Multi-state Modelling and the Estimation of Healthy Life Expectancy

Amsterdam, April 2016

Ardo van den Hout & Robson Machado

Department of Statistical Science
University College London
United Kingdom

Multi-state Modelling and Healthy Life Expectancy

Day	
1	**Introduction (both ways)**
	Practical I. Using simulated data
	Data considerations and software specifics
	Practical II. Using your own data
2	**Multi-state survival models**
	Practical III. Maximum likelihood estimation
	Life expectancies
	Practical IV. Estimation of life expectancies
	Results & issues in practicals so far
	Practical V
	Advanced topics
	Practical VI

Some literature for this workshop:

- Kalbfleisch & Lawless (1985). The analysis of panel data under a Markov assumption. *JASA*
- Jackson (2011). Multi-state models for panel data: The msm package for R. *Journal of Statistical Software*
- Kulkarni (2011). *Introduction to Modeling and Analysis of Stochastic Systems (2nd Ed.)* New York: Springer
- Van den Hout (*summer 2016*). *Multi-state Survival Models for Interval-Censored Data*. CRC/Chapman and Hall

Lecture 1

Introduction: multi-state modelling and healthy life expectancy (HLE)

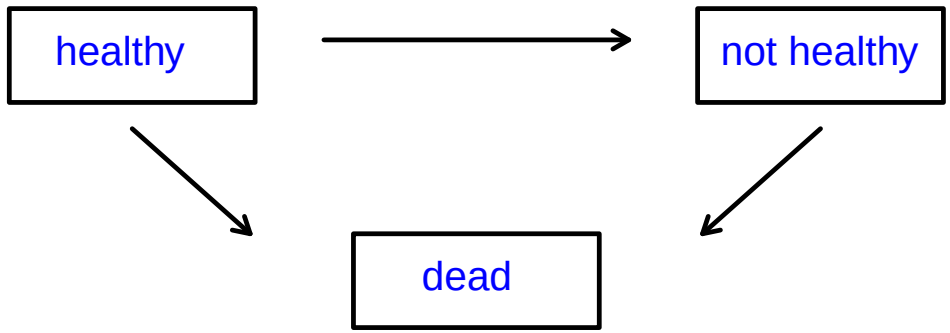
Outline:

1. Research interest & data
2. Example of a multi-state survival model
3. Estimation of life expectancies using a fitted model
4. Software
5. Conclusion
6. *Participants/interest/data*

1. Research interest & data

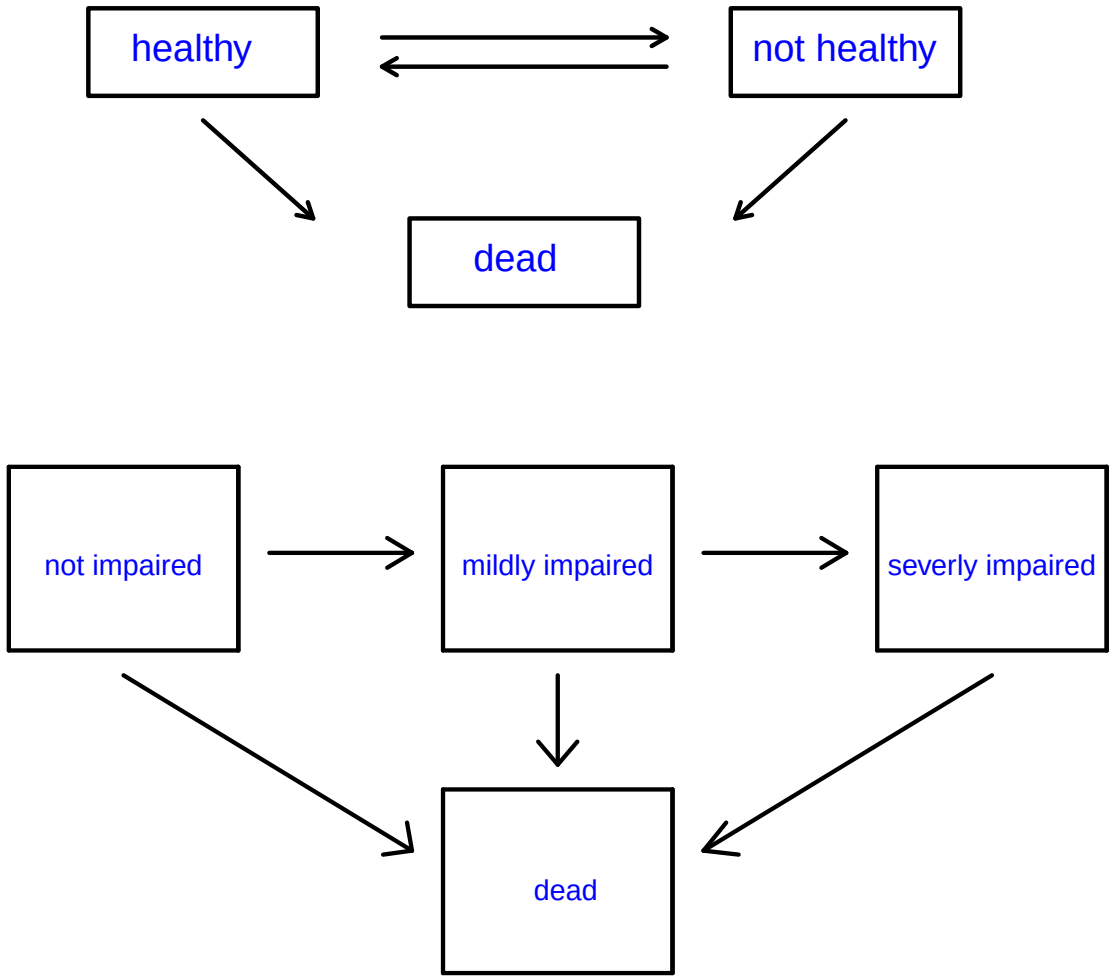
Aim: to estimate healthy life expectancy

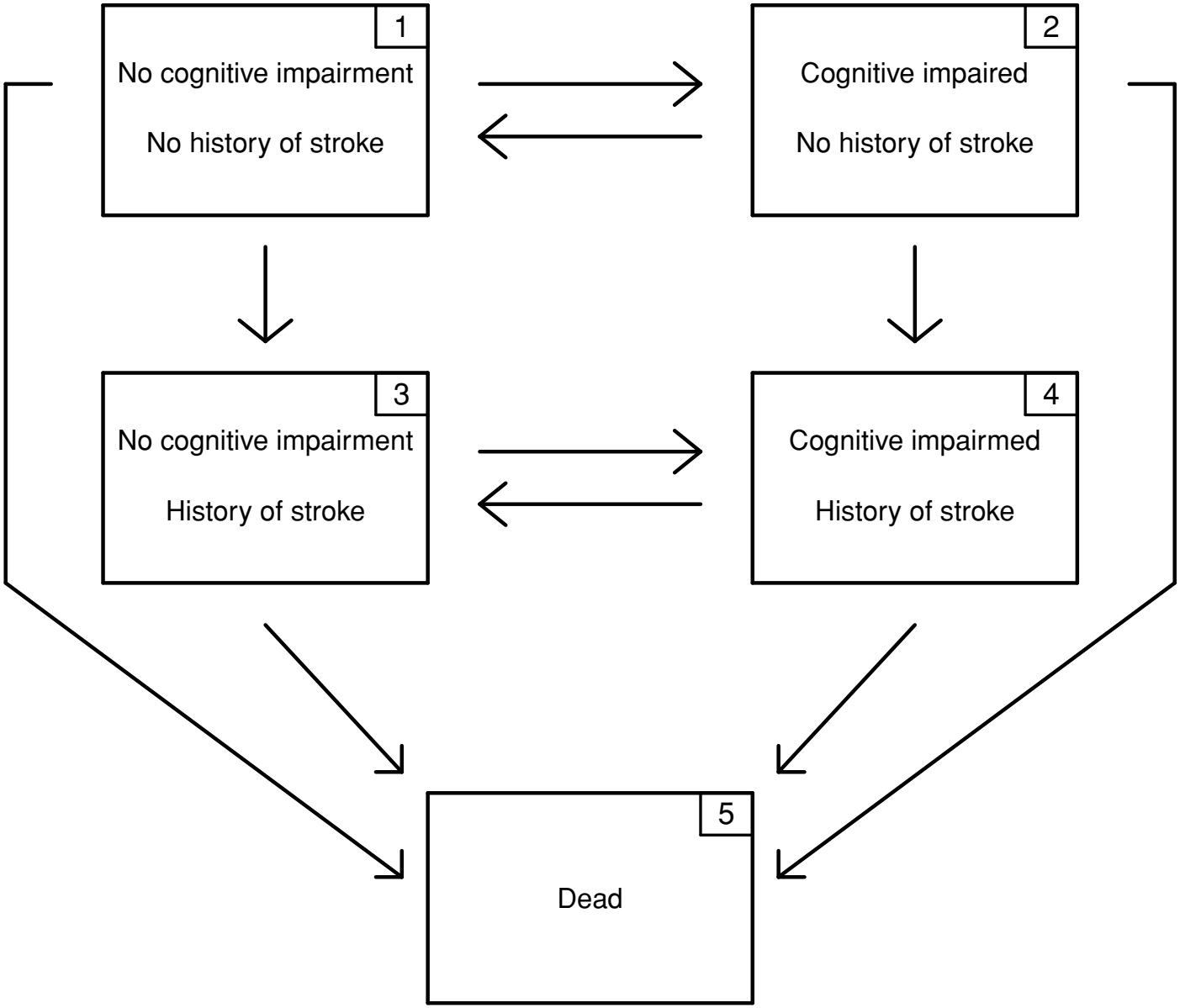
That is; expected number of years in a specific state remaining at a given age



Using longitudinal data and statistical modelling

Aim for method that can deal with various multi-state processes:




1
No cognitive impairment
No history of stroke
2
Cognitive impaired
No history of stroke
3
No cognitive impairment
History of stroke
4
Cognitive impairmed
History of stroke
5
Dead

Aim for method that ...

- can deal with irregular spaced observation times
- can deal with right censoring
- can deal with interval censoring
- allows for effect of covariates
- allows for proper statistical inference
- can be applied using existing software

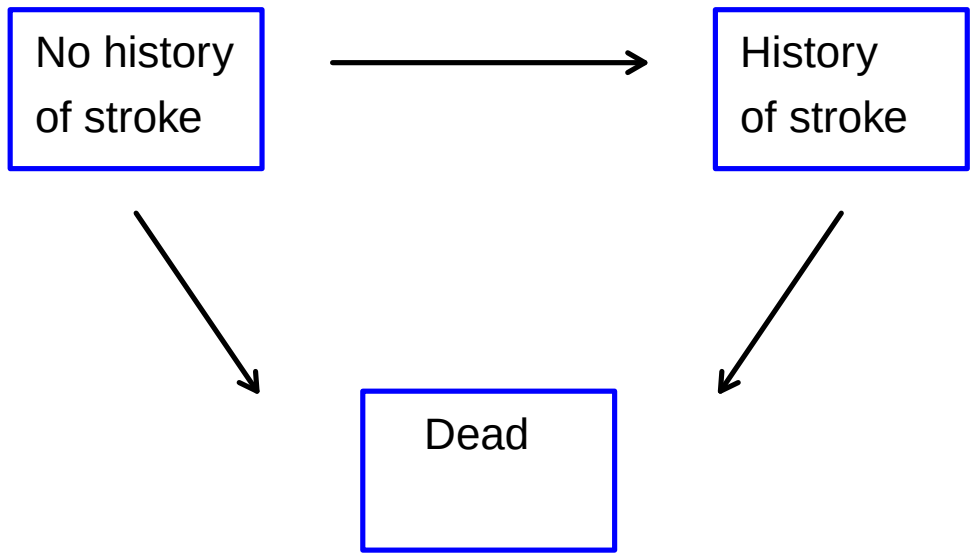
Estimation of HLE will be based on a fitted multi-state model. Hence...

