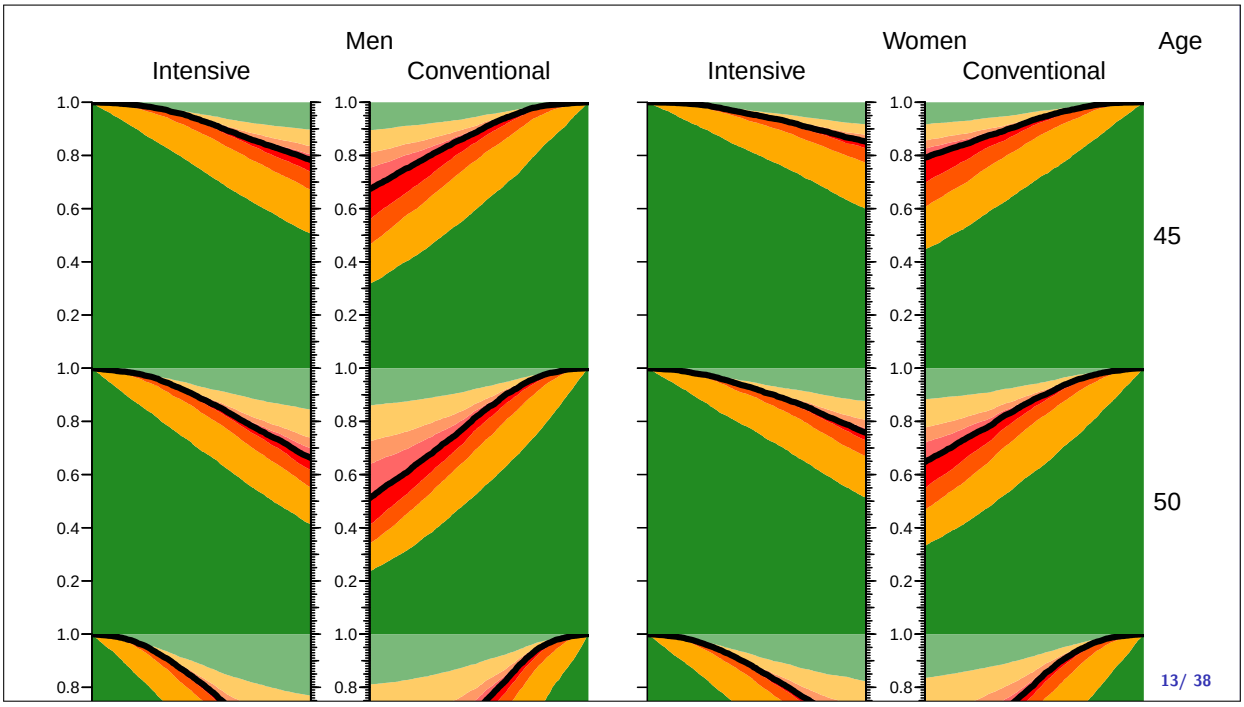
Expected lifetime and YLL (well, gained)

- Expected lifetime (years) in the Steno 2 cohort during the first 20 years after baseline by treatment group and CVD status.

State	where	Int.	Conv.	Int.−Conv.
Alive	under black line	15.6	14.1	1.5
No CVD	green area	12.7	10.0	2.6
Any CVD	orange area	3.0	4.1	−1.1

- What does “expected” mean?
- Expectation w.r.t. age and sex-distribution in the Steno 2 study!
- Computed as areas under survival curves



