The Analysis of Interval-Censored, Semi-Competing Risks Data in the Presence of Informative Loss-to-Followup

J. Barrett, F. Siannis and V. Farewell

MRC Biostatistics Unit, Cambridge, UK

30th Annual Conference of the International Society for Clinical Biostatistics

Outline

1. The MRC Cognitive Function and Ageing Study (MRC CFAS)
2. Modelling the MRC CFAS Data
3. Results
4. Discussion

The MRC Cognitive Function and Ageing Study (MRC CFAS)

- UK-based longitudinal multicentre study looking at health and cognitive function of older people. The centres are Newcastle, Nottingham, Liverpool, Cambridgeshire, Gwynedd and Oxford.
- We are interested in modelling transitions from the healthy state to cognitive impairment and/or death, and investigating the dependence of transition rates on explanatory variables.
- Cognitive impairment (CI) was assessed using the Mini-Mental State Examination (MMSE), a widely used test of memory and cognitive function with scores ranging from 0 to 30. A score ≤ 21 was taken to indicate CI.
- All participants were flagged in the National Health Service Central Registry, so exact death times are known.
- Participants who were assessed as CI at the prevalence screen, plus a random sample of participants assessed as healthy, attended assessments every 1 to 2 years over a 10-year period. The remaining participants were assessed less frequently with gaps of up to 8 years between assessments.

Modelling the MRC CFAS Data

We wish to model two processes: cognitive impairment (CI) and fatality.

- The CI process is censored by the fatal process.
- But the fatal process is not censored by the CI process.

So we have semi-competing risks, which we will model using a multi-state model.

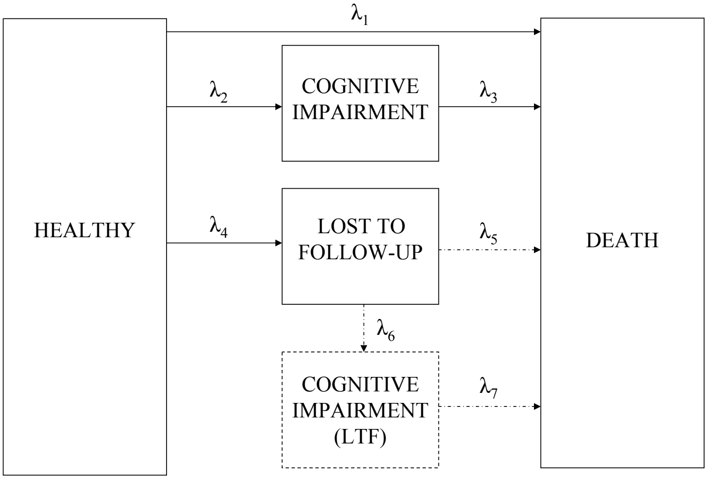
Censoring for the CI process may be informative, whereas the fatal process is censored only at the end of the study.

There are gaps of up to 8 years between interviews, so we must consider our data to be interval-censored for the CI process. We have exact transition times for the fatal process.

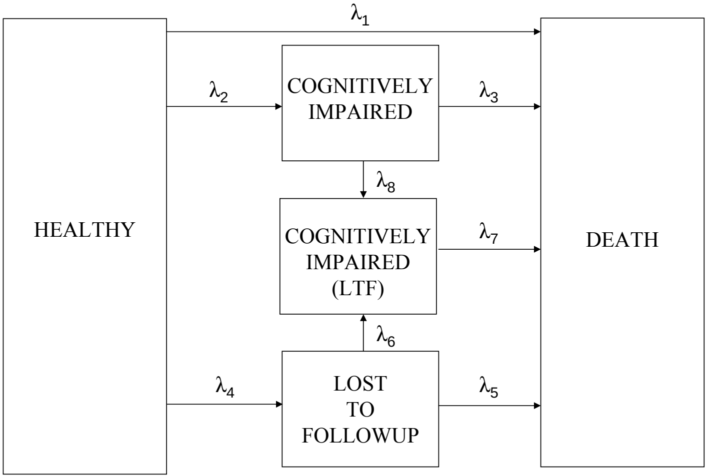
Siannis, Farewell and Head have fitted a five-state model for semi-competing risks data with informative censoring for the case where exact transition times are known (*Statistics in Medicine* 2007; **26**; 426-442).