2. Example of a multi-state survival model

- Model for history of stroke in the older population in the UK:
  - Risk factors for stroke: age, sex, education, and year of birth
  - Total residual life expectancy (TLE):

    TLE = stroke-free LE + LE with a history of stroke

- Longitudinal data available from the MRC Cognitive Function and Ageing Study (CFAS, `www.cfas.ac.uk`)
- No (reliable) data on exact time of stroke
- History of stroke = one or more strokes in the past
- Three-state model for history of stroke:



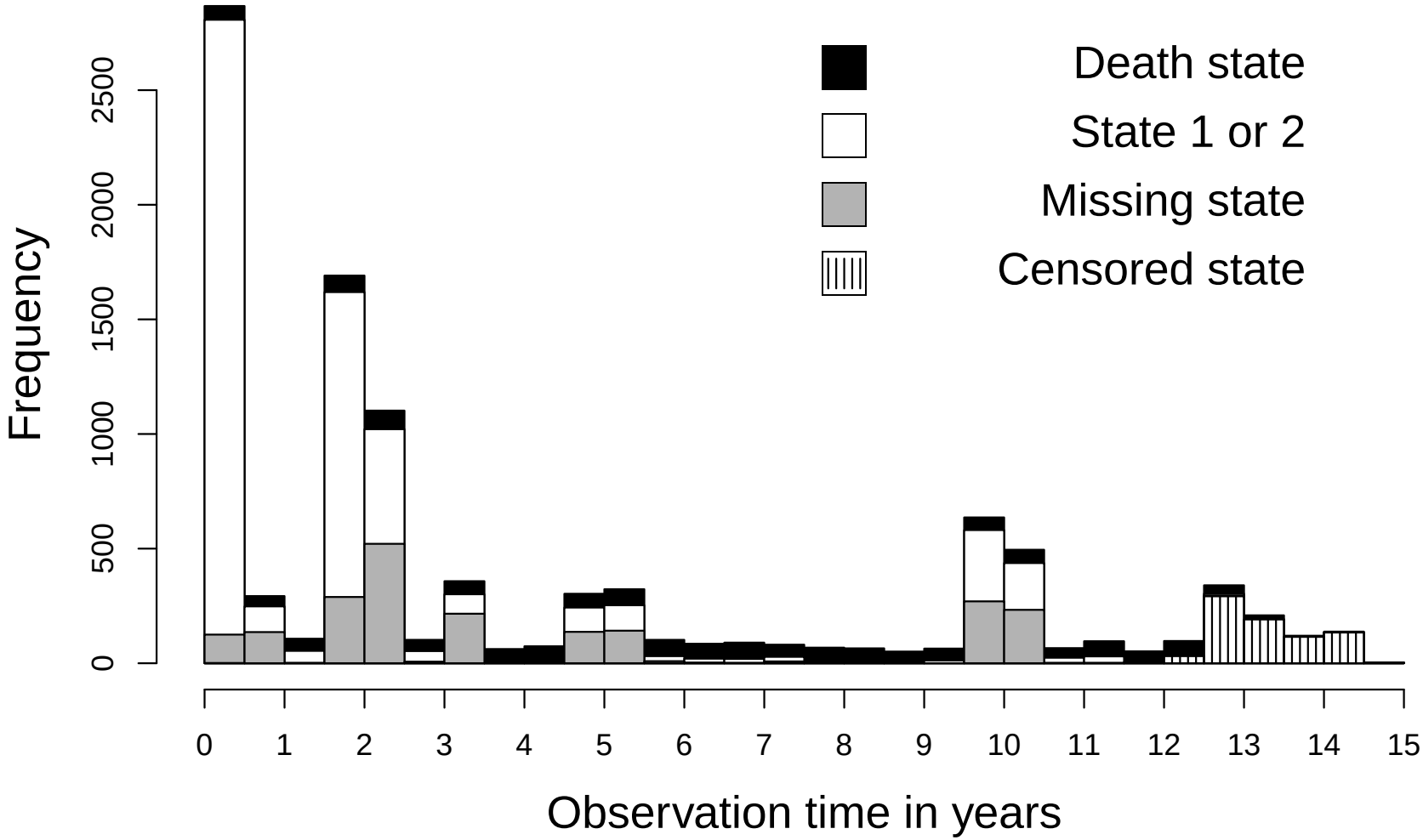
Censoring

- Occurrence of stroke is process in continuous time
- Pre-scheduled interviews: transitions between living states are interval-censored (*panel data*)
- In CFAS: death times are known
- Right censoring at end of 14 years of follow-up

Missing data

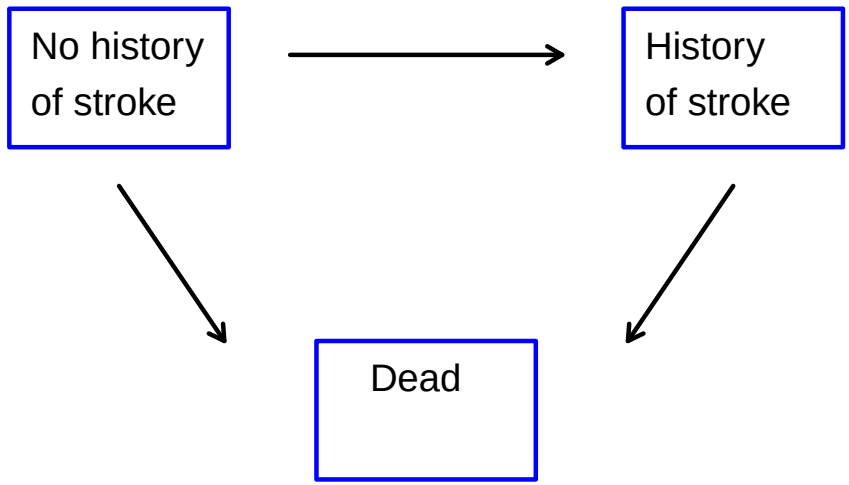
- Pre-scheduled interviews
- A missed interview: missing data
- Missing values for state can be seen as interval-censored data
  But...
  - Info on time-dependent covariates is also missing
  - If reason for missing an interview is related to the process under investigation, then ignoring this can lead to bias

- Panel data CFAS. For subset ($N = 2321$) in application:



Continuous-time model

- Three-state time-dependent model for stroke:



- Three transition-specific hazards: $q_{12}(t), q_{13}(t)$, and $q_{23}(t)$
- Log-linear model: $q_{rs}(t) = \exp\left(\boldsymbol{\beta}_{rs}^{\top}\mathbf{z}(t)\right)$

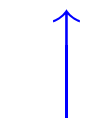
- Age as time scale

- Log-linear model in application:

$$q_{rs}(\text{age}) = \exp\left(\beta_{rs.0} + \beta_{rs.1}\texttt{age} + \beta_{rs.2}\texttt{ybirth} + \beta_{rs.3}\texttt{sex} + \beta_{rs.4}\texttt{educ}\right)$$

- Or, equivalently with $t = \text{age}$,

$$q_{rs}(t) \quad = \quad \lambda_{rs} \exp(\gamma_{rs} t) \exp(\boldsymbol{\alpha}_{rs}^{\top} \mathbf{z})$$



Gompertz baseline hazard

- Continuous-time model with age as time scale

- Maximum likelihood estimation using a piecewise-constant approximation of the continuous-time hazards
  - **Option 1**. Use data to define grid. Example: for individual i with observation times $t_{i1}, t_{i2}, t_{i3}, t_{i4}$, hazards are assumed to be constant within $(t_{i1}, t_{i2}]$, $(t_{i2}, t_{i3}]$, $(t_{i3}, t_{i4}]$

  - **Option 2**. Can also use a predefined grid and embed observation times in this grid

    Option 1 is used in this example and in workshop

Missing states

- Example: say there are three scheduled observations times: t_1, t_2, and t_3, and the state at t_2 is missing

  Law of total probability:

  $$\mathrm{P}(X_{t_1}, X_{t_3}) = \sum_{x \in \{1,2\}} \mathrm{P}(X_{t_1}, X_{t_2} = x, X_{t_3})$$

  That is, distribution of observed states can be expressed as a sum over all possible latent states at time t_2

- Leaving missing states in the data improves piecewise-constant approximation! Instead of one interval $(t_1, t_3]$, the model is fitted using intervals $(t_1, t_2]$ and $(t_2, t_3]$

Data statistics

- *State table* contains frequencies for the number of times each pair of states are observed at successive observation times

- Subset of CFAS: data from Newcastle
  State table:

	To				
From	1	2	3	Missing	Right-censored
1	2942	105	837	855	382
2	0	304	176	60	43
Missing	24	8	542	1200	341

- $N = 837 + 176 + 542 + 382 + 43 + 341 = 2321$

- 1441 women, 880 men

- Frequencies for age groups at baseline:

< 70	$(70, 75]$	$(75, 80]$	$(80, 85]$	> 85
761	559	541	318	142

- Frequencies for number of records per individual for living states:

1	2	3	4	5	6	7	8	9
566	788	629	151	99	54	21	12	1

Fitted model

- Time scale age is age in years minus 78.5
- Equations for the three transition hazards:

$$q_{rs}(\text{age}) = \exp\left(\beta_{rs.0} + \beta_{rs.1}\texttt{age} + \beta_{rs.2}\texttt{ybirth} + \beta_{rs.3}\texttt{sex} + \beta_{rs.4}\texttt{educ}\right)$$

age		ybirth		sex (men $\equiv$ 1)	
$\beta_{12.1}$	0.11 (0.05)	$\beta_{12.2}$	0.03 (0.05)	$\beta_{12.3}$	0.40 (0.20)
$\beta_{13.1}$	0.09 (0.01)	$\beta_{13.2}$	$<$ 0.01 (0.01)	$\beta_{13.3}$	0.36 (0.08)
$\beta_{23.1}$	0.05 (0.02)	$\beta_{23.2}$	-0.01 (0.02)	$\beta_{23.3}$	0.43 (0.13)
educ (10 or more yrs of educ $\equiv$ 1)					
$\beta_{12.4}$	-0.02 (0.23)	$\beta_{13.4}$	-0.27 (0.10)	$\beta_{23.4}$	0.16 (0.16)

3. Estimation of HLE using a fitted model

- Life expectancy (LE): years remaining conditional on current age. Hence *residual life expectancy*

- Multi-state survival: LE in state s given state r at age t_0:

$$e_{rs}(t_0) = \int_0^\infty \mathrm{P}(X_{t+t_0} = s |\ X_{t_0} = r, \mathcal{Z}) dt$$

- The transition probability in the integral is derived from the multi-state model

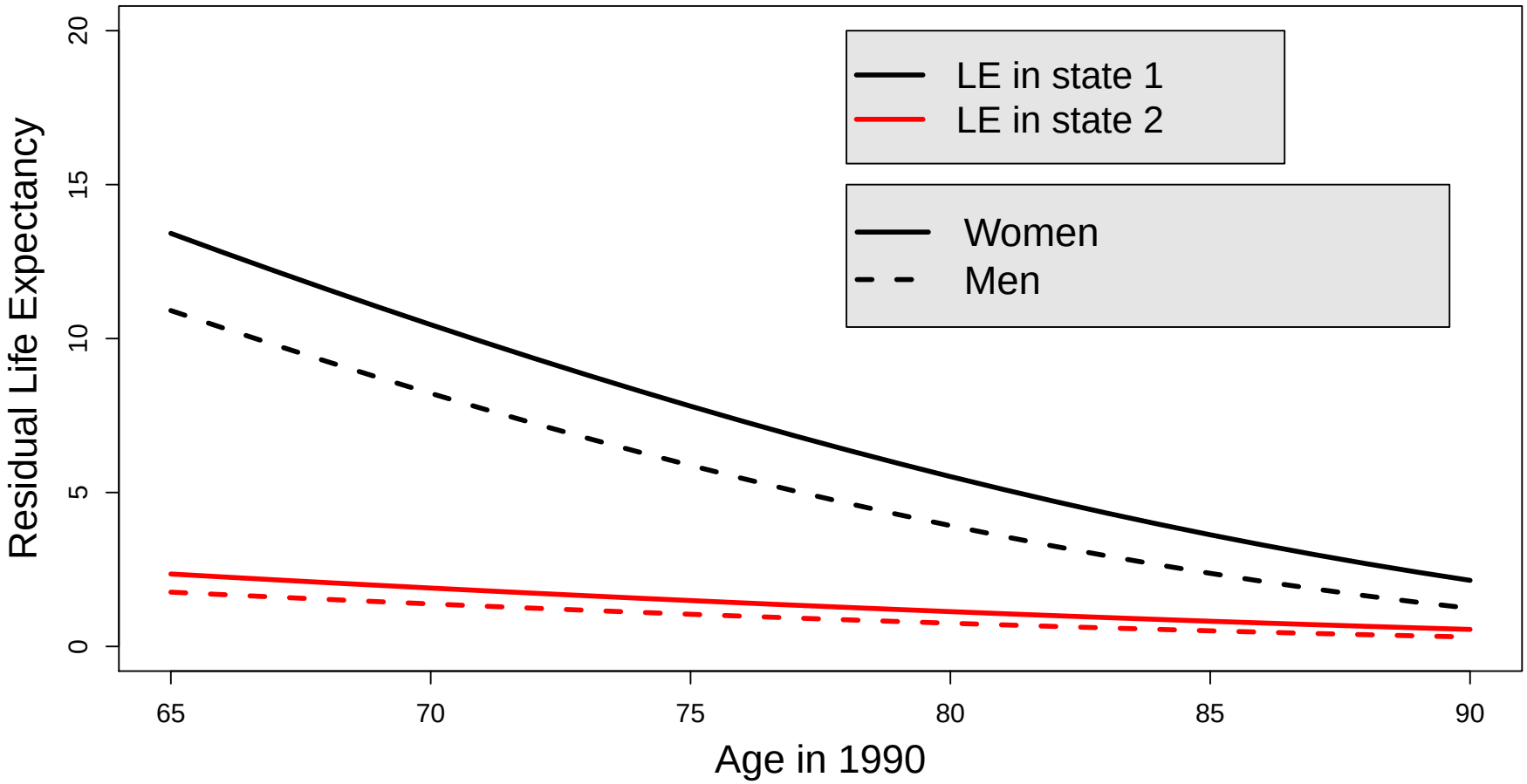
- Example: for a man age 70, with more than 10 years of education, born in 1920, and without a history of stroke, $e_{11}(70 - 78.5) = e_{11}(-8.5)$ is the HLE in years (expected remaining time spent in state 1)

