A Semi-Markov model for interval-censored data

Venediktos Kapetanakis[1], Ardo van den Hout, Fiona E Matthews

MRC Biostatistics Unit, Institute of Public Health, Cambridge, UK

[1] E-mail: venediktos.kapetanakis@mrc-bsu.cam.ac.uk

A Semi-Markov model for interval-censored data

Overview

1. Background Information.

2. Statistical Methods.

   Method A: Integration.
   Method B: Multiple Imputation.

3. Results.

4. Discussion & Future work.

Background Information
- Multi-state modelling
- Stroke
- Aims

Statistical Methods

Method A

Method B

Results

Discussion
& Future work

Acknowledgements

Background Information

MRC Biostatistics Unit
UNIVERSITY OF CAMBRIDGE

Background Information
- Multi-state modelling
- Stroke
- Aims

Statistical Methods

Method A

Method B

Results

Discussion & Future work

Acknowledgements

Multi-state modelling

Multi-state modelling is a method of analysing longitudinal data when the observed outcome is a categorical variable.

- Useful in medical applications where the levels of a disease can be regarded as the states of the model (*Kalbfleisch & Lawless, 1985; Commenges, 1999*) .
- A common assumption is that the data satisfy the first order Markov assumption under which the transition to the next state depends only on the current state, i.e. the history of the process is ignored.
- This assumption may often be inappropriate.

Background Information
- Multi-state modelling
- Stroke
- Aims

Statistical Methods

Method A

Method B

Results

Discussion & Future work

Acknowledgements

Information about stroke

- A stroke is the loss of brain function as a result of a disorder in the blood supply to the brain.
- Non-fatal stroke may cause permanent neurological damage and adult disability.
- It is one of the leading causes of death worldwide.

MRC Biostatistics Unit
UNIVERSITY OF CAMBRIDGE

Background Information
- Multi-state modelling
- Stroke
- Aims

Statistical Methods

Method A

Method B

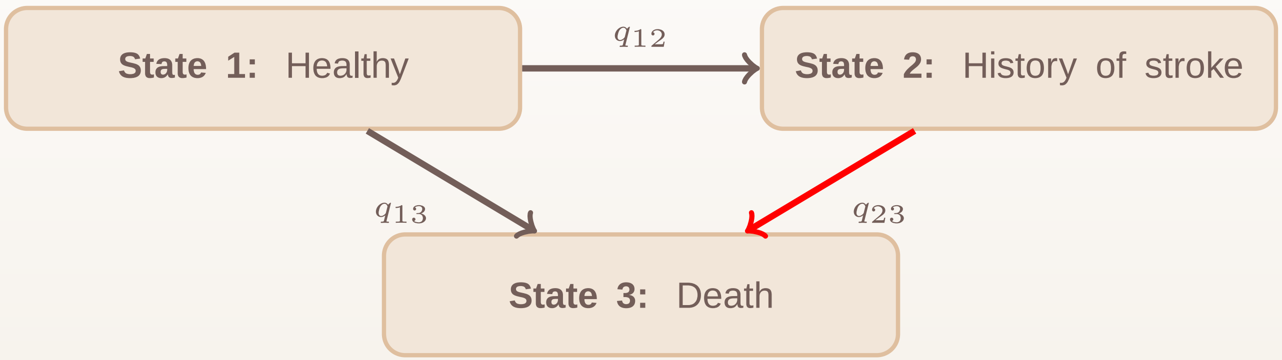
Results

Discussion & Future work

Acknowledgements

Aims

- Investigate the transition intensities in the three-state model:



- Relax the Markov assumption by adjusting the transition intensity from state 2 to state 3 for the time spent in state 2.

Method A: Integrate out all possible times for the transition from state 1 to state 2 in the estimation of the likelihood for every individual.

Method B: Estimate the unknown transition time from state 1 to state 2.

Background Information
Statistical Methods
- Data source
- Primary outcome
- Model specification
- Data patterns

Method A
Method B
Results
Discussion & Future work
Acknowledgements

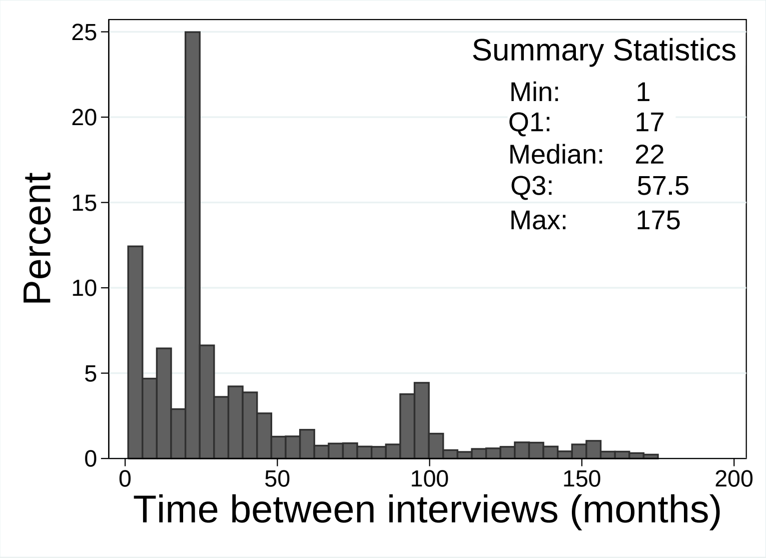
Statistical Methods

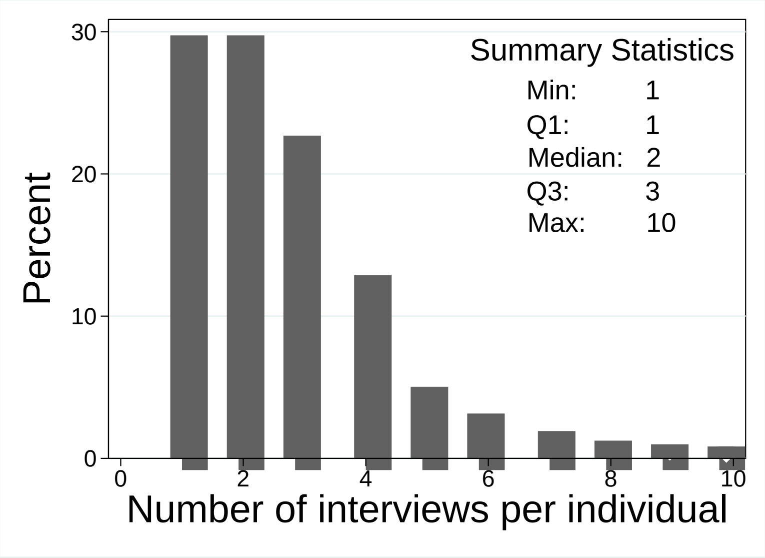
Background Information
Statistical Methods
- Data source
- Primary outcome
- Model specification
- Data patterns

Method A
Method B
Results
Discussion & Future work
Acknowledgements

Data source

- We used data from the UK Medical Research Council Cognitive Function and Ageing Study[1].
- 2321 individuals aged 65 and above had up to 9 interviews from 1991 to date where they were asked if they had any previous history of stroke.





