Cognitive impaired life expectancy: Multistate models including misclassification

Ardo van den Hout[1], Fiona Matthews[1] & Carol Jagger[2]

[1] *MRC Biostatistics Unit, Institute of Public Health, Cambridge, U.K.*

[2] *Department of Health Sciences, University of Leicester, U.K.*

Plan:

1. Data & research question

2. Multistate models & misclassification

3. Estimating life expectancies

4. Implementation & results

5. Conclusion & future work

1. Data & research questions

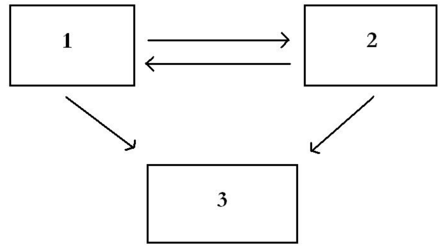
- Data from the Medical Research Council Cognitive Function and Ageing Study (MRC CFAS) - a longitudinal cohort study of 13,004 individuals aged 65 and above who have been followed over a 10 year period.

- Cognitive impairment measured using Mini-Mental State Examination with states $1 \equiv$ *not impaired* and $2 \equiv$ *impaired*. In addition, $3 \equiv$ *death*.

- Interested in: healthy life expectancy $\equiv$ expected remaining lifetime spent free of cognitive impairment.

2. Multistate models & misclassification

- Multistate models describe transitions between the states over time.



- Given a set of possible states, misclassification means that the observed state is not the latent (true) state.

- Misclassification probabilities

$$\mathbb{P}(\text{observed state} = s | \text{latent state} = r)$$

have to be estimated.

- Markov assumption: transition to the next state only depends on current state.
  - Discrete-time Markov model, see, e.g., Lièvre et al. (2003) and the software IMaCH.
  - Continuous-time Markov model, see, e.g., Jackson et al. (2003) and the R package `msm`.

- Markov models that allow for misclassification of the states are called Hidden Markov models (Jackson et al., 2003, and Bureau et al., 2003).

- Data format

ptnum	*sex*	*age*	*time*	*state*	*educ*
34	1	6	0	1	1
34	1	8	27	1	1
34	1	20	167	-2	1
35	1	11	0	2	1
35	1	13	29	3	1

age is in years minus 60,
state = -2 is censored (not dead),
time is in months,
educ = 0, 1, 2 is years of education (< 9, 9, > 9 respectively).

- Interpretation of a Markov model via transition probabilities such as

$$\mathbb{P}\Big(S_{t_2} = s \Big| S_{t_1} = r, \boldsymbol{z}(t_1)\Big),$$

  which is the probability of moving from state r to state s in the time interval $(t_1, t_2]$ given covariates $\boldsymbol{z}(t_1)$.

- Fitting a continuous-time Markov model:
  - Via transition intensities q_{sr}, i.e., instantaneous hazards of progression to state s given current state r.
  - Covariates are related to the intensities by $q_{sr}(t, \boldsymbol{z}(t)) = \exp\Big(\boldsymbol{\beta}_{rs}^T \boldsymbol{z}(t)\Big)$.

- <u>Note</u>: Age as covariate is time dependent.

3. Estimating life expectancies

- Life expectancy in state s, for an individual who begins in state r aged x is given by

$$e_{rs}(x) = \mathbb{E}\left[\int_0^\infty \mathbf{1}_{\{S_t=s\}} dt \middle| X_0 = r, x\right] = \int_0^\infty \mathbb{P}(S_t = s|S_0 = r, x) dt.$$

- Life expectancy in state s irrespectively of the initial state is given by

$$e_{.s}(x) = \sum_r \mathbb{P}(X_0 = r|x) e_{rs}(x),$$

where the r summation is over the not-dead states.

4. Implementation in the Cognitive Function and Ageing Study

1. Estimate the (hidden) Markov model using the `msm` package by Jackson.

2. Approximate $\mathbb{P}(S_t = s | S_0 = r, \boldsymbol{z})$ by piece-wise constant transition probabilities since age is a time-dependent covariate.