Using fitted model, ELECT software output:

```
For covariates values specified as:
  age ybrth   sex  educ
 -8.5  20.0   1.0   1.0
Using simulation with  1000 replications

Point estimates, and mean, SEs, and quantiles
from simulation:
      pnt    mn   se 0.025q 0.975q
e11 11.91 11.74 0.69  10.36  12.98
e12  1.69  1.77 0.52   0.88   2.95
e21  0.00  0.00 0.00   0.00   0.00
e22  7.08  7.12 0.93   5.46   9.06
e1  10.45 10.30 0.58   9.18  11.43
e2   1.90  1.95 0.43   1.23   2.89
e   12.35 12.25 0.48  11.35  13.25
```

- For $\geqslant$ 10 years of education:



4. Software

- Need software to fit the model, and software to compute LEs:
  - The `msm` package for R written by Chris Jackson (*Journal of Statistical Software*, 2011). Available on the R web site `www.r-project.org`
  - `ELECT`: Estimating Life Expectancies in Continuous Time. Suite of R functions, available on `www.homepages.ucl.ac.uk/~ucakadl/`

Fit model using msm

```
q <- 0.05
Q <- rbind(c(0,q,q), c(0,0,q), c(0,0,0))
covariates <- as.formula("~age+ybrth+sex+educ")

model <- msm(state~age, subject=id, data=dta,
    center=FALSE, qmatrix=Q, death=TRUE,
    covariates=covariates, censor=c(-1,-2),
    censor.states=list(c(1,2),c(1,2)),method="BFGS",
    control=list(trace=0, fnscale=80000))
```

Compute LEs using ELECT

```
sddata  <- dta[dta$state%in%c(1,2),]
age     <- 70 - 78.5
age.max <- 115 - 78.5
ybrth   <- 1920 - 1900
educ    <- 1
sex     <- 0

LEs <- elect(model=model, b.covariates=list(age=age,
   ybrth=ybrth, sex=sex, educ=educ), statedistdata=sddata,
   h=0.5, time.scale.msm="years", age.max=age.max, S=1000)
```

5. Conclusion

- Continuous-time illness-death model for stroke. Flexibility w.r.t. censoring and inclusion of covariates
- Missing values for state can be taken into account
- LEs using estimated model parameters. Complete info on uncertainty. Alternative to multi-state life-table methods
- Computation with available software

Continued...

- Aspects to be discussed later:
  - Model validation
  - Functional form of regression equation
  - Piecewise-constant hazards approximation
  - Maximising the likelihood function
  - The conditional Markov assumption
  - Definition of residual life expectancy
  - Estimation of residual life expectancy
  - Details of the software

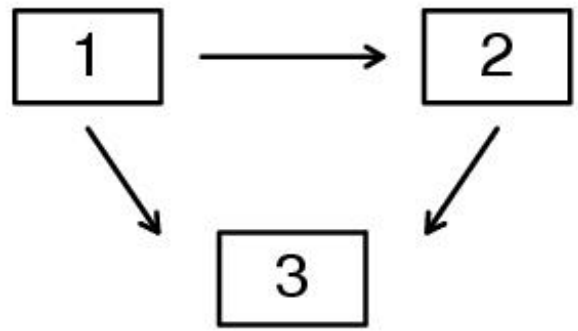
Participants

Interest

Data

Practical 1. Using simulated data

- Aim: to explore the software `msm` and `ELECT` without running into data problems
- Three-state process:



- Made up data!

Lecture 2

Data considerations and software specifics

Outline:

1. Censoring
2. Study design
3. The `msm` software
4. Model fitting: numerical issues and tips

1. Censoring

- A censored quantity is one whose exact value is unknown, but known to be in a certain set of values
- **Interval censoring:** The transition time is known to be in a bounded interval
  - For example, onset of disease diagnosed in a series of visits to a hospital
- **Right censoring:** The current state is known to be a state in the set of values
  - For example, at the end of a chronic-disease study, patients are known to be alive but in an unknown state
  - This is different from interval censoring where *observation times* are censored, in right censoring *states* are censored

2. Study Design

Study design takes into account the volatility of the process. Possible study designs include:

- **Fixed:** Each patient is observed at fixed intervals specified in advance
- **Random:** The sampling times vary randomly, independently of the current state
- **Doctor's care:** More severely ill patients are monitored more closely. The next sampling time is chosen on the basis of the current state
- **Patient self-selection:** A patient may decide to visit the doctor on occasions when they are in a poor condition

3. The msm package

The following is based on also Jackson's msm *Vignette* (2016, cran.r-project.org)

- The R software msm uses maximum likelihood estimation for general multi-state models for continuous time
- The msm package allows multi-state models to be fitted to data from processes with interval-censored observation times (panel data), exactly-observed transition times, exact death times and censored states, or a mixture of these schemes

4. Model fitting: numerical issues and tips

- If over-complex models are applied with insufficient data, then the parameters of the model will not be identifiable
- It is inadvisable to include several covariates in a models simultaneously
- The defaults of many functions in `msm` are for time-homogeneous models
- Mind the function arguments when fitting time-dependent models fitted with `msm`

Model simplification:

- Make sure there are enough data for each transition
- Consecutive states could be merged if necessary
- It is possible to apply constrains on the intensities or the covariates so that parameters are equal (or zero) for certain transitions
- Understand the difference between *exact* and *interval-censored* time of transitions
- Although individuals can be in state 1 at time t, and state 3 at time $t + \Delta$, that does not mean that instantaneous transitions from 1 to 3 should be permitted

Initial values:

- Make sure that a sensible set of initial values have been chosen
- Run the model for several different initial values to ensure that the estimation is not stuck at a local optimum

Scaling:

- It often helps to apply a scaling factor to standardise the likelihood (`fnscale`) or certain individuals parameters (`parscale`)

- It is advisable to analyse covariates on a standardised scale

  For example, working in terms of a time unit of months or years instead of days, when the data range over thousands of days

  (When using `ELECT`, use time scale weeks, months, or years)

Convergence criteria:

- "False convergence" in which `optim` reports convergence of the optimisation but the hessian is not positive definite

- When the hessian is not positive definite, standard errors can not be calculated

  Try to solve this problem by tightening the convergence criteria

  For example, `control = list(reltol = 1e-16)`

  Default is `reltol=1e-8`

Choice of algorithm:

- By default `msm` uses the BFGS method in `optim`
- BFGS uses function values and gradients to optimise a function. This speeds up the optimisation
- In case of divergence, it possible to use more robust methods such as Nelder-Mead

How to diagnose the problem:

`optim` **"function cannot be evaluated at initial parameters"** ?

- Run `msm` again with `fixedpars=TRUE` set
- It gives the value of -2 log-likelihood at initial values
- This will be probably be `Inf`
- Call `logLik.msm(x, by.subject = TRUE)` to show the contribution of individual subjects
- It is possible that data from only a few individuals cause the log-likelihood to be infinite

Practical 2. Using your own data

Checklist:

- Explore the data; see Practical 1 for `R` code
- Decide on the transitions that you want to model, and specify the generator matrix $\mathbf{Q}$. Three-state examples:
  - Reversible: `Q <- rbind(c(0,q,q),c(q,0,q),c(0,0,0))`
  - Progressive: `Q <- rbind(c(0,q,q),c(0,0,q),c(0,0,0))`
- Classify observation types: censored states/exact transition times? For a censored state, what are possible latent states?
- Start with fitting the intercept-only model
- If you run into optimisation problems, see hints in Lecture 2
- Extended models: mind the piecewise-constant approximation
- Check estimated effects and estimated standard errors
- When using `ELECT`: is estimated total LE reasonable?

Lecture 3

Multi-state survival model

Outline:

1. English Longitudinal Study of Ageing (ELSA)
2. Multi-state survival model and likelihood function
3. Maximising the likelihood function
4. Analysing ELSA data
5. Conclusion

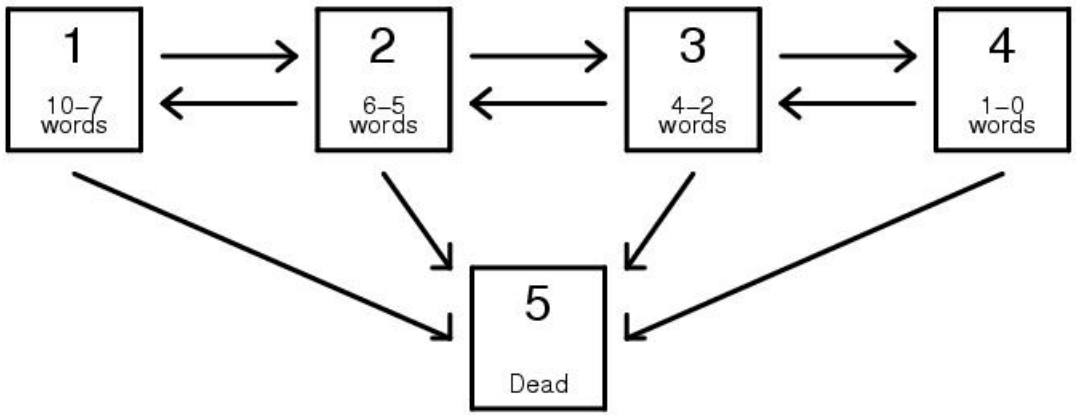
1. English Longitudinal Study of Ageing (ELSA)

- English Longitudinal Study of Ageing (ELSA)
- Info on health, economic position, and quality of life. Data in waves 1 - 5 (2002-2010)
- Random sample $N = 1000$ with 544 women and 456 men. Age $\geqslant 50$ years
- Longitudinal response for analysis: number of words remembered in a recall. Response $\in \{0, 1, ..., 10\}$