[1] www.cfas.ac.uk

Background Information

Statistical Methods
- Data source
- Primary outcome
- Model specification
- Data patterns

Method A

Method B

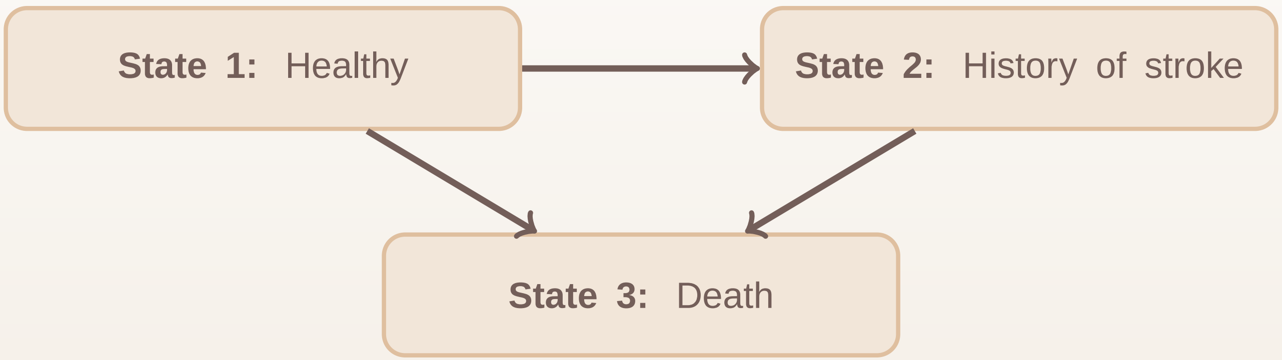
Results

Discussion & Future work

Acknowledgements

Primary outcome

- Our primary outcome was a categorical variable with three levels:



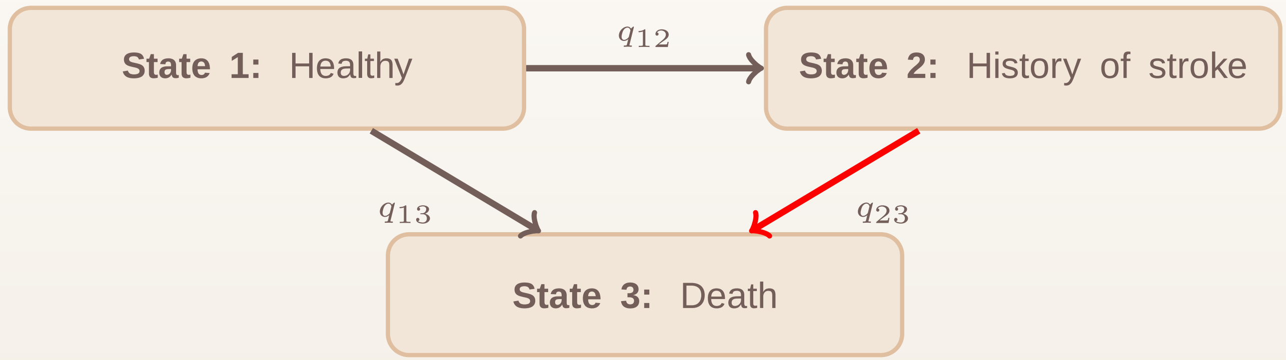
- Observed number of transitions:

	To:	Healthy	Stroke	Dead	Censored	Total
From:	Healthy	2966	113	1331	711	5121
	Stroke	0	304	224	55	583
	Total	2966	417	1555	766	5704

MRC Biostatistics Unit
UNIVERSITY OF CAMBRIDGE

Model specification

We used these data to fit the three-state Semi-Markov model, bellow:



$$\begin{aligned}\log(q_{1s}) = b_0^{1s} \quad &+ \quad b_1^{1s}(\text{age}) + b_2^{1s}(\text{male}) + b_3^{1s}(\text{education}) \\ &+ \quad b_4^{1s}(\text{current smoker}), \qquad \text{s = 2 or 3}\end{aligned}$$

$$\begin{aligned}\log(q_{23}) = b_0^{23} \quad &+ \quad b_1^{23}(\text{age}) + b_2^{23}(\text{male}) + b_3^{23}(\text{education}) \\ &+ \quad b_4^{23}(\text{current smoker}) + b_5^{23}(\text{time spent in state 2})\end{aligned}$$

Background Information
Statistical Methods
● Data source
● Primary outcome
● Model specification
● Data patterns
Method A
Method B
Results
Discussion & Future work
Acknowledgements

Observed data patterns

Without loss of generality we may assume that the observed data patterns can be summarised by the following table:

Data patterns	Observed states		
	First	**Intermediate**	**Last**
A	1	2	death
B	1	2	censored
C	1	1	death
D	1	1	censored
E	2	2	death
F	2	2	censored

Background Information
Statistical Methods
● Data source
● Primary outcome
● Model specification
● Data patterns
Method A
Method B
Results
Discussion & Future work
Acknowledgements

Observed data patterns

Without loss of generality we may assume that the observed data patterns can be summarised by the following table:

Data patterns		Observed states: First		Intermediate		Last
A		1	$\ast$	2		death
B		1	$\ast$	2		censored
C		1		1	?	death
D		1		1	?	censored
E	$\ast$	2		2		death
F	$\ast$	2		2		censored

$\ast$: Transition from state 1 to state 2.

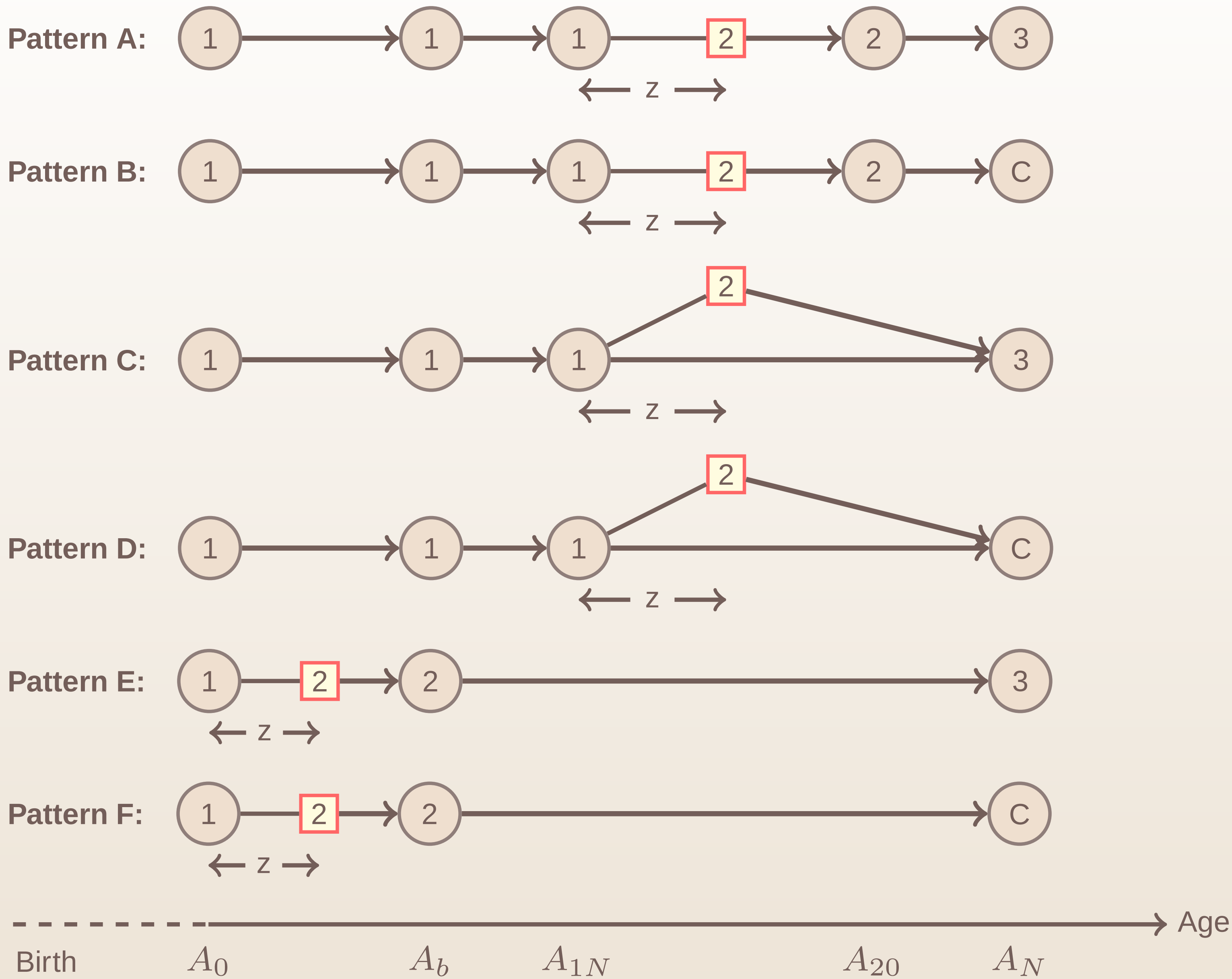
? : Possible transition from state 1 to state 2.