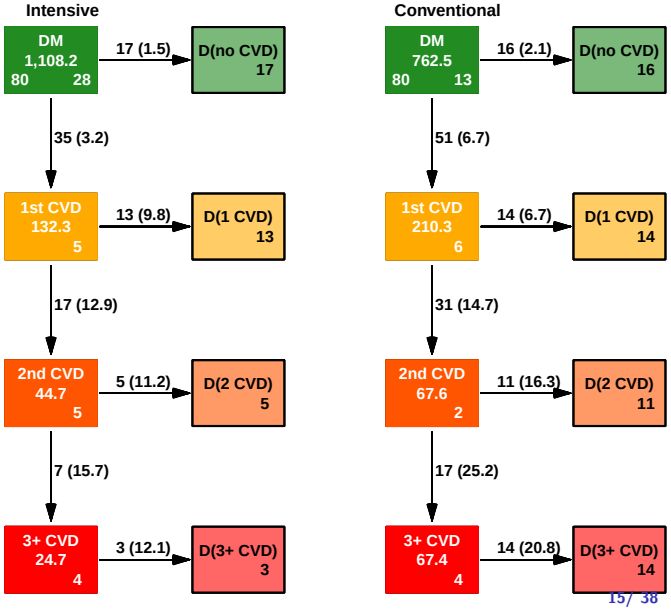


Expected lifetime (years) during the first 20 years after baseline by sex, age, treatment group and CVD status.

sex		Men			Women		
state	age	Int.	Conv.	Int.−Conv.	Int.	Conv.	Int.−Conv.
Alive	45	18.5	17.5	1.0	19.1	18.4	0.7
	50	17.2	16.1	1.1	18.0	17.2	0.8
	55	15.6	13.8	1.8	17.4	15.9	1.6
	60	13.9	11.6	2.2	15.5	13.7	1.8
	65	11.2	9.5	1.8	13.3	11.4	2.0
No CVD	45	14.9	12.5	2.4	15.8	14.3	1.5
	50	14.0	11.1	2.9	15.1	12.9	2.2
	55	12.2	9.7	2.5	14.3	11.6	2.7
	60	10.9	8.2	2.7	12.4	9.9	2.6
	65	9.0	6.7	2.2	10.7	8.3	2.4

Multistate models in practice:

- Representation:
  - States
  - Transitions
  - Sojourn times
  - Rates
- Analysis of rates:
  - Cox-model
  - Poisson model
- Reporting
  - Rates
  - HRs
  - Probabilities
  - Expected lifetime



Representation of multistate FU: Lexis

- Allowing multiple time scales
  - time-scale variables — the starting point on each time scale
  - sojourn time variable `lex.dur` — risk time, exposure
    — the same on all time scales
- Allowing multiple states requires state variables:
  - `lex.Cst` — the state **in** which follow-up (`lex.dur`) is
  - `lex.Xst` — the state **to** which transition occur

Representation of multistate FU: `Lexis`

Multiple records per person:
One record for each **transient** state (*i.e.* state with FU-time)

```
lex.id      per    age   dur   tsb lex.dur lex.Cst  lex.Xst   allocation sex
     5 1993.162 57.169 6.816 0.000   0.797      DM  1st CVD Conventional   M
     5 1993.959 57.966 7.613 0.797   0.698 1st CVD  2nd CVD Conventional   M
     5 1994.657 58.664 8.311 1.495   3.389 2nd CVD D(2 CVD) Conventional   M
```


Likelihood for transition through states

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

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

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

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

Likelihood contributions reflected in `Lexis` object

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

lex.id	time	lex.dur	lex.Cst	lex.Xst
1	t_0	t_B-t_0	A	B
1	t_B	t_C-t_B	B	C
1	t_C	t_x-t_C	C	C

constant rate in interval $\Rightarrow$ log-likelihood term is Poisson:
$d\log(\lambda) - \lambda y = (\texttt{lex.Xst!} = \texttt{lex.Cst}) \times \log(\lambda) - \lambda \times \texttt{lex.dur}$

Likelihood for multiple states

- **Product** of likelihoods for each state
  — each one as for a survival model
- **conditional** on being alive at (observed) entry to current state
- **Risk time** is the risk time in the **C**urrent (`lex.Cst`) state
- **Events** are transitions to the e**X**it state (`lex.Xst`)
- All other transitions **out** of `lex.Cst` are treated as **censorings** (but they are not)
- Fit models separately for each transition
- . . . or jointly for all or some
  — may require restructuring of data

Analysis of rates in multistate models

- Each transition modeled:
  - Cox model
  - Poisson model with log-PY as offset
- either one requires that you decide on a time-scale:
  age / time since study start / time since current state. . .
- Poisson model allows smooth baseline hazards
- — requires that follow-up is split in smaller pieces
  (so small that the assumption of constant rates is reasonable)
- — also allows modeling of several time scales simultaneously
- — simple to access baseline hazard without further ado