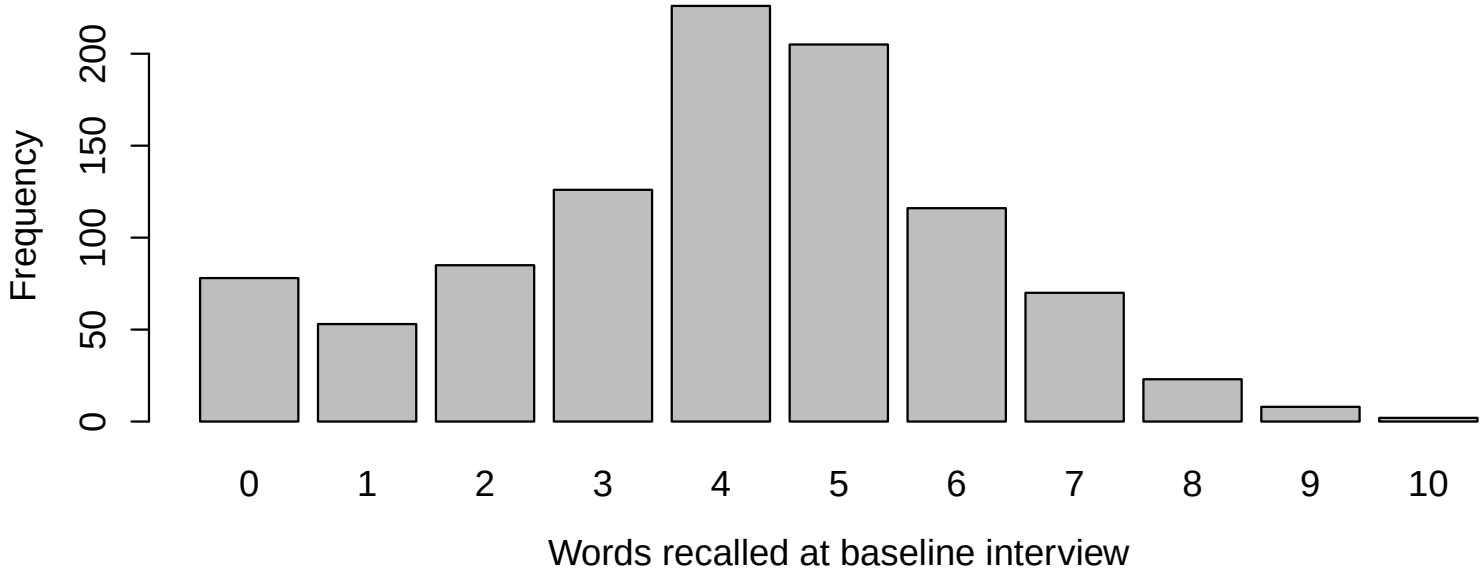
- Dropout due to death is about 20%

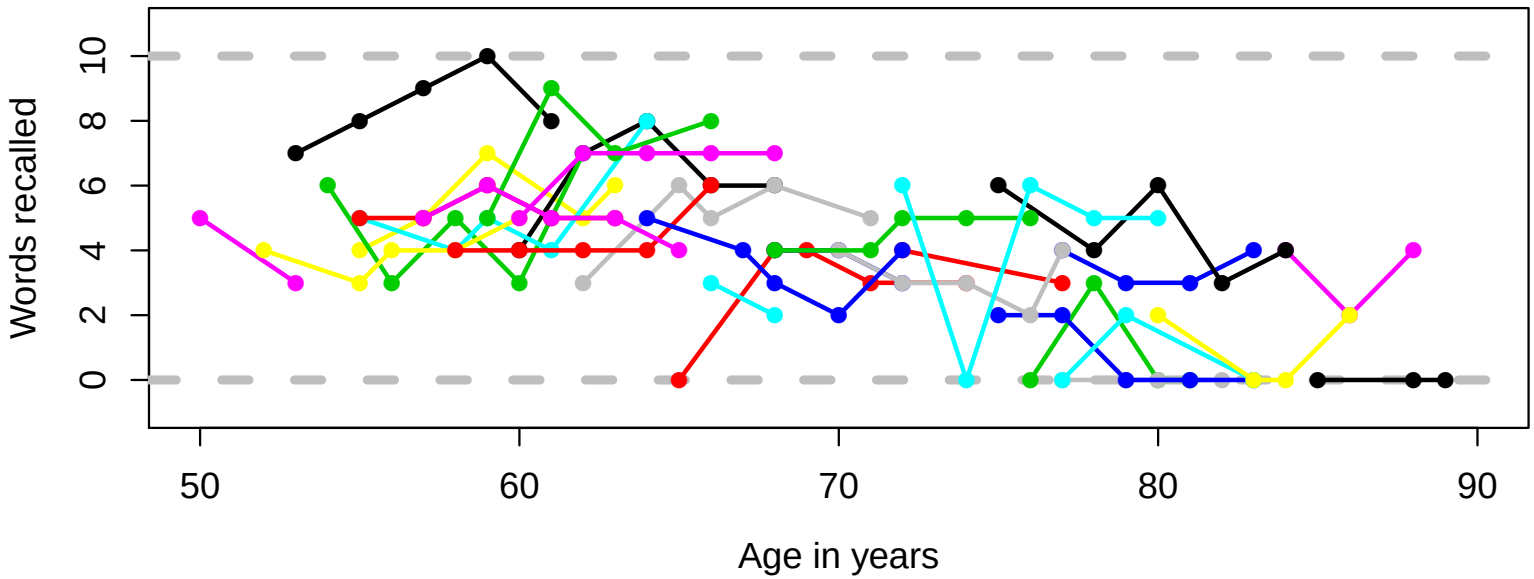
- Analysis of longitudinal data for ageing processes cannot ignore dropout due to death
  - One option: combine a longitudinal measurement model with a survival model (*joint model* with random effects)
  - Another option: use a (fixed-effects) multi-state survival model

- Multi-state survival model:



- Interval-censored data for living states, exact times for death


Frequency
0
50
100
150
200
0
1
2
3
4
5
6
7
8
9
10
Words recalled at baseline interview


Words recalled
0
2
4
6
8
10
50
60
70
80
90
Age in years

State table for the ELSA data: number of times each pair of states are observed at successive observation times

	To				
From	10-7 words	6-5 words	4-2 words	1-0 words	Dead
10-7 words	164	150	49	12	8
6-5 words	156	440	303	48	40
4-2 words	52	336	616	151	85
1-0 words	11	35	114	149	72

2. Multi-state survival model and likelihood function

- Regression models for transition-specific hazards
- Likelihood for interval-censored data constructed using transition probabilities for time intervals. Known time of death can be accounted for
  (Cox & Miller 1965; Kalbfleisch & Lawless 1985; Jackson 2011)
- Framework is extended to models with parametric time dependency (Van den Hout 2016)

- The time scale in a survival model is important. The time origin specifies from which moment survival is measured. (Bull and Spiegelhalter 1997; Korn et al. 1997)

  Typical choices of the time scale:
  - time from randomisation in a clinical trial
  - time from exposure in a study of infectious disease
  - age in an observational study

- In the workshop examples, the time scale t is age

- Individual i at age t_i
- Hazard model for transition from state r to state s is

$$h_{rs}(t_i) = h_{rs.0}(t_i) \exp(\gamma^\top_{rs} \mathbf{x}_i)$$

- Parametric baseline hazards can be chosen from, e.g.,

$$\begin{aligned} \text{Weibull:} \quad h_{rs.0}(t) &= \lambda_{rs} \tau_{rs} t^{\tau_{rs}-1} \\ \text{Gompertz:} \quad h_{rs.0}(t) &= \lambda_{rs} \exp(\xi_{rs} t) \end{aligned}$$

- Choose parametric shape to be able to predict
- **In this workshop**: use `msm` to fit Gompertz models

Likelihood

- **Markov assumption**: all information about the future is contained in the present state. This implies

$$\begin{aligned}\mathrm{P}\left(Y_3 = y_3, Y_2 = y_2 | Y_1 = y_1\right) \\ = \mathrm{P}\left(Y_3 = y_3 | Y_2 = y_2\right) \mathrm{P}\left(Y_2 = y_2 | Y_1 = y_1\right)\end{aligned}$$

  for Y_j denoting the state at time t_j

- Likelihood contribution of individual i with J observations. In absence of death or right-censoring:

$$\begin{aligned} L_i(\boldsymbol{\theta}|\mathbf{y}, \mathbf{x}_i) &= \mathrm{P}\left(Y_J = y_J, ..., Y_2 = y_2 | Y_1 = y_1, \boldsymbol{\theta}, \mathbf{x}_i\right) \\ &= \prod_{j=2}^{J} \mathrm{P}\left(Y_j = y_j | Y_{j-1} = y_{j-1}, \boldsymbol{\theta}, \mathbf{x}_i\right) \end{aligned}$$

- Adjustments in case of right-censored state and exact times of death (see, e.g., Jackson 2011)

  Say states are denoted by $1, 2, .., D$, with D the dead state

  Examples:

$$\mathrm{P}\left(Y_2 = \text{right-censored} | Y_1 = y_1\right) = \sum_{y \in \{1,..,D-1\}} \mathrm{P}\left(Y_2 = y | Y_1 = y_1\right)$$

$$\mathrm{P}\left(Y_2 = D | Y_1 = y_1\right) = \sum_{y \in \{1,..,D-1\}} h_{yD}(t_1) \mathrm{P}\left(Y_2 = y | Y_1 = y_1\right)$$

- A multi-state process is a Markov chain if all information about the future is contained in the present state

- Many processes in real life are not Markovian, but a statistical model based upon a Markov chain may still provide a good approximation

- Multi-state models in this workshop are not Markovian. By linking age and covariates values to transition hazards, information about the future is contained in the present state **and** current age and covariates values

3. Maximising the likelihood function

- Generator matrix $\mathbf{Q}(t) = \mathbf{Q}(t|\theta, \mathbf{x})$ contains hazards as defined at time t
- When hazard do not change in $(t_1, t_2]$, transition probabilities for $(t_1, t_2]$ are given by

$$\mathbf{P}(t_1, t_2) = \exp\Big((t_2 - t_1)\mathbf{Q}(t_1)\Big)$$

(solution to Kolmogorov differential eqns; see Cox & Miller 1965)

In the likelihood, time-dependency is dealt with by using a piecewise-constant approximation where hazard do not change within the interval defined by two individual consecutive observations times

Example with observation times u and t for individuals A and B:

$$A: \quad \underbrace{|------}_{u_1 = 0} \underbrace{|------}_{u_2} \underbrace{|--------|}_{u_3 \qquad\qquad u_4}$$

$$\mathbf{Q}_A(u_1) \qquad \mathbf{Q}_A(u_2) \qquad \mathbf{Q}_A(u_3)$$

$$B: \quad \underbrace{|----------|}_{t_1 = 0} \underbrace{|--------|}_{t_2} \underbrace{|--------|}_{t_3 \qquad\qquad t_4}$$

$$\mathbf{Q}_B(t_1) \qquad \mathbf{Q}_B(t_2) \qquad \mathbf{Q}_B(t_3)$$

- For efficient computation of

$$\mathbf{P}(t_1, t_2) = \exp\Big((t_2 - t_1)\mathbf{Q}(t_1)\Big)$$

see *Nineteen dubious ways to compute the exponential of a matrix, 25 years later*, Moler & Van Loan (2003)

- Computation is included in `msm`, and `ELECT` uses the function in `msm`

- Computation can be intensive which explains (partly) why fitting a model with `msm` may need some time

Likelihood maximisation

- The log-likelihood is defined by combining the likelihood contributions of all N individuals

$$\begin{aligned} \ell(\boldsymbol{\theta}|\mathbf{y}, \mathbf{x}_i) &= \log\left(\prod_{i=1}^{N} L_i(\boldsymbol{\theta}|\mathbf{y}, \mathbf{x}_i)\right) \\ &= \sum_{i=1}^{N} \log\left(L_i(\boldsymbol{\theta}|\mathbf{y}, \mathbf{x}_i)\right) \end{aligned}$$

- Maximisation in `msm` is undertaken using a general-purpose optimiser in `R`. No need to provide derivatives

- Most of the software for statistical computing include general-purpose optimisers. In `R`, it is the function `optim`

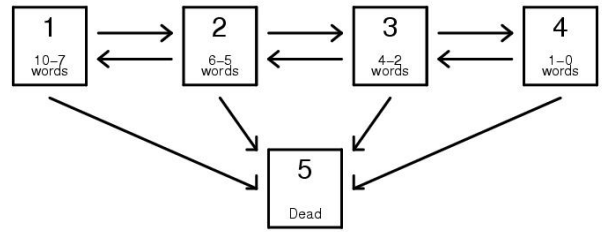
- Default method in `optim` is `Nelder-Mead`, which is relatively slow but robust. Another choice is `BFGS`, which is faster but less robust

- `optim` will return the Hessian matrix if requested. The Hessian can be used to derive the estimated covariance matrix for the maximum likelihood estimate

- It is recommended to explore varying starting values, using different methods for optimisation, taking note of convergence assessment, and inspecting the estimated variance-covariance matrix (all possible with `msm`)

4. Analysing ELSA data

- Multi-state survival model:



- $N = 1000$. Interval-censored data for living states, exact times for death

- State table:

From	*To* 10-7 words	6-5 words	4-2 words	1-0 words	Dead
10-7 words	164	150	49	12	8
6-5 words	156	440	303	48	40
4-2 words	52	336	616	151	85
1-0 words	11	35	114	149	72

Five-state model for
$(r, s) \in \{(1,2),(1,5),(2,1),(2,3),(2,5),(3,2),(3,4),(3,5),(4,3),(4,5)\}$

- Intercept-only model: $q_{rs}(t) = \exp\left(\beta_{rs.0}\right)$
- Model with age:

$$q_{rs}(t) = \exp\left(\beta_{rs.0} + \xi_{rs} t\right),$$

where $\xi_{21} = \xi_{32} = \xi_{43} = 0$ and $\xi_{15} = \xi_{25} = \xi_{35} = \xi_{45}$

- Extended model with binary covariates :

$$q_{rs}(t) = \exp\left(\beta_{rs.0} + \xi_{rs} t + \beta_{rs.1}\ sex + \beta_{rs.2}\ education\right),$$

where $\beta_{15.1} = \beta_{25.1} = \beta_{35.1} = \beta_{45.1}$, and $\beta_{r5.2} = 0$ for $r = 1, 2, 3, 4$

- Likelihood function takes known death times into account (no right censoring in ELSA)

- Maximise the likelihood function
  - by using-written scoring algorithm (Van den Hout 2016)
  - by coding the likelihood in `R` and using `optim`
  - by using `msm` which uses `optim` internally

    Parameter constraints can be coded by using `constraint` and `fixedpars` in `msm`-function call

- Using `msm`:

```
covars     <- as.formula("~age+sex+educ")
constraint <- list( age=c(1,2,3,4,2,5,6,2,7,2),
                    sex=c(1,2,3,4,2,5,6,2,7,2) )
fixedpars  <- c( 10+c(3,5,7), 17+c(3,5,7),
                 24+c(2,3,5,6,8,9,10) )
method     <- "BFGS"

model <- msm( state~age, subject=id, data=dta,
          center=FALSE, qmatrix=Q, death=TRUE,
          covariates=covars, constraint=constraint,
          fixedpars=fixedpars, method=method,
          control=list(trace=0,REPORT=1,maxit=3000,
          fnscale=4000) )
```

- Comparison between models:

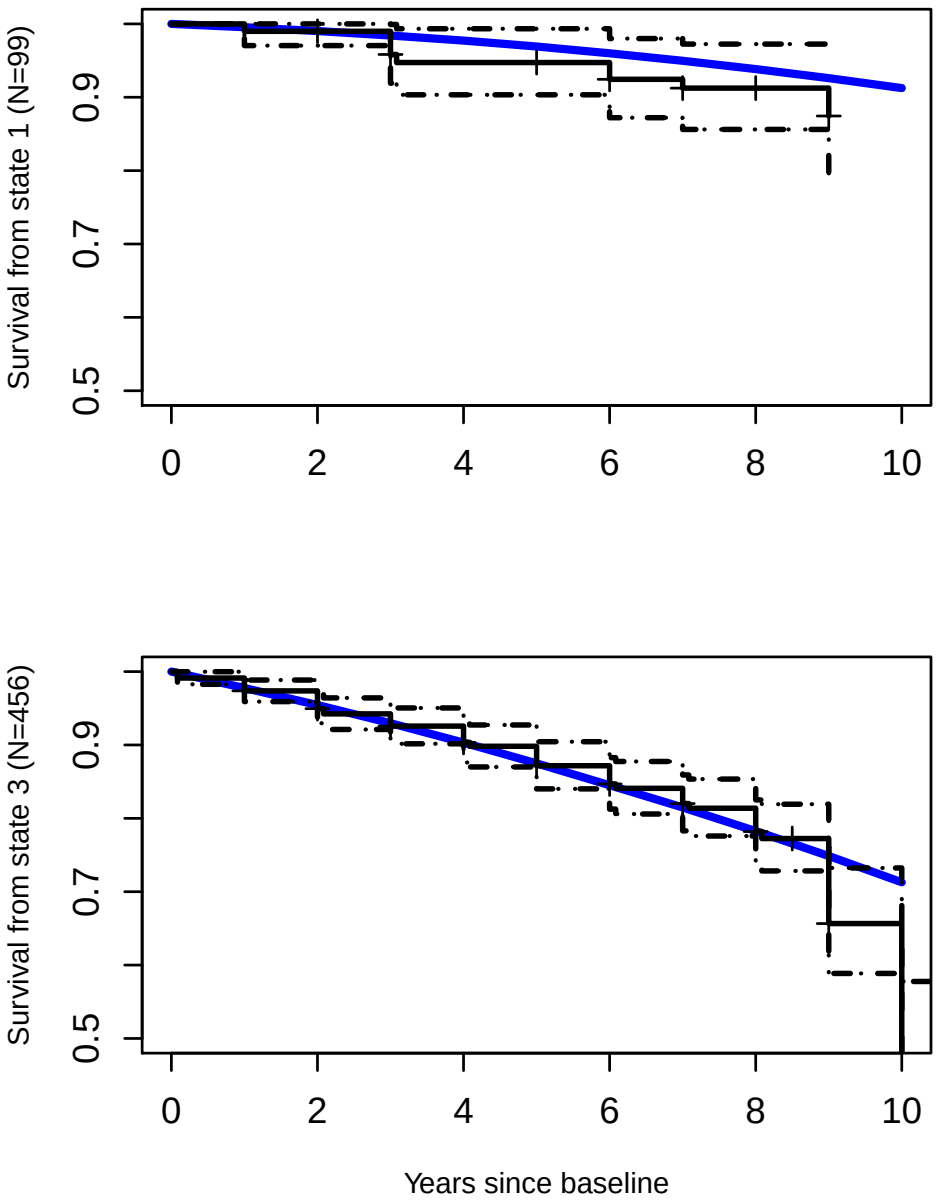
Model	Baseline hazards	#p	$-2\log(L_{\max})$	AIC
Intercept-only	exponential	10	8089.5	8109.5
t	Gompertz	14	7777.0	7805.0
t, *sex*, *education*	Gompertz	21	7661.4	7703.4

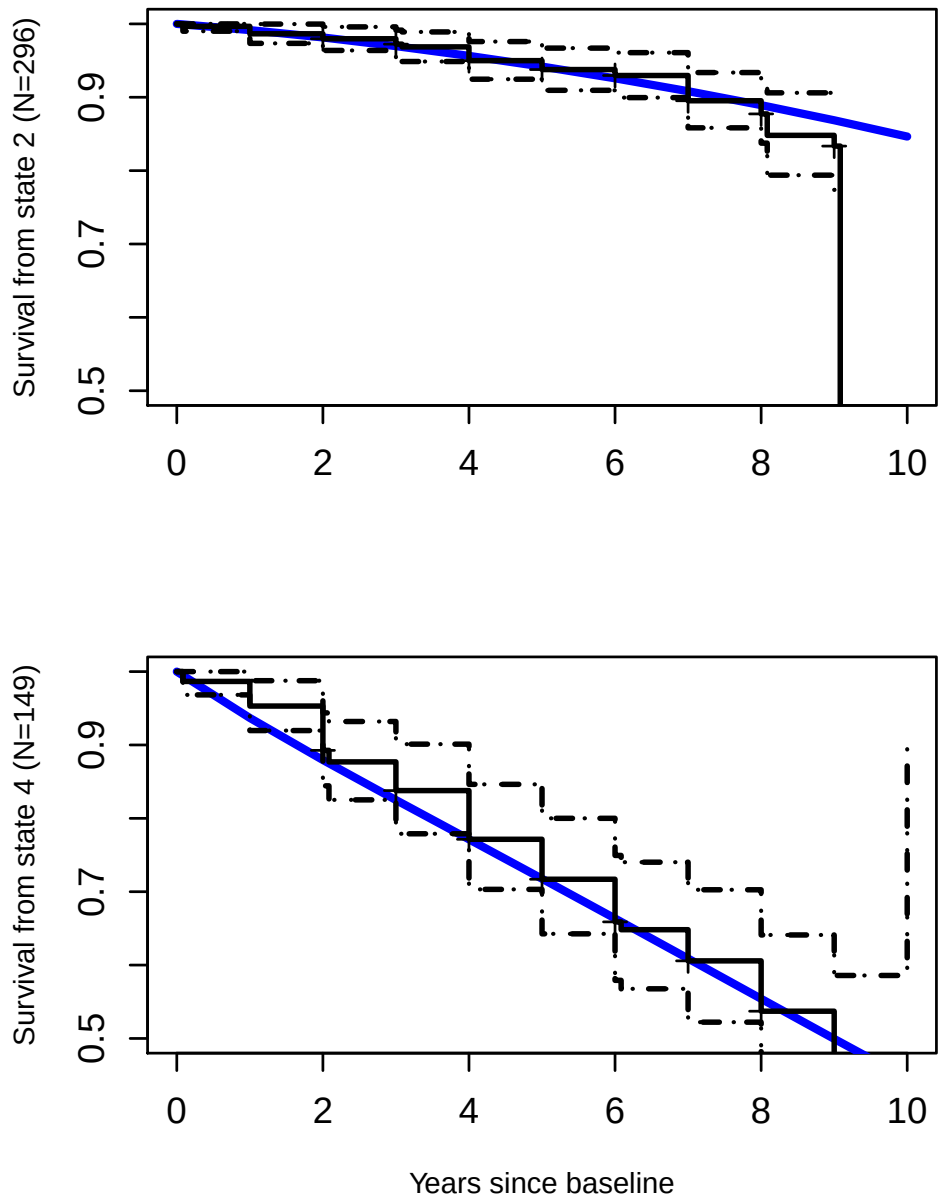
- The best model as selected by the AIC can still be a bad model

- Checking goodness of fit of a multi-state model is hampered by the interval censoring

- In ELSA: no interval censoring for death times
  - Produce Kaplan-Meier estimates for survival conditional on baseline state
  - For each individual, predict model-based survival conditional on baseline state
  - Compare the non-parametric Kaplan-Meier with the parametric survival

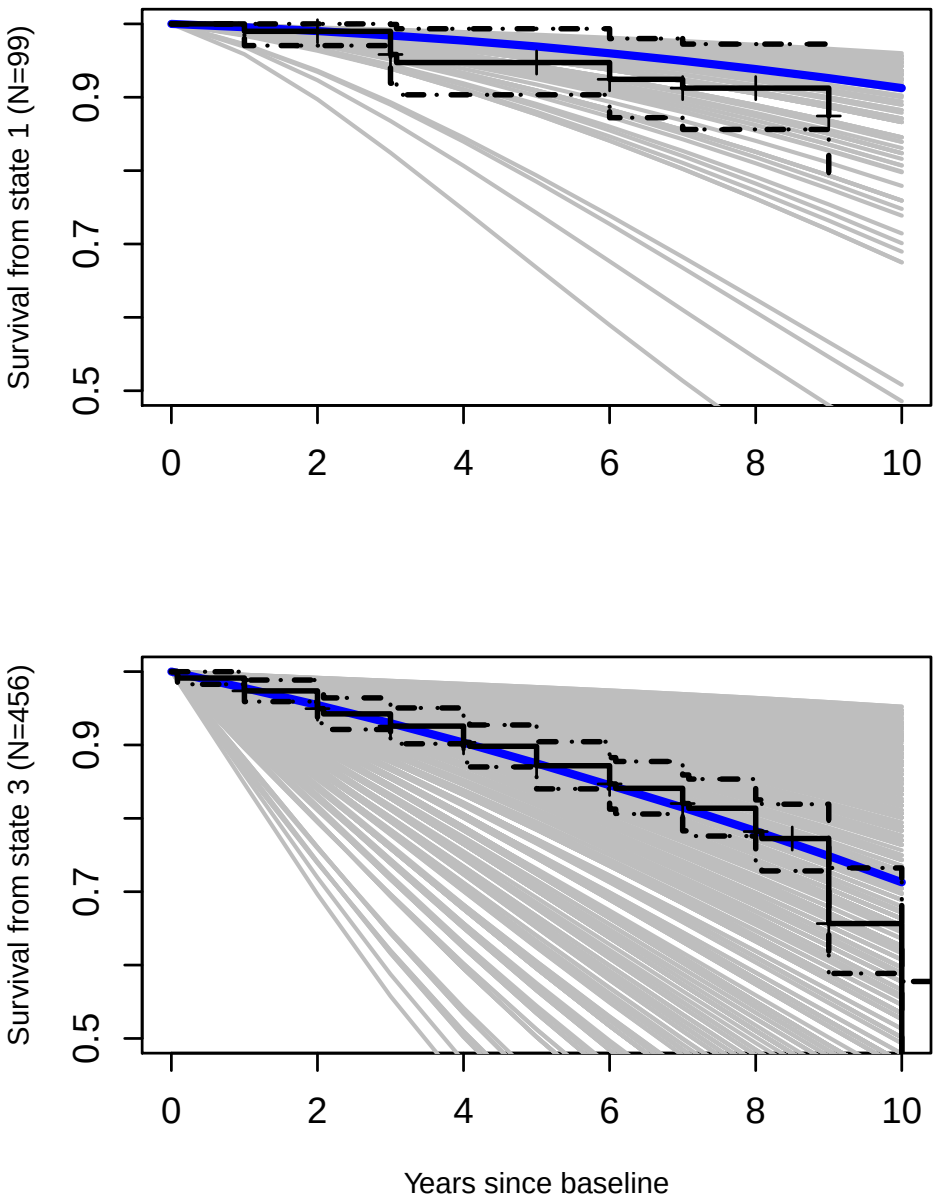
    (Gentleman *et al. 1994; Titman and Sharples 2010)*

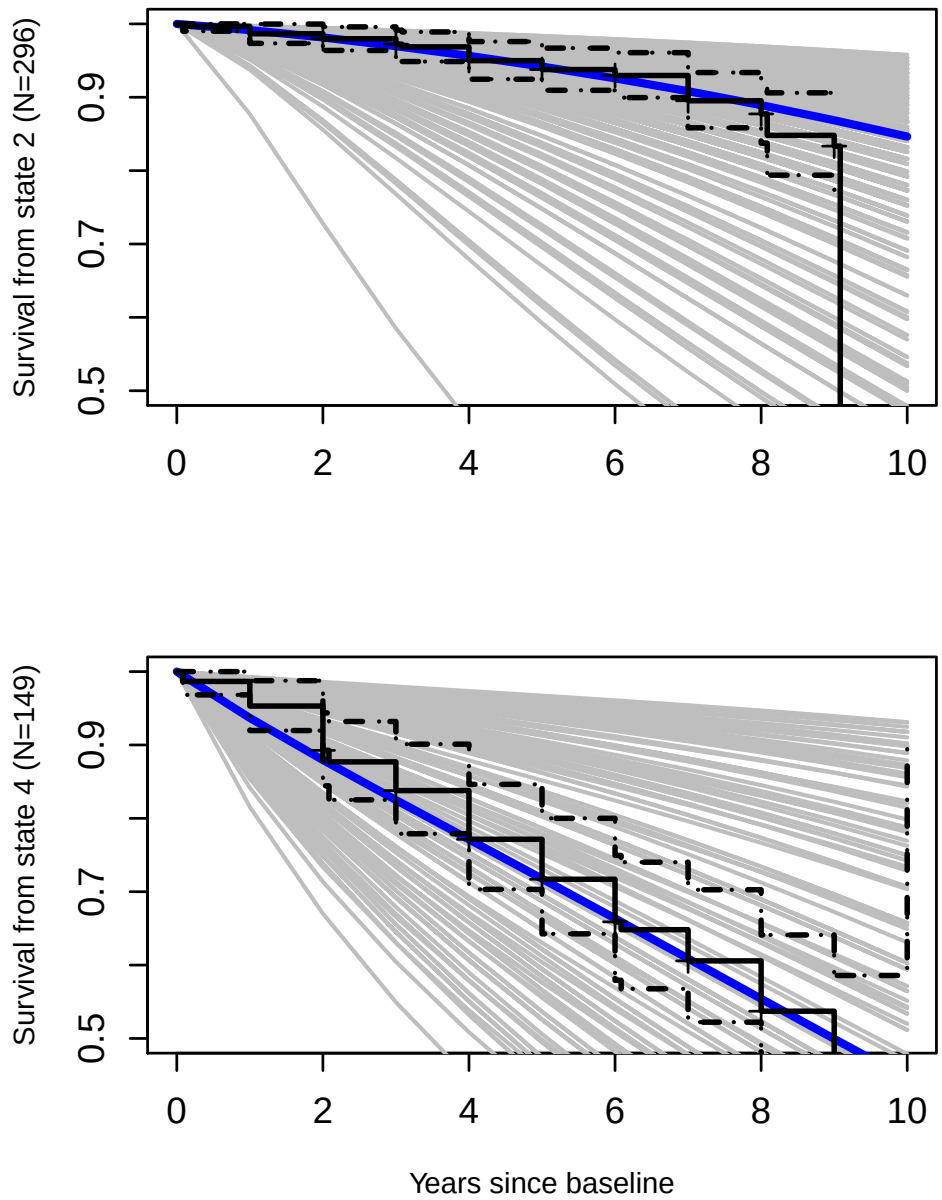

Survival from state 1 (N=99)
0.5
0.7
0.9
0
2
4
6
8
10


Survival from state 2 (N=296)
0.5
0.7
0.9
0
2
4
6
8
10

Survival from state 3 (N=456)
0.5
0.7
0.9
0
2
4
6
8
10
Years since baseline

Survival from state 4 (N=149)
0.5
0.7
0.9
0
2
4
6
8
10
Years since baseline


Survival from state 1 (N=99)
0.5
0.7
0.9
0
2
4
6
8
10


Survival from state 2 (N=296)
0.5
0.7
0.9
0
2
4
6
8
10

Survival from state 3 (N=456)
0.5
0.7
0.9
0
2
4
6
8
10
Years since baseline

Survival from state 4 (N=149)
0.5
0.7
0.9
0
2
4
6
8
10
Years since baseline

- Parameters estimates and SEs. Time scale t is age in years minus 49:

$$q_{rs}(t) = \exp\left(\beta_{rs.0} + \xi_{rs}t + \beta_{rs.1}\ sex + \beta_{rs.2}\ education\right)$$