MRC Biostatistics Unit
UNIVERSITY OF CAMBRIDGE

Background Information
Statistical Methods
Method A
- Transition patterns
- Likelihood contributions
- Piecewise-constant approach
Method B
Results
Discussion & Future work
Acknowledgements

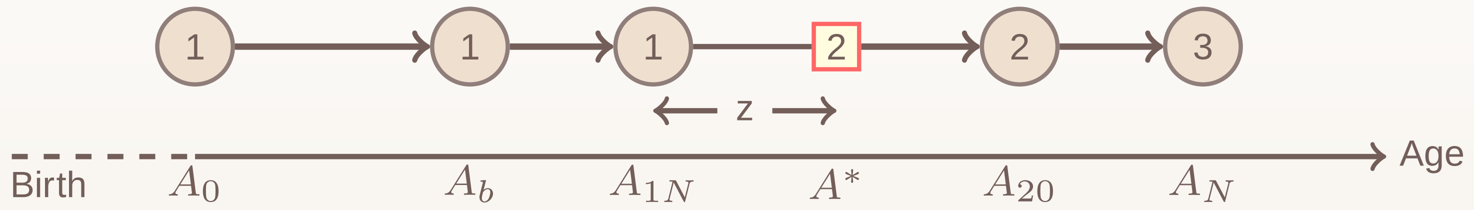
Method A

Transition patterns



Likelihood contributions (Pattern A)

Pattern A



$$L_A \quad = \quad \int_0^{A_{20}-A_{1N}} f(A_{1N} - A_0 + z; q_{12} + q_{13})\pi_{12}\, f(A_N - A_{1N} - z; q_{23})\pi_{23}\, dz$$

where

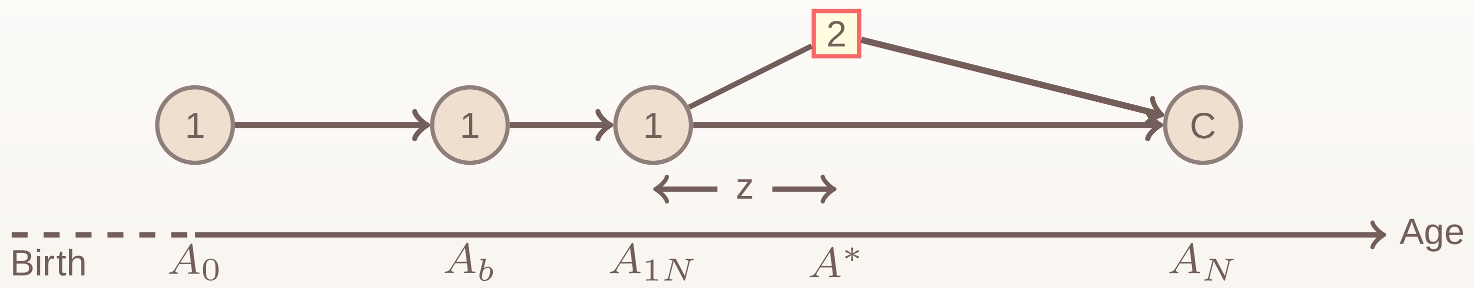
$$f(t;\lambda) = \begin{cases} \lambda \exp(-\lambda t) & t \geq 0 \\ 0 & t < 0 \end{cases} \quad , \qquad F(t;\lambda) = \begin{cases} 1 - \exp(-\lambda t) & t \geq 0 \\ 0 & t < 0 \end{cases}$$

and

$$\Pi = \begin{pmatrix} \pi_{11} & \pi_{12} & \pi_{13} \\ \pi_{21} & \pi_{22} & \pi_{23} \\ \pi_{31} & \pi_{32} & \pi_{33} \end{pmatrix} = \begin{pmatrix} 0 & \frac{q_{12}}{q_{12}+q_{13}} & \frac{q_{13}}{q_{12}+q_{13}} \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}$$

Likelihood contributions (Pattern D)

Pattern D



$$L_D = \int_0^{A_N - A_{1N}} f(A_{1N} - A_0 + z; q_{12} + q_{13})\pi_{12}\left[1 - F(A_N - A_{1N} - z; q_{23})\right] dz + $$
$$+ \left[1 - F(A_N - A_0; q_{12} + q_{23})\right]$$

where

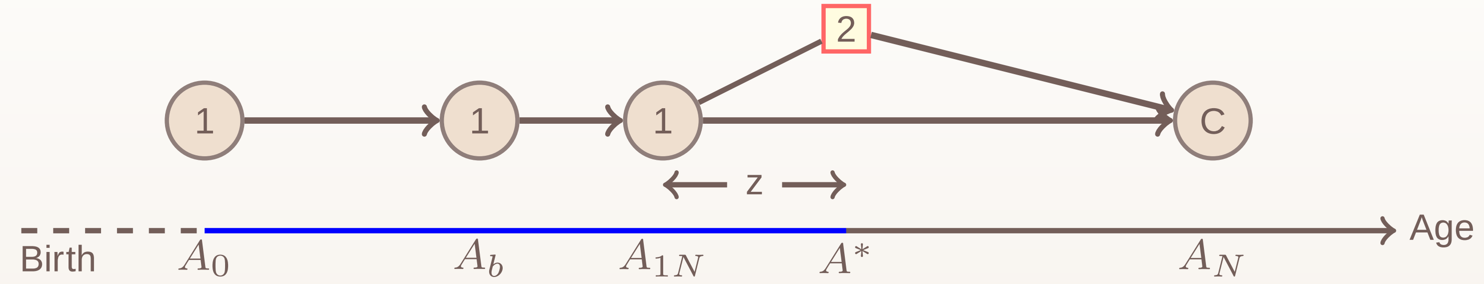
$$f(t;\lambda) = \begin{cases} \lambda \exp(-\lambda t) & t \geq 0 \\ 0 & t < 0 \end{cases}, \quad F(t;\lambda) = \begin{cases} 1 - \exp(-\lambda t) & t \geq 0 \\ 0 & t < 0 \end{cases}$$

and

$$\Pi = \begin{pmatrix} \pi_{11} & \pi_{12} & \pi_{13} \\ \pi_{21} & \pi_{22} & \pi_{23} \\ \pi_{31} & \pi_{32} & \pi_{33} \end{pmatrix} = \begin{pmatrix} 0 & \frac{q_{12}}{q_{12}+q_{13}} & \frac{q_{13}}{q_{12}+q_{13}} \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}$$

Likelihood contributions (Pattern D)

Pattern D



$$L_D = \int_0^{A_N - A_{1N}} f(A_{1N} - A_0 + z; q_{12} + q_{13})\pi_{12}\left[1 - F(A_N - A_{1N} - z; q_{23})\right] dz + $$
$$+ \left[1 - F(A_N - A_0; q_{12} + q_{23})\right]$$

where

$$f(t;\lambda) = \begin{cases} \lambda \exp(-\lambda t) & t \geq 0 \\ 0 & t < 0 \end{cases} \quad , \quad F(t;\lambda) = \begin{cases} 1 - \exp(-\lambda t) & t \geq 0 \\ 0 & t < 0 \end{cases}$$