Representation of multistate FU: Lexis I

Using `splitLexis` to obtain:

lex.id	per	age	dur	tsb	lex.dur	lex.Cst	lex.Xst	allocation	sex
5	1993.162	57.169	6.816	0.000	0.083	DM	DM	Conventional	M
5	1993.246	57.252	6.899	0.083	0.083	DM	DM	Conventional	M
5	1993.329	57.336	6.983	0.167	0.083	DM	DM	Conventional	M
5	1993.412	57.419	7.066	0.250	0.083	DM	DM	Conventional	M
5	1993.496	57.502	7.149	0.333	0.083	DM	DM	Conventional	M
5	1993.579	57.586	7.233	0.417	0.083	DM	DM	Conventional	M
5	1993.662	57.669	7.316	0.500	0.083	DM	DM	Conventional	M
5	1993.746	57.752	7.399	0.583	0.083	DM	DM	Conventional	M
5	1993.829	57.836	7.483	0.667	0.083	DM	DM	Conventional	M
5	1993.912	57.919	7.566	0.750	0.047	DM	1st CVD	Conventional	M
5	1993.959	57.966	7.613	0.797	0.037	1st CVD	1st CVD	Conventional	M
5	1993.996	58.002	7.649	0.833	0.083	1st CVD	1st CVD	Conventional	M
5	1994.079	58.086	7.733	0.917	0.083	1st CVD	1st CVD	Conventional	M
5	1994.162	58.169	7.816	1.000	0.083	1st CVD	1st CVD	Conventional	M

Representation of multistate FU: Lexis II

5	1994.246	58.252	7.899	1.083	0.083	1st CVD	1st CVD	Conventional	M
5	1994.329	58.336	7.983	1.167	0.083	1st CVD	1st CVD	Conventional	M
5	1994.412	58.419	8.066	1.250	0.083	1st CVD	1st CVD	Conventional	M
5	1994.496	58.502	8.149	1.333	0.083	1st CVD	1st CVD	Conventional	M
5	1994.579	58.586	8.233	1.417	0.078	1st CVD	2nd CVD	Conventional	M
5	1994.657	58.664	8.311	1.495	0.005	2nd CVD	2nd CVD	Conventional	M
5	1994.662	58.669	8.316	1.500	0.083	2nd CVD	2nd CVD	Conventional	M
5	1994.746	58.752	8.399	1.583	0.083	2nd CVD	2nd CVD	Conventional	M
5	1994.829	58.836	8.483	1.667	0.083	2nd CVD	2nd CVD	Conventional	M
5	1994.912	58.919	8.566	1.750	0.083	2nd CVD	2nd CVD	Conventional	M
5	1994.996	59.002	8.649	1.833	0.083	2nd CVD	2nd CVD	Conventional	M
5	1995.079	59.086	8.733	1.917	0.083	2nd CVD	2nd CVD	Conventional	M
5	1995.162	59.169	8.816	2.000	0.083	2nd CVD	2nd CVD	Conventional	M
5	1995.246	59.252	8.899	2.083	0.083	2nd CVD	2nd CVD	Conventional	M
5	1995.329	59.336	8.983	2.167	0.083	2nd CVD	2nd CVD	Conventional	M
5	1995.412	59.419	9.066	2.250	0.083	2nd CVD	2nd CVD	Conventional	M
5	1995.496	59.502	9.149	2.333	0.083	2nd CVD	2nd CVD	Conventional	M
5	1995.579	59.586	9.233	2.417	0.083	2nd CVD	2nd CVD	Conventional	M
5	1995.662	59.669	9.316	2.500	0.083	2nd CVD	2nd CVD	Conventional	M