Intercept			t			*sex*		
$\beta_{12.0}$	-1.133	(0.189)	ξ_{12}	0.052	(0.009)	$\beta_{12.1}$	0.543	(0.137)
$\beta_{15.0}$	-6.338	(0.862)	ξ_{23}	0.084	(0.011)	$\beta_{23.1}$	0.477	(0.151)
$\beta_{21.0}$	-1.357	(0.097)	ξ_{34}	0.047	(0.006)	$\beta_{34.1}$	0.192	(0.100)
$\beta_{23.0}$	-1.060	(0.152)	ξ_{5}	0.063	(0.008)	$\beta_{5.1}$	0.139	(0.145)
$\beta_{25.0}$	-6.064	(0.443)						
$\beta_{32.0}$	-0.774	(0.078)				*education*		
$\beta_{34.0}$	-2.984	(0.223)				$\beta_{12.2}$	-0.291	(0.144)
$\beta_{35.0}$	-5.429	(0.359)				$\beta_{23.2}$	-0.828	(0.102)
$\beta_{43.0}$	-0.749	(0.101)				$\beta_{34.2}$	-0.443	(0.160)
$\beta_{45.0}$	-5.001	(0.437)						

- Transition probabilities can help to interpret the model
- For men aged 60 with higher level of education, the two-year probabilities:

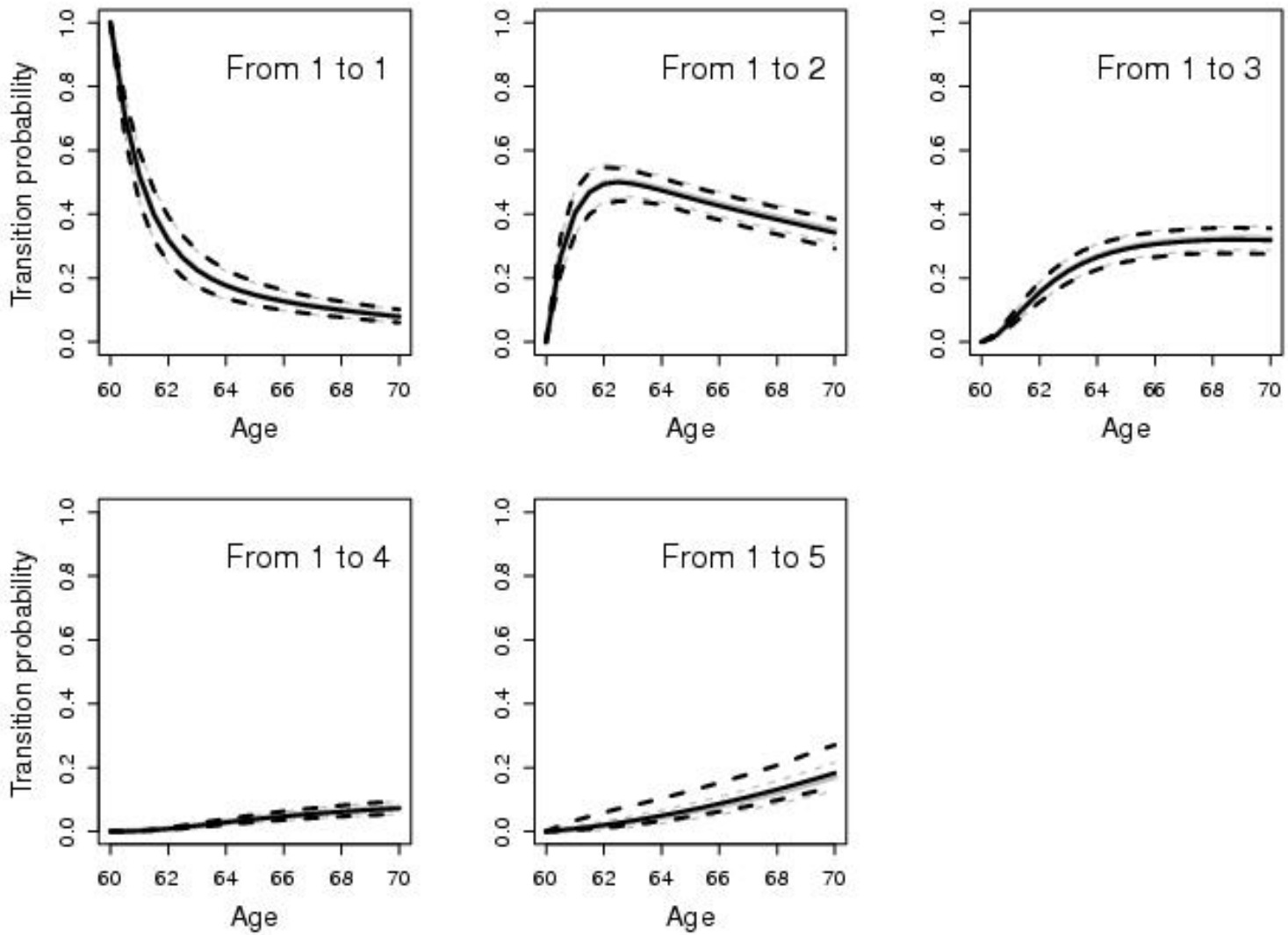
$$\widehat{\mathbf{P}}\begin{pmatrix} t_1 = 11, \\ t_2 = 13 \end{pmatrix} \Bigg| \begin{array}{l} sex = 1, \\ education = 1 \end{array} \Bigg) =$$

$$\begin{pmatrix} 0.335 & 0.493 & 0.147 & 0.007 & 0.017 \\ 0.173 & 0.539 & 0.249 & 0.017 & 0.021 \\ 0.077 & 0.373 & 0.460 & 0.059 & 0.031 \\ 0.025 & 0.165 & 0.370 & 0.394 & 0.046 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

(assuming that the hazard are constant witin those two years)

- Long-term prediction derived using a piecewise-constant approximation of the time-dependent hazards

  For men aged 60 with higher education level:

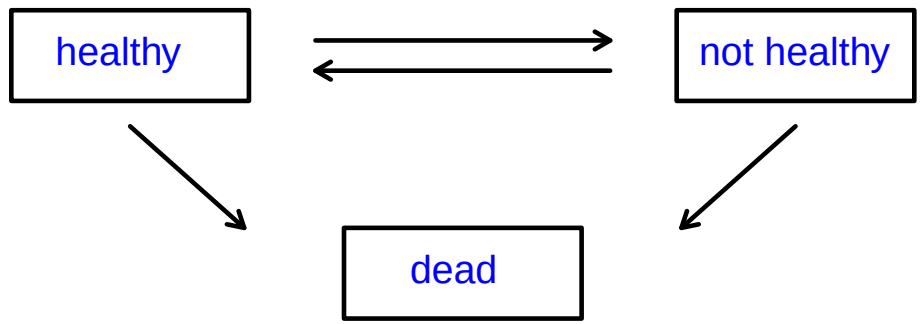


5. Conclusion

- Continuous-time survival models: there is theory & there is user-friendly `R` software for Gompertz models
- Piecewise-constant approximation of time dependency is important. This is not part of the model, but an aspect of the estimation
- Fitting the models is computationally intensive, and may need a bit of trial and error
- Methods which underlie the estimation of long-term transition probabilities (piecewise-constant approximation, and simulation for the CIs) are also the methods used to estimate state-specific life expectancies

Practical 3. Maximum likelihood estimation

- Aim: to understand the likelihood function and some of the computation that is involved
- Three-state process:



- Compare/check `msm` results
- Fit models with parameter constraints

Lecture 4

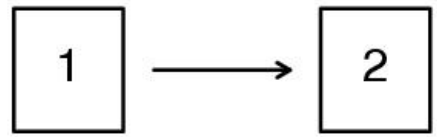
Life expectancies

Outline:

1. Mean survival & state-specific life expectancy
2. Marginal life expectancy
3. Estimation of life expectancies
4. The `ELECT` software
5. Conclusion

1. Mean survival & state-specific life expectancy

- In a standard survival model, mean survival is the expected time to be alive:



$$\mathrm{E}(T) = \int_0^\infty uf(u)du = \ldots = \int_0^\infty S(u)du$$

- Total residual life expectancy at age t_1 is the expected remaining years spent in state 1:

$$\mathrm{E}(T|t_1) = \int_0^\infty S(t_1+u|t_1)du = \int_0^\infty \mathrm{P}(Y_{t_1+u} = 1|Y_{t_1} = 1)du$$

State-specific residual life expectancy

- In a multi-state survival process, you can be alive in more than one state:
  - Assume that covariate values are known from age t_1 onwards; that is, $\mathcal{X}_{t_1} = \{\mathbf{x}(t) |\ t \geqslant t_1\}$ is known
  - Residual life expectancy (LE) in state s given state r at t_1:

$$e_{rs}(t_1) = e_{rs}(t_1|\mathcal{X}_{t_1}) = \int_0^\infty \mathrm{P}(Y_{t_1+u} = s |\ Y_{t_1} = r, \mathcal{X}_{t_1}) du$$

- If we know transition probabilities $\mathrm{P}(Y_{t_1+u} = s |\ Y_{t_1} = r, \mathcal{X}_{t_1})$, then LEs can be computed
- Transition probabilities can be estimated from the fitted multi-state model

- Residual life expectancy (*again*):

$$e_{rs}(t_1) = \int_0^{\infty} \mathrm{P}(Y_{t_1+u} = s |\ Y_{t_1} = r, \mathcal{X}_{t_1}) du$$

  - If $r \neq s$, then $e_{rs}(t_1)$ does not include any specification of time spent in r or in states other than state s

  - Say there are four living states. If $e_{14}(t_1)$ is two years, then this means only that starting in state 1 at specified age t_1, the expected time spent in state 4 is two years. How the rest of the time spent in the other living states is divided over the these states can only be derived from additional computation of $e_{11}(t_1)$, $e_{12}(t_1)$, and $e_{13}(t_1)$

2. Marginal life expectancy

- Marginal residual life expectancy is defined by

$$e_{\bullet s}(t_1) = \sum_{r \neq D} e_{rs}(t_1) \mathrm{P}(Y_{t_1} = r | \mathcal{X}_{t_1})$$

  It is the expected number of years spent in state s, irrespective of the initial state at age t_1

- $\mathrm{P}(Y_{t_1} = r | \mathcal{X}_{t_1})$ is the distribution of the living states at age t_1. Estimate this distribution using a multinomial logistic regression model

- Total residual life expectancy at age t_1 is defined by

$$e(t_1) = \sum_{s \neq D} e_{\bullet s}(t_1)$$

3. Estimation of life expectancies

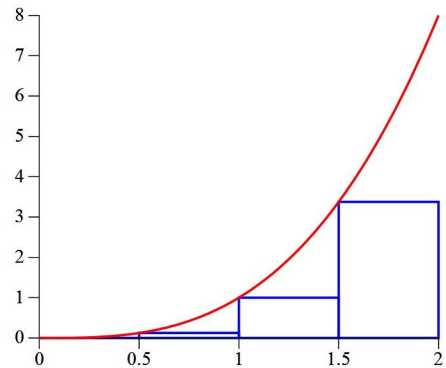
- The life expectancies can be derived from the fitted multi-state model and an additionally fitted multinomial logistic regression model. Of specific importance:
  - Specification of covariate values $\mathcal{X}_{t_1}$
  - Modelling of the time dependency of the multi-state process
  - Fixing assumed maximum age
- Computation of LEs can be based on integration or micro-simulation. In this workshop, only the first is discussed

- Estimation of LEs only as good as the estimated multi-state model!
  - If there is bias in the estimated model, then this bias is propagated in the LEs
  - Often, the LEs imply an extrapolation of the model beyond the age range in the study
  - Be careful with extrapolation across birth cohorts in LEs inference

- Residual life expectancy (*again*):

$$e_{rs}(t_1) = \int_0^\infty \mathrm{P}(Y_{t_1+u} = s |\ Y_{t_1} = r, \mathcal{X}_{t_1}) du$$