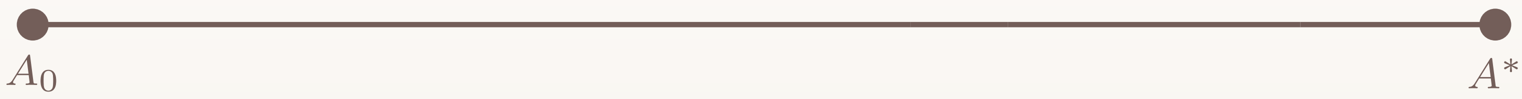
and

$$\Pi = \begin{pmatrix} \pi_{11} & \pi_{12} & \pi_{13} \\ \pi_{21} & \pi_{22} & \pi_{23} \\ \pi_{31} & \pi_{32} & \pi_{33} \end{pmatrix} = \begin{pmatrix} 0 & \frac{q_{12}}{q_{12}+q_{13}} & \frac{q_{13}}{q_{12}+q_{13}} \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}$$

Background Information
Statistical Methods
Method A
- Transition patterns
- Likelihood contributions
- Piecewise-constant approach

Method B
Results
Discussion & Future work
Acknowledgements

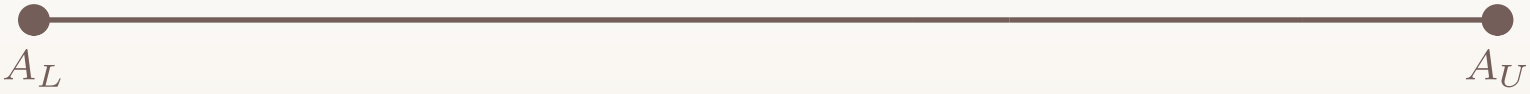
Piecewise-constant approach (1)



- We needed to calculate:

$$f(A^* - A_0; q_{12} + q_{13})$$

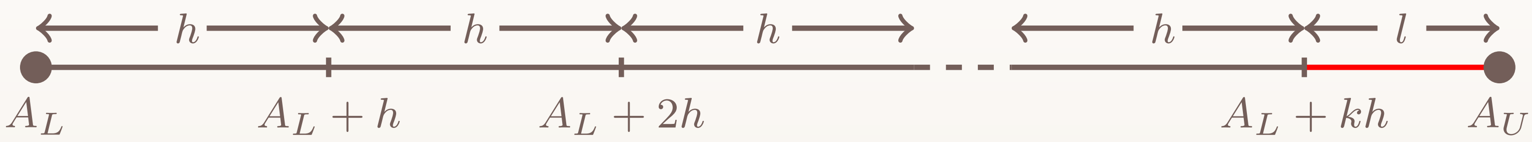
Piecewise-constant approach (2)



- Generally, we needed to calculate:

$$f(A_U - A_L; Q)$$
$$1 - F(A_U - A_L; Q)$$

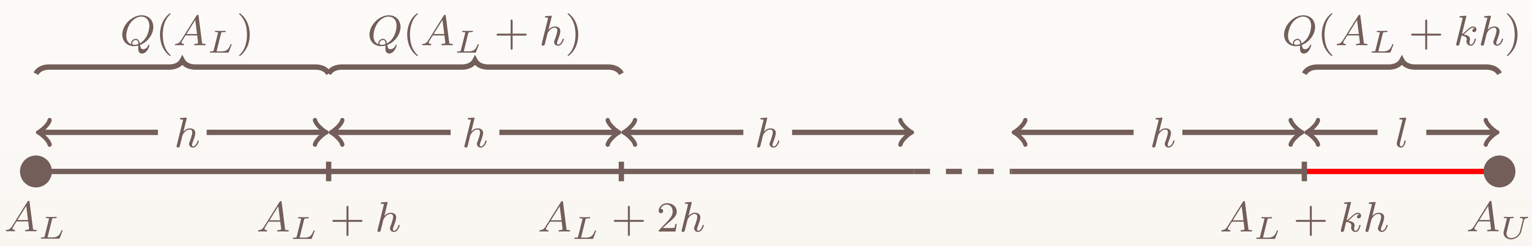
Piecewise-constant approach (3)



- We specified the resolution, h, for the piecewise-constant approach.
- We split the interval into k subintervals of length h so that: $(A_U - A_L) - kh = l \leq h$.

Background Information
Statistical Methods
Method A
- Transition patterns
- Likelihood contributions
- Piecewise-constant approach
Method B
Results
Discussion & Future work
Acknowledgements

Piecewise-constant approach (3)

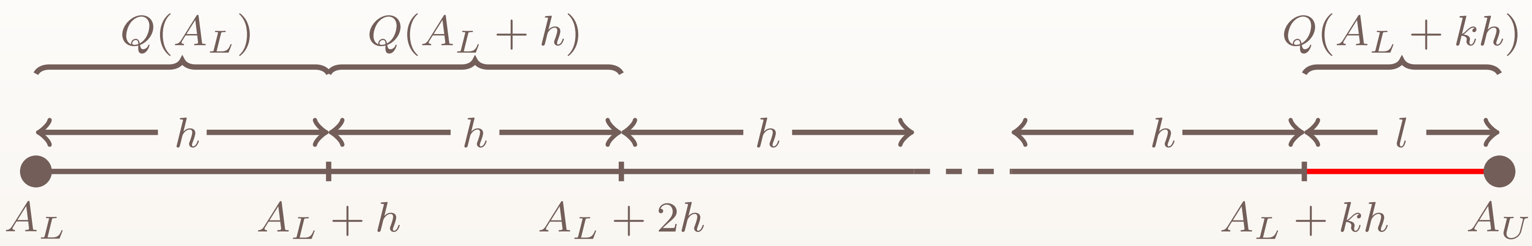


- We specified the resolution, h, for the piecewise-constant approach.
- We split the interval into k subintervals of length h so that: $(A_U - A_L) - kh = l < h$.
- In every subinterval, we evaluated the transition intensities $Q(\text{Age})$ at:

$$\text{Age} = \text{age at the left subinterval limit}$$

Background Information
Statistical Methods
Method A
- Transition patterns
- Likelihood contributions
- Piecewise-constant approach
Method B
Results
Discussion & Future work
Acknowledgements

Piecewise-constant approach (4)



- We obtained:

$$f(A_U - A_L; Q) = \left\{ \prod_{i=0}^{k-1} \Big[1 - F\big(h; Q(A_L + ih)\big) \Big] \right\} f\big(l; Q(A_L + kh)\big)$$

$$1 - F(A_U - A_L; Q) = \left\{ \prod_{i=0}^{k-1} \Big[1 - F\big(h; Q(A_L + ih)\big) \Big] \right\} \Big[1 - F\big(l; Q(A_L + kh)\big) \Big]$$

Background Information
Statistical Methods
Method A
Method B
- Data patterns
- Interval regression
- Multiple imputation
Results
Discussion & Future work
Acknowledgements

Method B

Background Information
Statistical Methods
Method A
Method B
● Data patterns
● Interval regression
● Multiple imputation
Results
Discussion & Future work
Acknowledgements

Observed data patterns

Data patterns		Observed states First		Intermediate		Last
A		1	$\ast$	2		death
B		1	$\ast$	2		censored
C		1		1	?	death
D		1		1	?	censored
E	$\ast$	2		2		death
F	$\ast$	2		2		censored

$\ast$: Transition from state 1 to state 2.

? : Possible transition from state 1 to state 2.