In the model of Siannis et al all participants were initially in the healthy state. For the MRC CFAS data we will model the probability of being cognitively impaired at prevalence screen.

Model 1


HEALTHY
COGNITIVE IMPAIRMENT
LOST TO FOLLOW-UP
COGNITIVE IMPAIRMENT (LTF)
DEATH
λ_1
λ_2
λ_3
λ_4
λ_5
λ_6
λ_7

Model 2


HEALTHY
COGNITIVELY IMPAIRED
COGNITIVELY IMPAIRED (LTF)
LOST TO FOLLOWUP
DEATH
λ_1
λ_2
λ_3
λ_4
λ_5
λ_6
λ_7
λ_8

Observable Routes

There are 12 observable routes in this model:

1. $\mathrm{H} \rightarrow \mathrm{H}$
2. $\mathrm{H} \rightarrow \mathrm{CI}$
3. $\mathrm{H} \rightarrow \mathrm{CI} \rightarrow \mathrm{D}$
4. $\mathrm{H} \rightarrow \mathrm{D}$
5. $\mathrm{H} \rightarrow \mathrm{LTF}$
6. $\mathrm{H} \rightarrow \mathrm{LTF} \rightarrow \mathrm{D}$
7. $\mathrm{CI} \rightarrow \mathrm{CI}$
8. $\mathrm{CI} \rightarrow \mathrm{D}$
9. $\mathrm{H} \rightarrow \mathrm{CI} \rightarrow \mathrm{CI(LTF)}$
10. $\mathrm{H} \rightarrow \mathrm{CI} \rightarrow \mathrm{CI(LTF)} \rightarrow \mathrm{D}$
11. $\mathrm{CI} \rightarrow \mathrm{CI(LTF)}$
12. $\mathrm{CI} \rightarrow \mathrm{CI(LTF)} \rightarrow \mathrm{D}$

Some observable routes correspond to more than one model trajectory e.g. observable route 5 corresponds to:

$$\mathrm{H} \rightarrow \mathrm{LTF} \quad \text{and} \quad \mathrm{H} \rightarrow \mathrm{LTF} \rightarrow \mathrm{CI(LTF)}$$

The Likelihood

The likelihood function is a product of the probabilities of observable routes

$$\ell = \prod_{i=1}^{n} \prod_{j=1}^{8} Q_{ji}^{I_{ji}}$$

- Q_{ji} is the probability of observing route j for subject i
- $I_{ji} = 1$ if subject i takes route j, and 0 otherwise

We assume a logistic model for the probability of a subject being CI at the prevalence screen.

We assume that transition rates depend only on the last state visited, and not on any previous history of states visited. We assume Weibull hazards:

$$\lambda_m(t|\mathbf{x}_i) = e^{\mathbf{v}_m' \mathbf{x}_i} \alpha_m t^{\alpha_m - 1}$$

- α_m is the shape parameter for transition m, $m = 1, \ldots, 8$
- $\mathbf{x}_i$ are explanatory variables for subject i
- $\mathbf{v}_m$ are explanatory variable parameters for transition m

Model Assumptions

Not all parameters in the model are identifiable, so we make the following constraints:

$$\lambda_3(t) = \lambda_7(t)$$

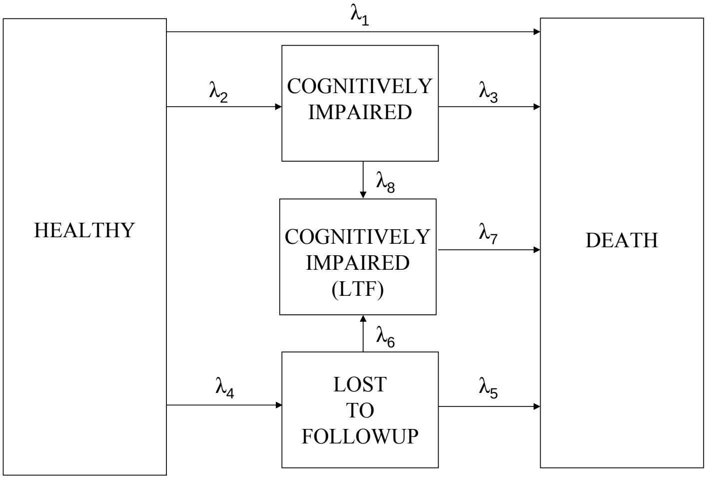
and

$$\frac{\lambda_1(t)}{\lambda_2(t)} = k\frac{\lambda_5(t)}{\lambda_6(t)}$$

We make the following additional assumptions:

- Cognitive impairment cannot be misclassified.
- For those who become both CI and LTF, the death rate is independent of the order in which CI and LTF occurred.
- Those who die within 2 years of their last healthy observation are assumed not to pass through the CI or LTF states prior to death.
- Similarly, those who die within 2 years of their last CI observation are assumed not to pass through the LTF state.

Model 2


HEALTHY
COGNITIVELY IMPAIRED
COGNITIVELY IMPAIRED (LTF)
LOST TO FOLLOWUP
DEATH
λ1
λ2
λ3
λ4
λ5
λ6
λ7
λ8

Proportional Hazards Analyses

We will fit our model to data from the Newcastle centre, with 2452 participants. We use age and sex as explanatory variables.

Proportional hazards analysis for time to fatal event with (a) CI and (b) LTF indicators as time-dependent explanatory variables:

Hazard ratio (95% CI)			
	(a)		(b)
CI-Indic	2.544 (2.11 - 3.063)	LTF-Indic	1.148 (1.029 - 1.280)
Age group		*Age group*	
Below average	1	Below average	1
Above average	2.617 (2.29 - 2.990)	Above average	2.544 (2.298 - 2.815)
Gender		*Gender*	
Male	1	Male	1
Female	0.762 (0.67 - 0.869)	Female	0.745 (0.672 - 0.826)

Multi-State Model Results

Maximum likelihood estimates with $k = 1$:

Death transition parameters:

	λ_1 $\mathrm{H} \rightarrow \mathrm{D}$	λ_3 $\mathrm{CI} \rightarrow \mathrm{D}$	λ_5 $\mathrm{LTF} \rightarrow \mathrm{D}$
Intercept	0.052 (0.005)	0.094 (0.015)	0.024 (0.007)
Age	0.082(0.007)	0.053 (0.007)	0.087 (0.011)
Sex	-0.602 (0.094)	-0.328 (0.112)	-0.552 (0.129)
Shape	1.156 (0.045)	1.321 (0.057)	1.636 (0.117)

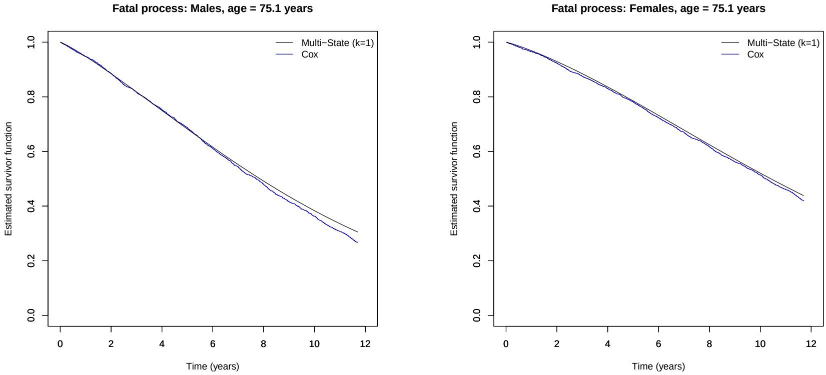
Other transition parameters:

	λ_2 $\mathrm{H} \rightarrow \mathrm{CI}$	λ_4 $\mathrm{H} \rightarrow \mathrm{LTF}$	λ_8 $\mathrm{CI} \rightarrow \mathrm{CI(LTF)}$
Intercept	0.013 (0.002)	0.137 (0.011)	1.710 (0.774)
Age	0.133 (0.011)	0.014 (0.006)	-0.023 (0.011)
Sex	0.455 (0.162)	0.404 (0.081)	0.096 (0.180)
Shape	1.163 (0.072)	0.515 (0.026)	0.092 (0.044)

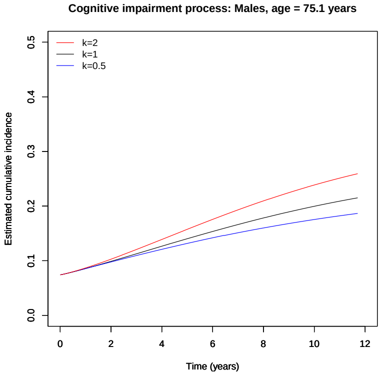
λ_6 and λ_7 are defined by the constraints.

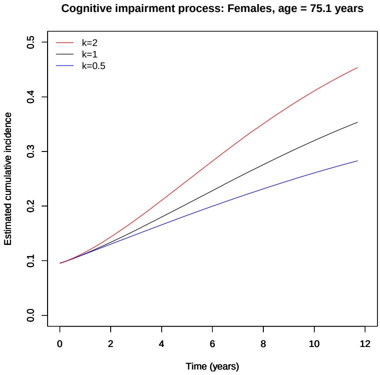
A likelihood ratio test comparing this model to a non-informative LTF model with $\lambda_1 = \lambda_5$ and $\lambda_2 = \lambda_6$ gave a test statistic of 16.3 on 4 degrees of freedom, p-value 0.003.

The Fatal Process



The Cognitive Impairment Process





Discussion

- We have fitted a multi-state model to data from the Newcastle centre of the MRC CFAS study, that takes into account informative loss-to-followup and interval censoring.
- Not all parameters in our model are identifiable, and it was therefore necessary to introduce some constraints. The parameter k in our model must be fixed, and it is therefore necessary to test the sensitivity of results to various values of k.
- Here we used sex and age as explanatory variables, although it would be easy to add more to our model. However, the number of explanatory variables might be constrained by the runtime of the program that optimizes the loglikelihood.
- In the future we may look at more complicated models, e.g. investigating the effect of having a stroke on the risk of cognitive impairment. Again, the complexity of the model may be constrained by computer runtime.

Discussion

- We have fitted a multi-state model to data from the Newcastle centre of the MRC CFAS study, that takes into account informative loss-to-followup and interval censoring.
- Not all parameters in our model are identifiable, and it was therefore necessary to introduce some constraints. The parameter k in our model must be fixed, and it is therefore necessary to test the sensitivity of results to various values of k.
- Here we used sex and age as explanatory variables, although it would be easy to add more to our model. However, the number of explanatory variables might be constrained by the runtime of the program that optimizes the loglikelihood.
- In the future we may look at more complicated models, e.g. investigating the effect of having a stroke on the risk of cognitive impairment. Again, the complexity of the model may be constrained by computer runtime.

Thank you for your attention.

The Newcastle Centre

We have fitted our model to data from the Newcastle centre, with 2452 participants. We use age and sex as covariates.

Number of males = 908 , Number of females = 1544

Mean age at prevalence screen = 75.1

Age range at prevalence screen = 64 - 103

Mean interview times:

S0	A0	F1	S2	C2	A2	F3	C6	C8	CX
0	0.1	1.1	2.0	2.1	2.3	3.2	5.2	7.0	10.1

Counts of observed events:

	CI at $t = 0$	$H \longrightarrow CI$ events	$H \longrightarrow F$ events	$CI \longrightarrow F$ events	$H \longrightarrow LTF$ events	$CI \longrightarrow LTF$ events	$LTF \longrightarrow F$ events
Male	75	50	373	64	225	46	161
Female	213	134	421	182	499	130	335
Total	288	184	794	246	724	176	496