  - Numerical integration using simple grid approximation (pic taken from `en.wikipedia.org`):



  - For the integrand, a piecewise-constant approximation is used for the time-dependent hazards

    The same grid is used for piecewise-constant approximation **and** for the numerical approximation of the integral

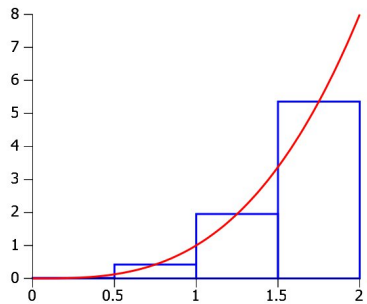
- Numerical integration provides a point estimate of LEs

- Asymptotic properties of the maximum likelihood estimator of the model parameters can be used to obtain standard errors or confidence intervals by simulation; see Mandel (2013)

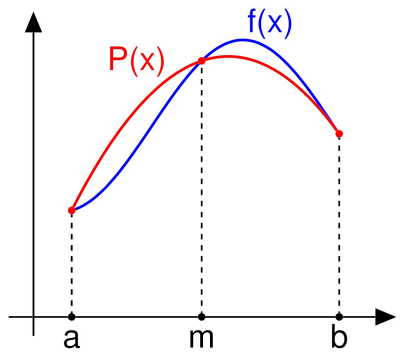
4. The ELECT software

- Functions for computation, summarising and plotting
- MLE simulation used to compute uncertainty
- Currently implemented for 3-state, 4-state, and 5-state models
- Newest version: integration method can be simple grid, or

Middle Riemann:



Simpson's:



(pics taken from `en.wikipedia.org`)

- There are three functions in `ELECT`:
  - `elect`
  - `summary.elect`
  - `plot.elect`

Function

`elect`

Description

Estimates state-specific and marginal life expectancies given fitted illness-death model in `msm`

Usage

```
elect(model, b.covariates, statedistdata,
time.scale.msm="years",h, age.max, S=0,
setseed=NULL)
```

Arguments

`model`	fitted `msm` model.
`b.covariates`	list with specified covariates values
`statedistdata`	data for distribution of living states
`time.scale.msm`	time scale in multi-state model ("years", ...)
`h`	grid parameter for integration
`max.age`	assumed maximum age in years
`S`	number of replications for simulation
`setseed`	seed for the random number generation

Function

`summary.elect`

Description

Summarises the life expectancies as estimated by `elect`

Usage

```
summary.elect(LEs,probs=c(.025,0.5,.975),digits=3,
print=TRUE)
```

Arguments

`LEs`	life expectancies estimated by `elect`
`probs`	numeric vector of probabilities for quantiles
`digits`	number of decimal places in output
`print`	TRUE for printing output to screen, FALSE otherwise

Function

`plot.elect`

Description

Graphical representation by smoothed densities of the distribution of the life expectancies as estimated by `elect`

Usage

```
plot.elect(LEs, kernel="gaussian", col="red",
lwd=2, cex.lab=1)
```

Arguments

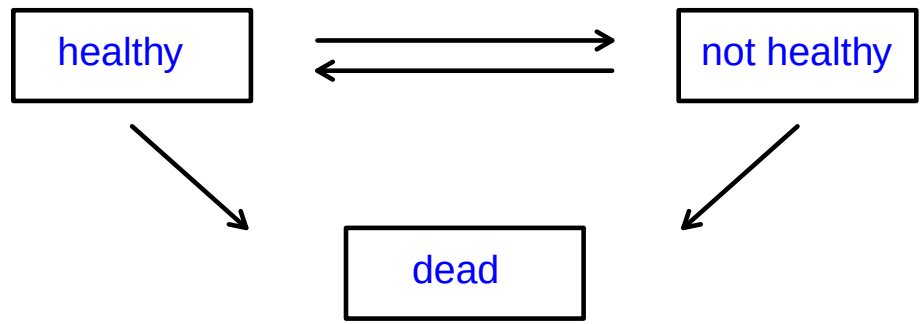
`LEs`	life expectancies estimated by `elect`
`kernel`	character string for smoothing kernel ("gaussian", ...)
`col`	colour of curve
`lwd`	line width of curve
`cex.lab`	magnification to be used for axis-labels

5. Conclusion

- Life expectancies are functions of parameters of the multi-state model. Software: `ELECT`
- Estimation of life expectancies is only as good as the fitted model
  - Bias in the estimated model is propagated in the estimation of LEs. Minor bias in the estimated effect of an explanatory variable may be innocuous, but when LEs are calculated the bias will be inflated
  - The definition of the LEs often implies an extrapolation of the model beyond age range in the data
- Micro-simulation is an alternative way to estimate LEs (Van den Hout *et al. 2014)*

Practical 4. Estimation of life expectancies

- Aim: to explore and understand the ELECT software
- Three-state process:



- How to deal with constraints on model parameters

Lecture 5

Results in practicals so far

Lecture 6

Advanced topics

Outline:

1. Misclassification of state
2. Multi-state models with B-splines
3. Missing data

1. Misclassification of state

- If a latent true state r is observed as state $r^* \neq r$, then this is called misclassification of state
  - **Example**: When a marker for a condition or disease is categorised in a series of living states using threshold values, misclassification can occur due to imprecise measurement
- If the Markov assumption is used for the latent process, then misclassification models are called *hidden Markov models*
- If misclassification of state is present, a model for the latent process has to be combined with a model for the misclassification

- Denote the observed process by Y_t^* and the latent by Y_t
- Misclassification probability for observed state r^* and latent state r is defined as

$$c_{rr^*} = \mathrm{P}(Y_t^* = r^* | Y_t = r)$$

- Commonly, misclassification probabilities are modelled using the logit link. An intercept-only model as an example:

$$c_{rr^*} = \frac{\exp(\alpha_{rr^*})}{1 + \exp(\alpha_{rr^*})}$$

- Fitting hidden Markov models is quite complex
- The `msm` software can estimate a wide range of hidden Markov models

- Whatever the model for the misclassification is, the model for the latent multi-state process is as discussed before
- With misclassification present, life expectancies can be estimated conditional on observed state **or** conditional on latent state
  - When the objective is to make a statement with regard to the whole population, for example concerning the planning of future health care, estimating LEs conditional on latent state makes more sense as need for care will be induced by the true latent state
  - When state-specific LEs for a given individual are the primary quantities of interest, it seems more reasonable to base the estimation on current observed state and taken into account the potential misclassification

- Be careful with comparing formally models with and without misclassification

- Check estimated misclassification probabilities. There may be an identifiability problem

- Including misclassification can be sometimes interpreted as smoothing data. For example, when measuring cognition in the older population, the measurement over time may result in a backward transition now and then, but the underlying latent cognitive process is assumed to be stable or decreasing

2. Multi-state models with B-splines

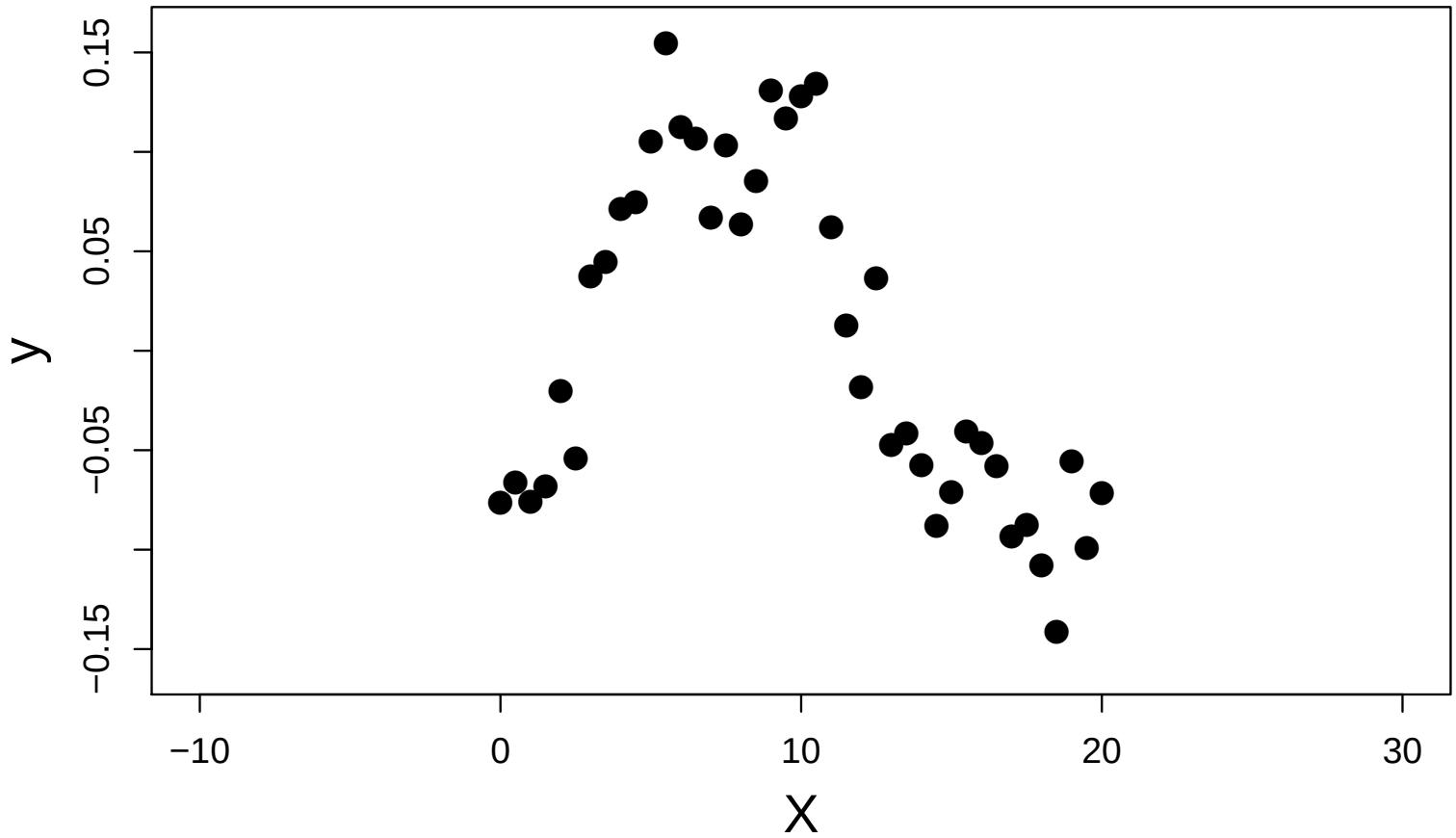
- More flexible multi-state models can be fitted using splines function to model the baseline hazard function:

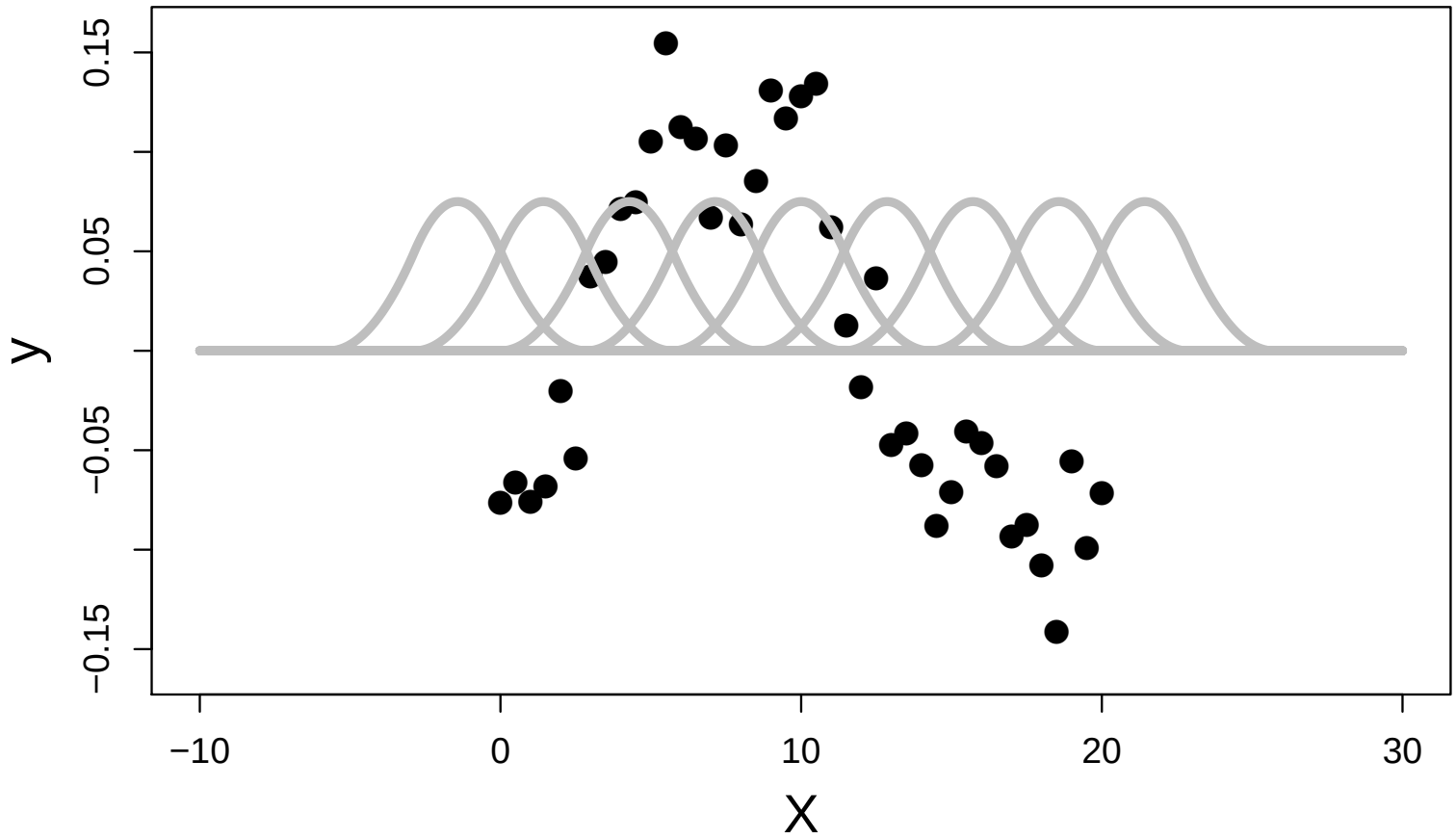
$$h_{rs.0}(t) \quad = \quad \exp\left(\sum_{k=1}^{K} \alpha_{rs.k} B_k(t)\right),$$