Background Information
Statistical Methods
Method A
Method B
- Data patterns
- Interval regression
- Multiple imputation
Results
Discussion & Future work
Acknowledgements

Estimation of the exact transition time (1)

- We fit an interval regression model for *age* at the transition time, A^*, adjusted for several covariates using data from patterns A, B, E and F:
  - age at baseline
  - sex
  - education status
  - smoking
  - data pattern
  - observed sojourn time in state 2
  - the interaction of the last two

MRC Biostatistics Unit
UNIVERSITY OF CAMBRIDGE

Background Information
Statistical Methods
Method A
Method B
- Data patterns
- Interval regression
- Multiple imputation
Results
Discussion & Future work
Acknowledgements

Estimation of the exact transition time (2)

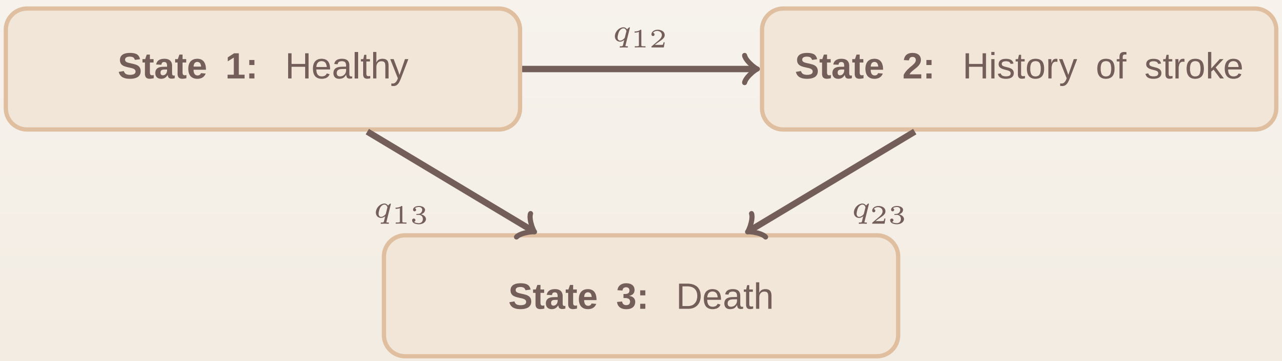
- Using mean and variance estimates conditional on the covariate specifications of a single individual, we assumed normality to find the density of A^*, $f(a^*)$ for that specific individual.

- We obtained
$$f(a^*|a^* \in I) = \frac{f(a^*)}{\mathbb{P}(a^* \in I)},$$
where I was the interval within which the transition from state 1 to state 2 occurred.

Background Information
Statistical Methods
Method A
Method B
- Data patterns
- Interval regression
- Multiple imputation
Results
Discussion & Future work
Acknowledgements

Multiple imputation

- 100 values were imputed from $f(a^*|a^* \in I)$, producing 100 imputed data sets in which the computation of the time spent in state 2 was straightforward.
- For each data set we fit the multi-state model:



- Results were combined using Rubin's multiple imputation rules (*Rubin, 1987*).

Background Information
Statistical Methods
Method A
Method B
Results
- Method A Results
- Method B Transition time
- Method B Results
- Methods Comparison

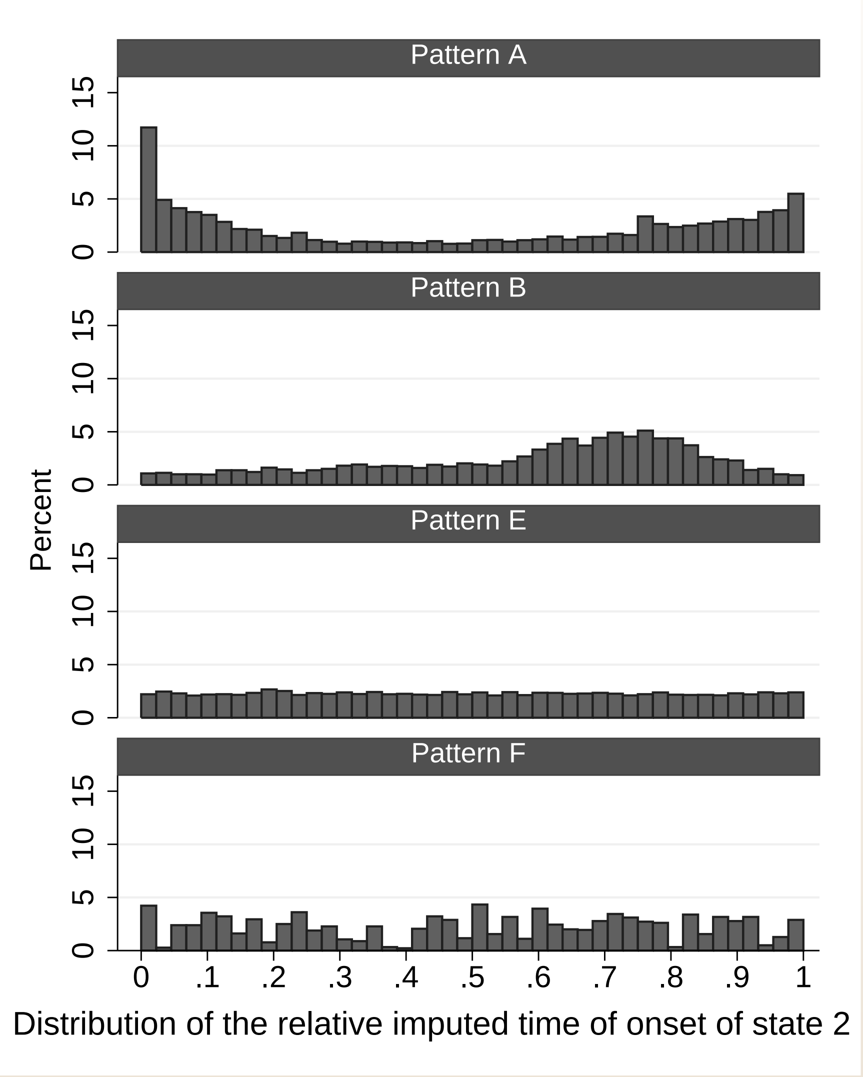
Discussion & Future work
Acknowledgements

Results

Method A	Coefficient	Mean (S.E.)		
Age	b_1^{12}	0.105	(0.007) $*$	
(years)	b_1^{13}	0.166	(0.004) $*$	
	b_1^{23}	0.104	(0.009) $*$	
Sex	b_2^{12}	0.437	(0.129) $*$	
(males versus females)	b_2^{13}	0.414	(0.070) $*$	
	b_2^{23}	0.457	(0.142) $*$	
Education	b_3^{12}	-0.300	(0.153) $[*]$	
(10 years or more)	b_3^{13}	-0.226	(0.077) $*$	
	b_3^{23}	0.202	(0.167)	
Smoking	b_4^{12}	0.336	(0.129) $*$	
(current versus never/ex)	b_4^{13}	0.538	(0.069) $*$	
	b_4^{23}	0.372	(0.139) $*$	
Time spent in State 2 (years)	b_5^{23}	-0.004	(0.013)	$\longleftarrow$ p-value = 0.78

Relative imputed time of onset of state 2 (1)

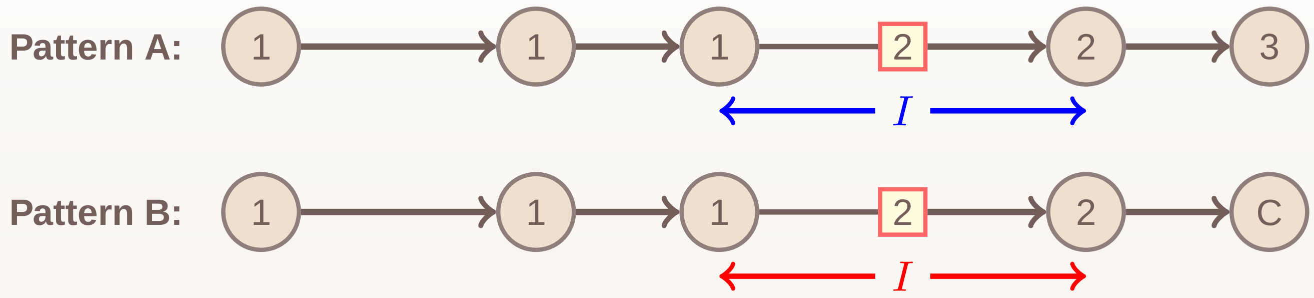
Distribution of the relative imputed time for the transition from state 1 to state 2 within I, by data pattern.