Representation of multistate FU: Lexis III

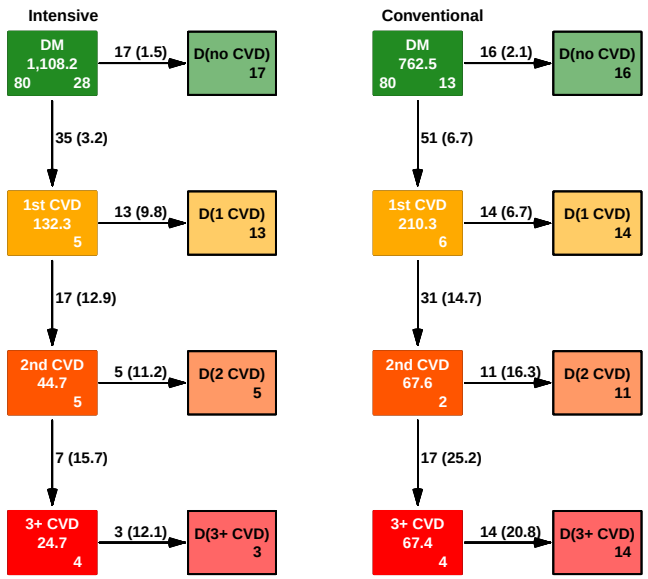
5	1995.746	59.752	9.399	2.583	0.083	2nd CVD	2nd CVD	Conventional	M
5	1995.829	59.836	9.483	2.667	0.083	2nd CVD	2nd CVD	Conventional	M
5	1995.912	59.919	9.566	2.750	0.083	2nd CVD	2nd CVD	Conventional	M
5	1995.996	60.002	9.649	2.833	0.083	2nd CVD	2nd CVD	Conventional	M
5	1996.079	60.086	9.733	2.917	0.083	2nd CVD	2nd CVD	Conventional	M
5	1996.162	60.169	9.816	3.000	0.083	2nd CVD	2nd CVD	Conventional	M
5	1996.246	60.252	9.899	3.083	0.083	2nd CVD	2nd CVD	Conventional	M
5	1996.329	60.336	9.983	3.167	0.083	2nd CVD	2nd CVD	Conventional	M
5	1996.412	60.419	10.066	3.250	0.083	2nd CVD	2nd CVD	Conventional	M
5	1996.496	60.502	10.149	3.333	0.083	2nd CVD	2nd CVD	Conventional	M
5	1996.579	60.586	10.233	3.417	0.083	2nd CVD	2nd CVD	Conventional	M
5	1996.662	60.669	10.316	3.500	0.083	2nd CVD	2nd CVD	Conventional	M
5	1996.746	60.752	10.399	3.583	0.083	2nd CVD	2nd CVD	Conventional	M
5	1996.829	60.836	10.483	3.667	0.083	2nd CVD	2nd CVD	Conventional	M
5	1996.912	60.919	10.566	3.750	0.083	2nd CVD	2nd CVD	Conventional	M
5	1996.996	61.002	10.649	3.833	0.083	2nd CVD	2nd CVD	Conventional	M
5	1997.079	61.086	10.733	3.917	0.083	2nd CVD	2nd CVD	Conventional	M
5	1997.162	61.169	10.816	4.000	0.083	2nd CVD	2nd CVD	Conventional	M
5	1997.246	61.252	10.899	4.083	0.083	2nd CVD	2nd CVD	Conventional	M

Representation of multistate FU: Lexis IV

```
5 1997.329 61.336 10.983 4.167   0.083 2nd CVD  2nd CVD Conventional   M
5 1997.412 61.419 11.066 4.250   0.083 2nd CVD  2nd CVD Conventional   M
5 1997.496 61.502 11.149 4.333   0.083 2nd CVD  2nd CVD Conventional   M
5 1997.579 61.586 11.233 4.417   0.083 2nd CVD  2nd CVD Conventional   M
5 1997.662 61.669 11.316 4.500   0.083 2nd CVD  2nd CVD Conventional   M
5 1997.746 61.752 11.399 4.583   0.083 2nd CVD  2nd CVD Conventional   M
5 1997.829 61.836 11.483 4.667   0.083 2nd CVD  2nd CVD Conventional   M
5 1997.912 61.919 11.566 4.750   0.083 2nd CVD  2nd CVD Conventional   M
5 1997.996 62.002 11.649 4.833   0.051 2nd CVD D(2 CVD) Conventional   M
```