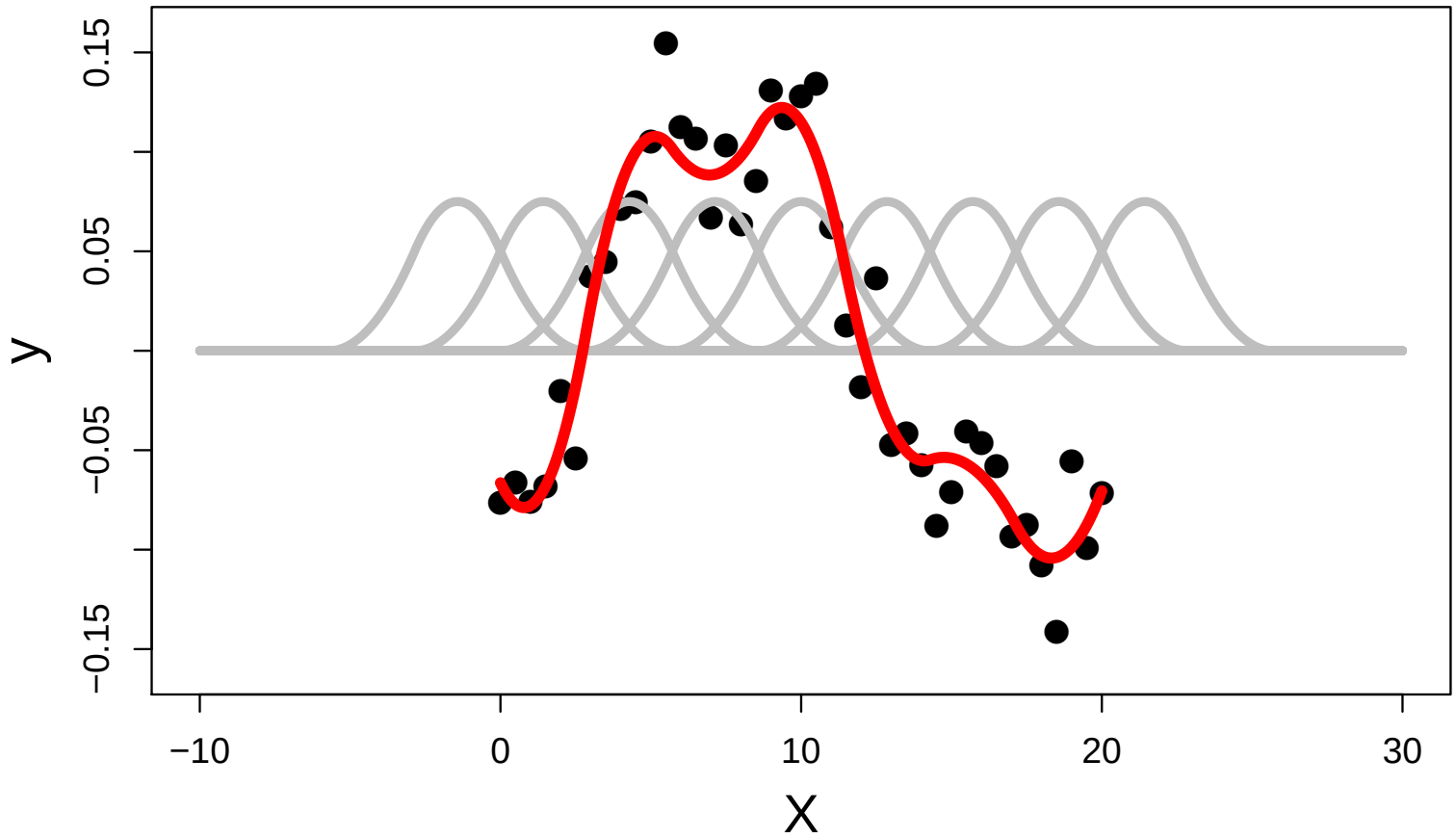
  where B_ks are determined by range of t in data, and αs are the parameters (Eilers and Marx 1996; Titman 2011)

- B-splines are polynomials pieces connected in a particular way

- They are easy to implement and allow for good flexibility

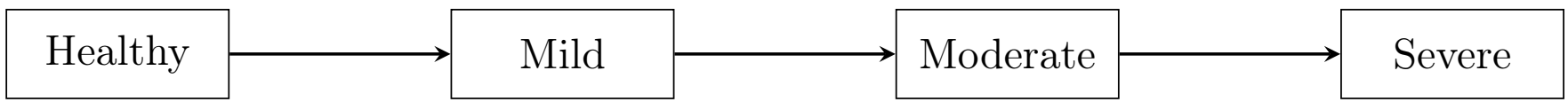

0.15
0.05
−0.05
−0.15
y
−10
0
10
20
30
X


0.15
0.05
−0.05
−0.15
y
−10
0
10
20
30
X


0.15
0.05
−0.05
−0.15
y
−10
0
10
20
30
X

Application to aneurysm data

- Abdominal aortic aneurysm is common in elderly men in the UK (Jackson at al. 2003)
- The aneur data consist of measurements of grades of aortic aneurysms
- The states represent successive degrees of aneurysm severity
- Modelling transition process is important because screening policies are defined w.r.t. the risk of moving across states
- Severe aneurysms are repaired by surgery

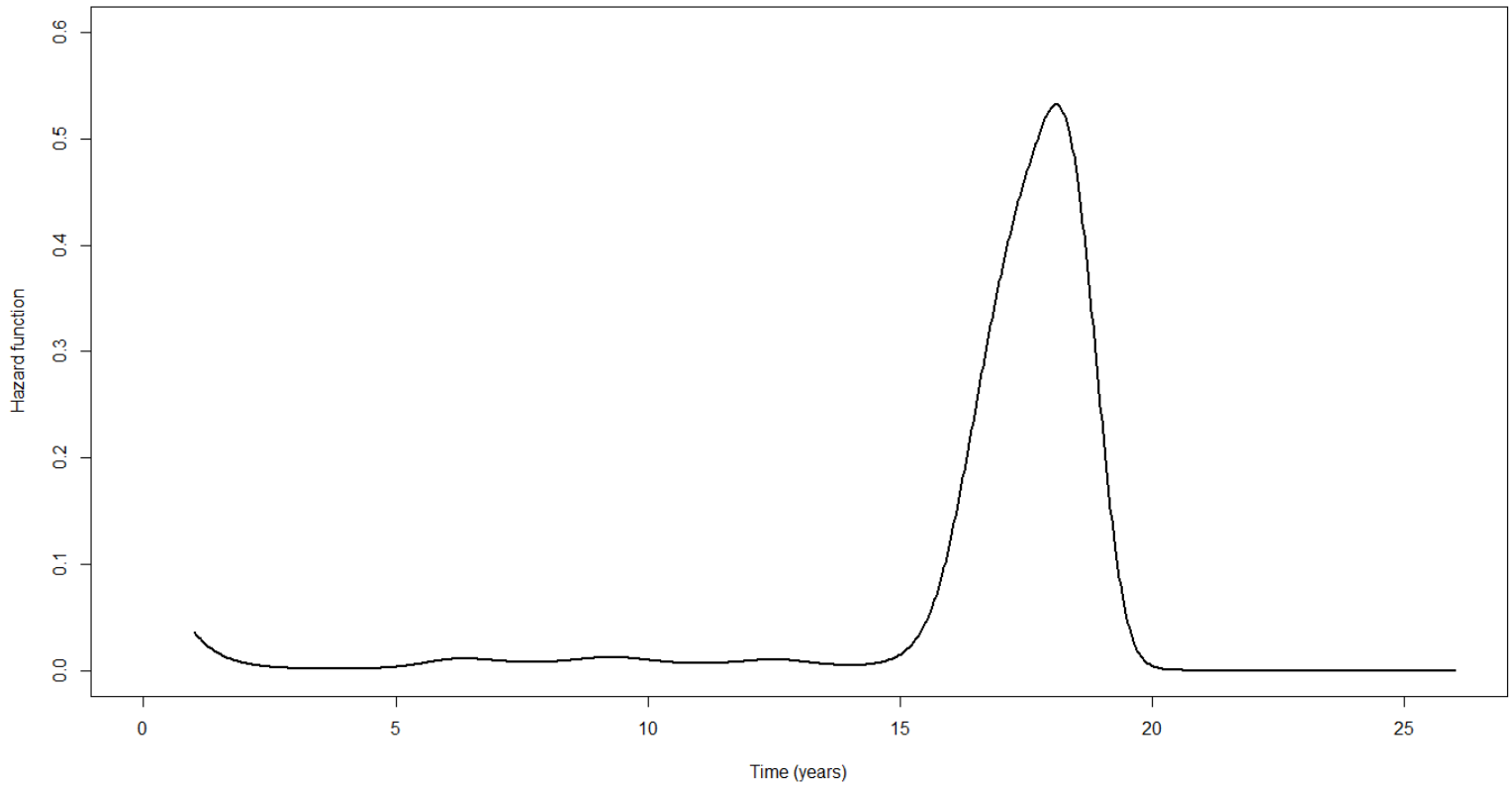


- Flexible multi-state models with B-splines can be used to estimate risk of diameter progression

- Hazards functions are specified in terms of a large set of B-splines basis

- Estimation is undertaken using (penalised) maximum likelihood estimation

- Fitted hazard transition from state 1 to state 2



3. Missing data

- Restrict discussion to models where data are subject to missing values of the state variable:
  - *Intermittently missing values*
  - Missing values due to *lost to follow-up*

- Intermittently missing values can be seen as interval-censored data with the caveat that the time interval for which no data are available is wider than planned for in the study design

- What to do when missing data is *non-ignorable*; that is, when the missing-data mechanism and the multi-state process are interlinked? Ignoring the missing data may lead to biased inference

- Approach for non-ignorable missingness: model probability of observing a state at time t conditional on the state at time t (Cole *et al. 2005; Van den Hout and Matthews 2010)*

- Let R_t denote the observation indicator at time t such that $R_t = 1$ if state Y_t is observed and $R_t = 0$ otherwise

  For an interval $(t_1, t_2]$, state Y_{t_2} and indicator R_{t_2} are modelled jointly:

$$\begin{aligned} \mathrm{P}(Y_{t_2}, R_{t_2} | Y_{t_1}) &= \mathrm{P}(R_{t_2} | Y_{t_2}, Y_{t_1}) \mathrm{P}(Y_{t_2} | Y_{t_1}) \\ &= \mathrm{P}(R_{t_2} | Y_{t_2}) \mathrm{P}(Y_{t_2} | Y_{t_1}) \end{aligned}$$

- Note that for $\mathrm{P}(Y_{t_2} | Y_{t_1})$ we have the multi-state model

Joint model (*again*):

$$\mathrm{P}(Y_{t_2}, R_{t_2} | Y_{t_1}) = \mathrm{P}(R_{t_2} | Y_{t_2}) \mathrm{P}(Y_{t_2} | Y_{t_1})$$

- We need a model for $\mathrm{P}(R_{t_2} | Y_{t_2})$. Use logistic regression models

- **Example** for the three-state model with two living states:
  - One model for $R_t = 1$ conditional on $Y_t = 1$, and one model for $R_t = 1$ conditional on $Y_t = 2$
  - Logistic regression for $y \in \{1, 2\}$

$$\mathrm{P}(R_t = 1 | Y_t = y, x_t) = \frac{\exp(1 + \alpha_y x_t)}{1 + \exp(1 + \alpha_y x_t)}$$

- If we have the missing values, then we can use the models above

- The missing values are missing—that is the problem

- Can write the likelihood contribution of an individual by summing over all possible missing values

  Example: For a progressive three-state survival model, only patterns with monotone increase are possible. For example, if $\mathbf{y} = (1, ?, ?, 2, 3)$, then the possible trajectories are given by $\{(1,1,1,2,3), (1,1,2,2,3), (1,2,2,2,3)\}$

- Satten and Longini (1996) present elegant matrix method for the above. Can use a general-purpose optimiser for likelihood inference. This may take a while