- In pattern A, the unknown transition seems more likely to take place closer to the interval limits.
- In pattern B, the distribution is skewed to the left $(p < 0.001)$.
- In patterns E and F, the unknown transition seems to be uniformly distributed within the interval.

Background Information
Statistical Methods
Method A
Method B
Results
- Method A Results
- Method B Transition time
- Method B Results
- Methods Comparison
Discussion & Future work
Acknowledgements

Relative imputed time of onset of state 2 (2)



Distribution of the *length* of the interval (in months) within which the transition from state 1 to state 2 could happen:

	N	Min	Q_1	Median	Q_3	Max
Pattern A	76	1	22	24	60	104
Pattern B	37	12	24	88	96	119

Method B	Coefficient	Mean (S.E.)	
Age	b_1^{12}	0.071	(0.003) $*$
(years)	b_1^{13}	0.123	(0.003) $*$
	b_1^{23}	0.090	(0.005) $*$
Sex	b_2^{12}	0.257	(0.126) $*$
(males versus females)	b_2^{13}	0.323	(0.080) $*$
	b_2^{23}	0.309	(0.123) $*$
Education	b_3^{12}	-0.275	(0.152) $\dagger$
(10 years or more)	b_3^{13}	-0.240	(0.090) $*$
	b_3^{23}	0.110	(0.146)
Smoking	b_4^{12}	0.352	(0.126) $*$
(current versus never/ex)	b_4^{13}	0.487	(0.078) $*$
	b_4^{23}	0.389	(0.124) $*$
Time spent in State 2 (years)	b_5^{23}	-0.001	(0.0005) $*$

Wald test taking into account the multiple imputation (*Reiter & Raghunathan, 2007*)

$\longleftarrow$ p-value = 0.02

Covariate	Coefficient	Method A Mean (S.E.)	Method B Mean (S.E.)
Age	b_1^{12}	0.105 (0.007) $*$	0.071 (0.003) $*$
(years)	b_1^{13}	0.166 (0.004) $*$	0.123 (0.003) $*$
	b_1^{23}	0.104 (0.009) $*$	0.090 (0.005) $*$
Sex	b_2^{12}	0.437 (0.129) $*$	0.257 (0.126) $*$
(males versus females)	b_2^{13}	0.414 (0.070) $*$	0.323 (0.080) $*$
	b_2^{23}	0.457 (0.142) $*$	0.309 (0.123) $*$
Education	b_3^{12}	-0.300 (0.153) $[*]$	-0.275 (0.152) $\dagger$
(10 years or more)	b_3^{13}	-0.226 (0.077) $*$	-0.240 (0.090) $*$
	b_3^{23}	0.202 (0.167)	0.110 (0.146)
Smoking	b_4^{12}	0.336 (0.129) $*$	0.352 (0.126) $*$
(current versus never/ex)	b_4^{13}	0.538 (0.069) $*$	0.487 (0.078) $*$
	b_4^{23}	0.372 (0.139) $*$	0.389 (0.124) $*$
Time spent in State 2 (years)	b_5^{23}	-0.004 (0.013)	-0.001 (0.0005) $*$

Background Information
Statistical Methods
Method A
Method B
Results
Discussion & Future work
- Conclusions
- Extensions
Acknowledgements

Discussion & Future work

Background Information
Statistical Methods
Method A
Method B
Results
Discussion & Future work
- Conclusions
- Extensions
Acknowledgements

Conclusions

- Semi-Markov models that incorporate history of the process may improve inference when the Markov assumption is inappropriate.
- Assuming that the unknown transition time from state 1 to state 2 occurs midway *may* be inappropriate for interval-censored data.

Background Information
Statistical Methods
Method A
Method B
Results
Discussion & Future work
- Conclusions
- Extensions
Acknowledgements

Extensions and work in progress

- In Method B, adjust for possible unobserved $1 \rightarrow 2$ transitions in patterns C and D:
  - Simulate state at the time just before death or at censoring using Bernoulli trials.
  - If the simulated state is state 2, impute age at the transition time using interval regression.
- Sensitivity analysis for the age where everyone is assumed to be healthy.
- Run a simulation study to compare:
  - The more theoretical model (Method A).
  - The model where the transition time is imputed using interval regression (Method B).
  - The model where the transition time is assumed to occur midway.

Background Information
Statistical Methods
Method A
Method B
Results
Discussion & Future work
Acknowledgements

Acknowledgements

- We would like to thank all study participants and their relatives for their invaluable contribution.
- Funding source: MRC CFAS[1] is funded by the UK Medical Research Council.

[1] www.cfas.ac.uk

Background Information
Statistical Methods
Method A
Method B
Results
Discussion & Future work
Acknowledgements
Thank you!

Thank you!

Background Information
Statistical Methods
Method A
Method B
Results
Discussion & Future work
Acknowledgements
Extra Material
- Markov models
- Data patterns
- Likelihood contributions
- Interval regression (1)
- Interval regression (2)
- References

Extra Material

Background Information
Statistical Methods
Method A
Method B
Results
Discussion & Future work
Acknowledgements
Extra Material
- Markov models
- Data patterns
- Likelihood contributions
- Interval regression (1)
- Interval regression (2)
- References

Markov models

- *Homogeneous Markov* model: $q_{rs} = q_{rs}$ for all t.
- *Non-homogeneous Markov* model: $q_{rs} = q_{rs}(t)$.
- *Semi-Markov* model: $q_{rs} = q_{rs}(\tau)$ where τ is the time spent in the present state.
- *Partial-Markov* model: $q_{rs} = q_{rs}(\mathbf{z}_s)$ where $\mathbf{z}_s$ is a multivariate predictable process with components which may be stochastic or deterministic.

Background Information
Statistical Methods
Method A
Method B
Results
Discussion & Future work
Acknowledgements
Extra Material
- Markov models
- Data patterns
- Likelihood contributions
- Interval regression (1)
- Interval regression (2)
- References

Observed data patterns

Without loss of generality we may assume that the observed data patterns can be summarised by the following table:

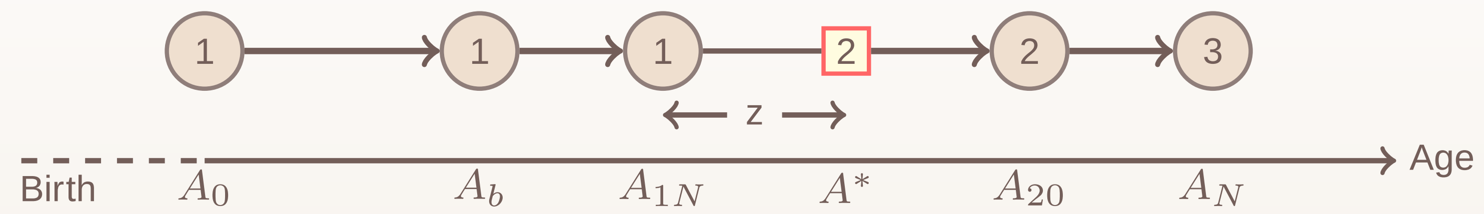
Data patterns	**N**		**Observed states** First		**Observed states** Intermediate		**Observed states** Last
A	76		1	$\ast$	2		death
B	37		1	$\ast$	2		censored
C	1331		1		1	?	death
D	711		1		1	?	censored
E	148	$\ast$	2		2		death
F	18	$\ast$	2		2		censored

$\ast$: Transition from state 1 to state 2.

? : Possible transition from state 1 to state 2.