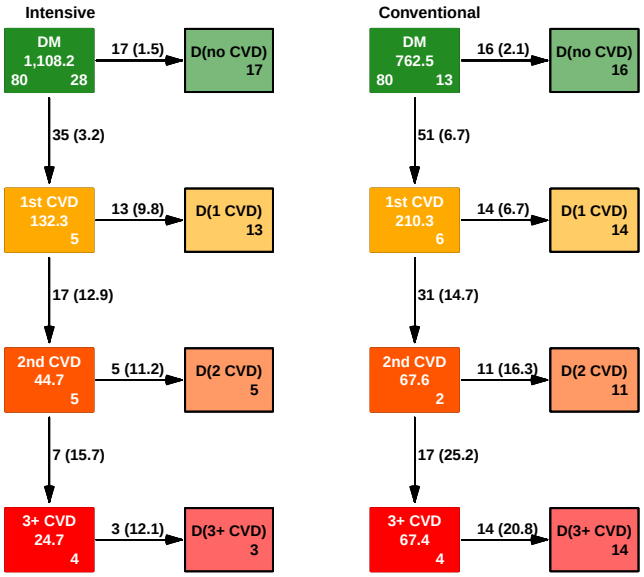

Representation of multistate FU: Lexis

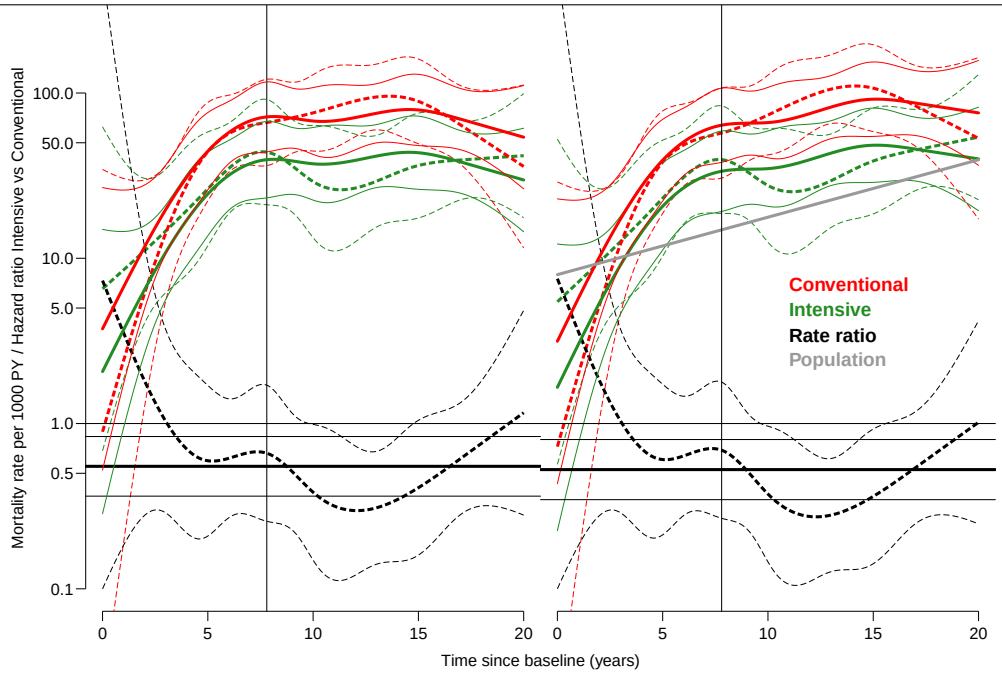
```
lex.id      per    age    dur   tsb lex.dur lex.Cst  lex.Xst   allocation sex
     5 1993.162 57.169  6.816 0.000   0.083      DM       DM Conventional   M
     5 1993.246 57.252  6.899 0.083   0.083      DM       DM Conventional   M
   ...
     5 1993.829 57.836  7.483 0.667   0.083      DM       DM Conventional   M
     5 1993.912 57.919  7.566 0.750   0.047      DM  1st CVD Conventional   M
     5 1993.959 57.966  7.613 0.797   0.037 1st CVD  1st CVD Conventional   M
   ...
     5 1994.496 58.502  8.149 1.333   0.083 1st CVD  1st CVD Conventional   M
     5 1994.579 58.586  8.233 1.417   0.078 1st CVD  2nd CVD Conventional   M
     5 1994.657 58.664  8.311 1.495   0.005 2nd CVD  2nd CVD Conventional   M
   ...
     5 1994.746 58.752  8.399 1.583   0.083 2nd CVD  2nd CVD Conventional   M
     5 1994.829 58.836  8.483 1.667   0.083 2nd CVD  2nd CVD Conventional   M
   ...
     5 1997.912 61.919 11.566 4.750   0.083 2nd CVD  2nd CVD Conventional   M
     5 1997.996 62.002 11.649 4.833   0.051 2nd CVD D(2 CVD) Conventional   M
```

Modeling mortality rates in `Lexis` objects

```
> dlev <- c("D(no CVD)", "D(1 CVD)",  "D(2 CVD)",  "D(3+ CVD)")
> #
> m0 <- glm( (lex.Xst %in% dlev ) & (lex.Xst!=lex.Cst) ~
+            Ns( tsb, knots=d.kn ) + lex.Cst + allocation + sex + age,
+            offset = log(lex.dur),
+            family = poisson,
+            data = S1 )
> #
> m0i <- update( m0, . ~ . + allocation:lex.Cst )
> #
> m0n <- update( m0, . ~ . + allocation:tsb )
> #
> # Test for interactions (PropHaz)
> anova( m0i, m0, m0n, test="Chisq" )
```



Modeling CVD rates in `Lexis` objects

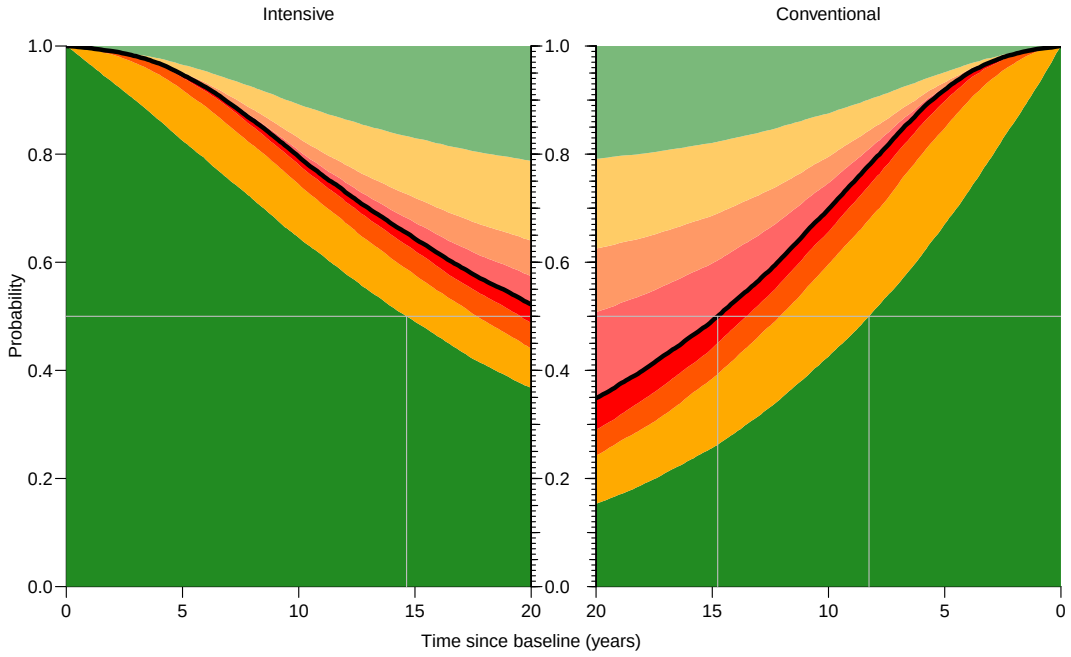
```
> clev <- c("1st CVD","2nd CVD","3+ CVD")
> #
> c0 <- glm( ( (lex.Xst %in% clev) & (lex.Xst!=lex.Cst) ) ~
+            Ns( tsb, knots=d.kn ) + lex.Cst + allocation + sex + age,
+            offset = log(lex.dur),
+            family = poisson,
+            data = subset( S1, lex.Cst!="3+ CVD" ) )
> #
> c0i <- update( c0, . ~ . + allocation:lex.Cst )
> #
> c0n <- update( c0, . ~ . + allocation:tsb )
> #
> # Test for interactions (PropHaz)
> anova( c0i, c0, c0n, test="Chisq" )
```