Likelihood contributions (Pattern A)

Pattern A



$$L_A \quad = \quad \int_0^{A_{20}-A_{1N}} f(A_{1N} - A_0 + z; q_{12} + q_{13})\pi_{12}\, f(A_N - A_{1N} - z; q_{23})\pi_{23}\, dz$$

where

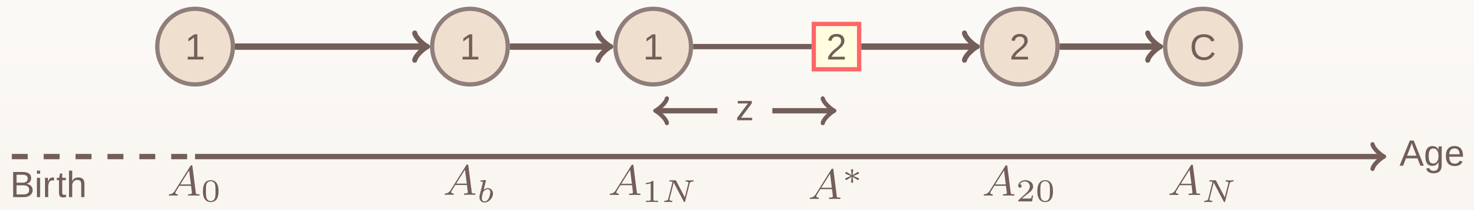
$$f(t;\lambda) = \begin{cases} \lambda \exp(-\lambda t) & t \geq 0 \\ 0 & t < 0 \end{cases} \quad , \qquad F(t;\lambda) = \begin{cases} 1 - \exp(-\lambda t) & t \geq 0 \\ 0 & t < 0 \end{cases}$$

and

$$\Pi = \begin{pmatrix} \pi_{11} & \pi_{12} & \pi_{13} \\ \pi_{21} & \pi_{22} & \pi_{23} \\ \pi_{31} & \pi_{32} & \pi_{33} \end{pmatrix} = \begin{pmatrix} 0 & \frac{q_{12}}{q_{12}+q_{13}} & \frac{q_{13}}{q_{12}+q_{13}} \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}$$

Likelihood contributions (Pattern B)

Pattern B



$$L_B = \int_0^{A_{20}-A_{1N}} f(A_{1N} - A_0 + z; q_{12} + q_{13})\pi_{12}\big[1 - F(A_N - A_{1N} - z; q_{23})\big] dz$$

where

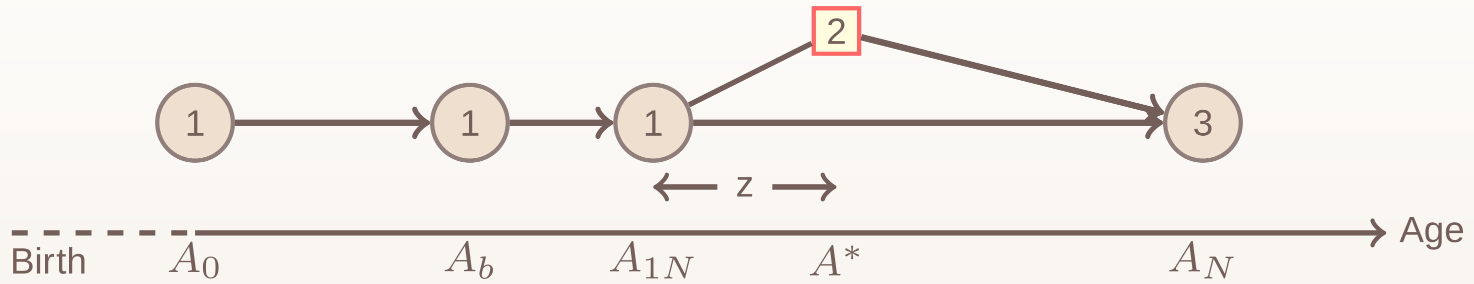
$$f(t;\lambda) = \begin{cases} \lambda \exp(-\lambda t) & t \geq 0 \\ 0 & t < 0 \end{cases} \quad , \qquad F(t;\lambda) = \begin{cases} 1 - \exp(-\lambda t) & t \geq 0 \\ 0 & t < 0 \end{cases}$$

and

$$\Pi = \begin{pmatrix} \pi_{11} & \pi_{12} & \pi_{13} \\ \pi_{21} & \pi_{22} & \pi_{23} \\ \pi_{31} & \pi_{32} & \pi_{33} \end{pmatrix} = \begin{pmatrix} 0 & \frac{q_{12}}{q_{12}+q_{13}} & \frac{q_{13}}{q_{12}+q_{13}} \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}$$

Likelihood contributions (Pattern C)

Pattern C



$$L_C = \int_0^{A_N - A_{1N}} f(A_{1N} - A_0 + z; q_{12} + q_{13})\pi_{12}\, f(A_N - A_{1N} - z; q_{23})\pi_{23}\, dz + $$

$$+ f(A_N - A_0; q_{12} + q_{23})\pi_{13}$$

where

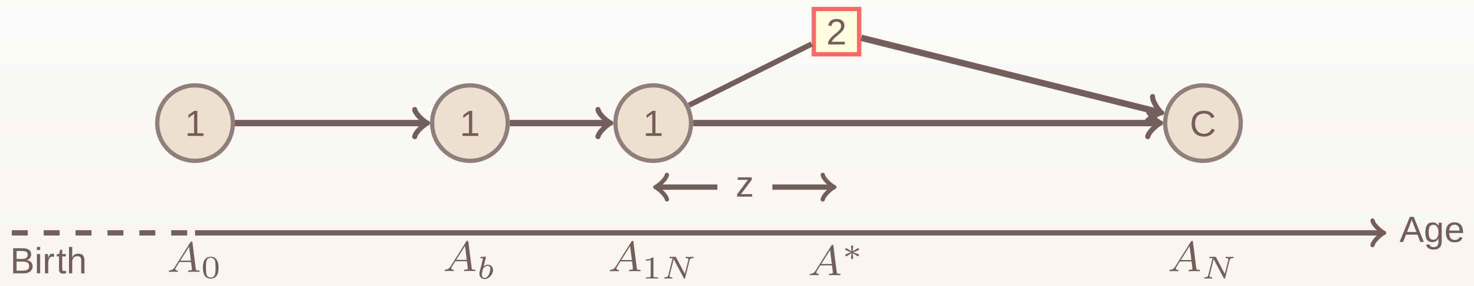
$$f(t;\lambda) = \begin{cases} \lambda \exp(-\lambda t) & t \geq 0 \\ 0 & t < 0 \end{cases} \quad , \qquad F(t;\lambda) = \begin{cases} 1 - \exp(-\lambda t) & t \geq 0 \\ 0 & t < 0 \end{cases}$$

and

$$\Pi = \begin{pmatrix} \pi_{11} & \pi_{12} & \pi_{13} \\ \pi_{21} & \pi_{22} & \pi_{23} \\ \pi_{31} & \pi_{32} & \pi_{33} \end{pmatrix} = \begin{pmatrix} 0 & \frac{q_{12}}{q_{12}+q_{13}} & \frac{q_{13}}{q_{12}+q_{13}} \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}$$

Likelihood contributions (Pattern D)

Pattern D



$$L_D = \int_0^{A_N - A_{1N}} f(A_{1N} - A_0 + z; q_{12} + q_{13})\pi_{12}\left[1 - F(A_N - A_{1N} - z; q_{23})\right] dz + \\ + \left[1 - F(A_N - A_0; q_{12} + q_{23})\right]$$

where

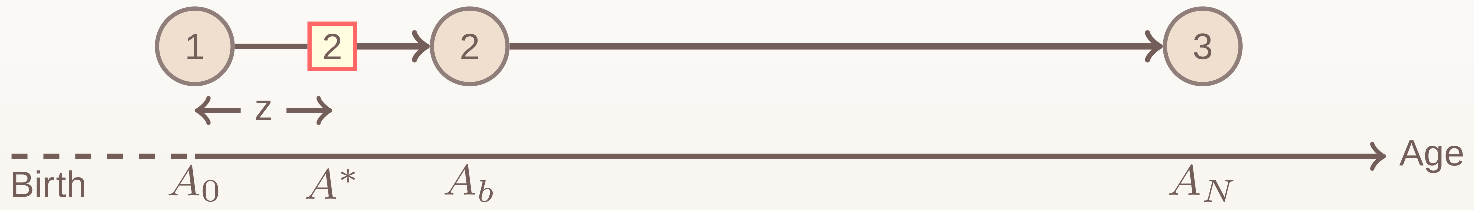
$$f(t;\lambda) = \begin{cases} \lambda\exp(-\lambda t) & t \geq 0 \\ 0 & t < 0 \end{cases} \quad , \quad F(t;\lambda) = \begin{cases} 1 - \exp(-\lambda t) & t \geq 0 \\ 0 & t < 0 \end{cases}$$

and

$$\Pi = \begin{pmatrix} \pi_{11} & \pi_{12} & \pi_{13} \\ \pi_{21} & \pi_{22} & \pi_{23} \\ \pi_{31} & \pi_{32} & \pi_{33} \end{pmatrix} = \begin{pmatrix} 0 & \frac{q_{12}}{q_{12}+q_{13}} & \frac{q_{13}}{q_{12}+q_{13}} \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}$$

Likelihood contributions (Pattern E)

Pattern E



$$L_E \quad = \quad \int_0^{A_b - A_0} f(z; q_{12} + q_{13}) \pi_{12} \, f(A_N - A_0 - z; q_{23}) \pi_{23} \, dz$$

where

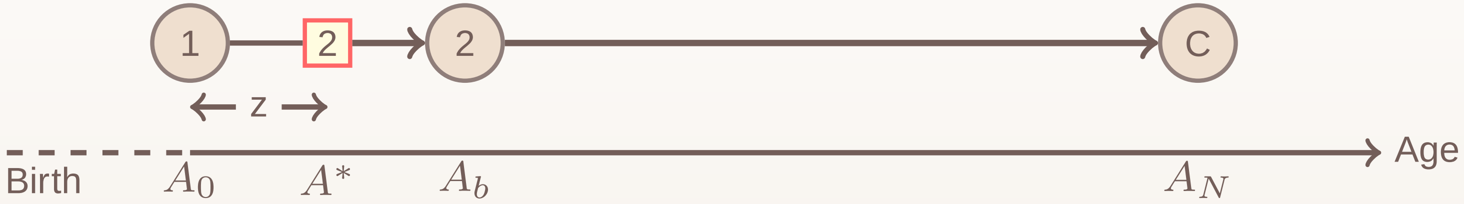
$$f(t;\lambda) = \begin{cases} \lambda \exp(-\lambda t) & t \geq 0 \\ 0 & t < 0 \end{cases} \quad , \qquad F(t;\lambda) = \begin{cases} 1 - \exp(-\lambda t) & t \geq 0 \\ 0 & t < 0 \end{cases}$$

and

$$\Pi = \begin{pmatrix} \pi_{11} & \pi_{12} & \pi_{13} \\ \pi_{21} & \pi_{22} & \pi_{23} \\ \pi_{31} & \pi_{32} & \pi_{33} \end{pmatrix} = \begin{pmatrix} 0 & \frac{q_{12}}{q_{12}+q_{13}} & \frac{q_{13}}{q_{12}+q_{13}} \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}$$

Likelihood contributions (Pattern F)

Pattern F



$$L_F \quad = \quad \int_0^{A_b - A_0} f(z; q_{12} + q_{13})\pi_{12}\big[1 - F(A_N - A_0 - z; q_{23})\big]\,dz$$

where

$$f(t;\lambda) = \begin{cases} \lambda \exp(-\lambda t) & t \geq 0 \\ 0 & t < 0 \end{cases} \quad , \quad F(t;\lambda) = \begin{cases} 1 - \exp(-\lambda t) & t \geq 0 \\ 0 & t < 0 \end{cases}$$

and

$$\Pi = \begin{pmatrix} \pi_{11} & \pi_{12} & \pi_{13} \\ \pi_{21} & \pi_{22} & \pi_{23} \\ \pi_{31} & \pi_{32} & \pi_{33} \end{pmatrix} = \begin{pmatrix} 0 & \frac{q_{12}}{q_{12}+q_{13}} & \frac{q_{13}}{q_{12}+q_{13}} \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}$$

Background Information
Statistical Methods
Method A
Method B
Results
Discussion & Future work
Acknowledgements
Extra Material
- Markov models
- Data patterns
- Likelihood contributions
- Interval regression (1)
- Interval regression (2)
- References

Interval regression (1)

- Interval regression fits models for data where each observation represents:
  - interval-censored data
  - left-censored data
  - right-censored data
  - point data
- In this type of regression, if the value for the i-th individual is known to lie within an interval $[y_{L_i}, y_{R_i}]$, then the likelihood contribution from this individual is simply $P(y_{L_i} \leq Y_i \leq y_{R_i})$.
- Similarly, for left-censored and right-censored data, the likelihood contribution consists of terms of the form $P(Y_i \leq y_{L_i})$ and $P(Y_i \geq y_{R_i})$, respectively.

MRC Biostatistics Unit
UNIVERSITY OF CAMBRIDGE

Background Information
Statistical Methods
Method A
Method B
Results
Discussion & Future work
Acknowledgements
Extra Material
- Markov models
- Data patterns
- Likelihood contributions
- Interval regression (1)
- Interval regression (2)
- References

Interval regression (2)

- Let $\mathbf{y} = \mathbf{X}\beta + \epsilon$ be the model.
- We assume that $\epsilon \sim N(\mathbf{0}, \sigma^2\mathbf{I})$.
- The log-likelihood is:

$$
\begin{aligned}
\log L = \quad & - \frac{1}{2}\sum_{i \in C}\left\{\left(\frac{y_i - x\beta}{\sigma}\right)^2 + \log \pi\sigma^2\right\} \\
& + \sum_{i \in L} \log \Phi\left(\frac{y_{L_i} - x\beta}{\sigma}\right) \\
& + \sum_{i \in R} \log\left\{1 - \Phi\left(\frac{y_{R_i} - x\beta}{\sigma}\right)\right\} \\
& + \sum_{i \in I} \log\left\{\Phi\left(\frac{y_{R_i} - x\beta}{\sigma}\right) - \Phi\left(\frac{y_{L_i} - x\beta}{\sigma}\right)\right\}
\end{aligned}
$$

Background Information
Statistical Methods
Method A
Method B
Results
Discussion & Future work
Acknowledgements
Extra Material
- Markov models
- Data patterns
- Likelihood contributions
- Interval regression (1)
- Interval regression (2)
- References

References

[1] Michael G. Kenward ; James Carpenter. Multiple imputation: current perspectives. *Statistical Methods in Medical Research*, 16:199–218, 2007.

[2] D. Commenges. Multi-state models in epidemiology. *Lifetime Data Analysis*, 5:315–327, 1999.

[3] William H. Greene. *Econometric Analysis*. Pearson Education International, 2003.

[4] C. Jackson. *Multi-state modelling with R: the msm package*.

[5] J. D. Kalbfleisch ; J. F. Lawless. The analysis of panel data under a markov assumption. *Journal of the American Statistical Association*, 80:863–871, 1985.

[6] J.R. Norris. *Markov Chains*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 1997.

[7] J. P. Reiter ; T. E. Raghunathan. The multiple adaptations of multiple imputation. *Journal of the American Statistical Association*, 102:1462–1471, 2007.

[8] Donald R. Rubin. *Multiple imputation for nonresponse in surveys*. John Wiley & Sons, Ltd., 1987.

[9] Stata Corporation. *STATA Manual*, STATA/SE 10 edition, 2007.