From rates to probabilities

- There is a one-to-one correspondence between:
  - all rates between states (by time) + initial state distribution
  - state distribution by time
- Model for rates
  $\Rightarrow$ probability of being in a given state at any given time
- **Analytically** this is a nightmare
- **Simulation** is the answer

From rates to probabilities: `simLexis`

- Assume a person is in "DM" initially
  - Simulate a time of death (transition to "D(no CVD)")
  - Simulate a time of CVD (transition to "1st CVD")
  - Choose the smaller as the transition
- If transition is to "1st CVD" simulate death / 2nd CVD, etc.
- Repeat for, say, 10,000 persons
  $\Rightarrow$ simulated cohort study
- `simLexis` does this for you, provided you have
  - initial state and covariates for all persons
  - models to predict (cumulative) rates
- Count how many is in each state at each time:
  $\Rightarrow$ state occupancy probabilities
- `nState` and `pState` does this for you

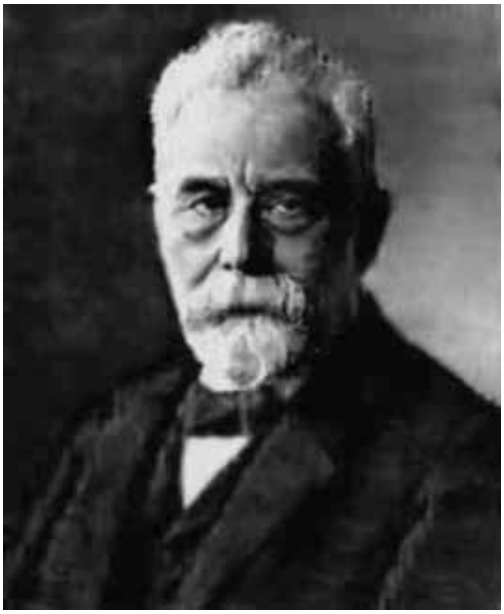


Using the `Lexis` machinery

- Allows estimation of fully parametric rate function
- Simple test for proportional hazards
- State occupancy probabilities requires simulation: `simLexis` — see vignette in `Epi` package
- Access to other measures such as expected residual lifetime.
- — similar machinery available in `Stata`:
  - `multistate`
  - Crowther, M. J. & Lambert, P. C.: Parametric multi-state survival models: flexible modeling allowing transition-specific distributions with application to estimating clinically useful measures of effect differences. Under review for Stats in Medicine
  - Only one timescale, however. . .

History

- `Epi` package grew out of "Statistical Practice in Epidemiology with R", annually since 2002 in Tartu Estonia
- `Lexis` machinery conceived by Martyn Plummer, IARC
- Naming originally by David Clayton & Michael Hills, `stlexis` in Stata, later renamed `stsplit`
- David Clayton wrote a `lexis` function for the `Epi` package. Obsolete now.



EINLEITUNG

IN DIE

THEORIE

DER

BEVÖLKERUNGSSTATISTIK

VON

W. LEXIS

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

STRASSBURG

KARL J. TRÜBNER

1875.

Summary of `Lexis`

- Proper representation of multistate data essential:
  States, transitions, risk time, occupancy
- Readable modeling code — but standard models
- Calculation of state probabilities requires simulation in any realistic situation
- Examples of practical multistate modeling in:
  `http://bendixcarstensen.com/AdvCoh/Lexis-ex/`
- Worked example in the `simLexis` vignette in `Epi` package

Thanks for your attention