3. Approximate the integral in $e_{rs}(x)$ by the trapezoidal rule.

4. Estimate the initial distribution in $e_{.s}(x)$ by logistic regression.

Model with misclassification

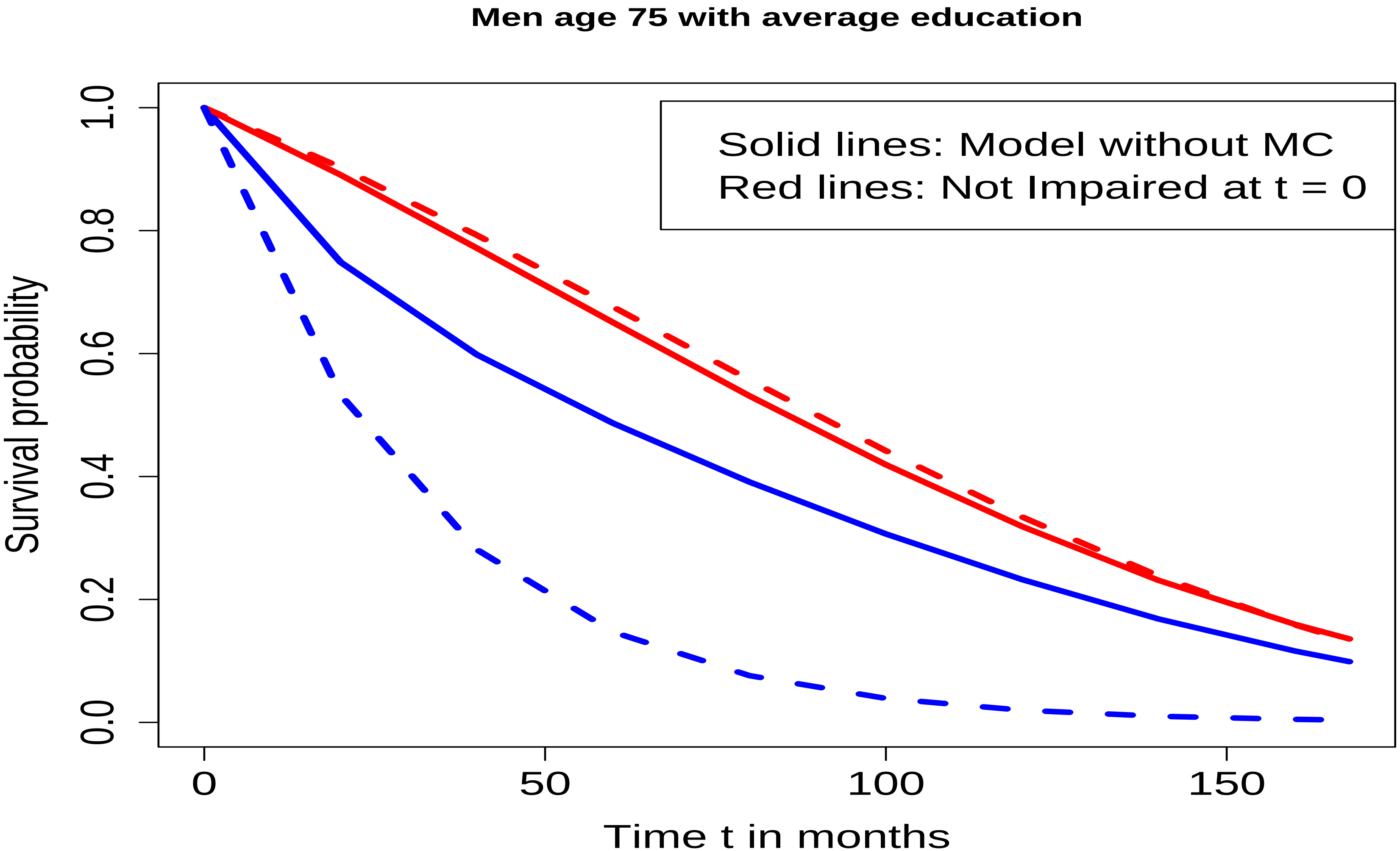
- Transition probabilities $p_{rs} = \mathbb{P}(S_t = s | S_0 = r)$ for men going from age 75 to age 85, average education:

$$\widehat{p}_{11} = 0.230 \qquad \widehat{p}_{12} = 0.102 \qquad \widehat{p}_{13} = 0.669$$
$$\widehat{p}_{21} < 0.001 \qquad \widehat{p}_{22} = 0.019 \qquad \widehat{p}_{23} = 0.980$$

- Misclassification probabilities $c_{rs} = \mathbb{P}(O = s | S = r)$ and 95% confidence intervals:

$$\widehat{c}_{11} = 0.892 \ \ (0.888, 0.895) \qquad \widehat{c}_{12} = 0.108 \ \ (0.104, 0.113)$$
$$\widehat{c}_{21} = 0.078 \ \ (0.061, 0.099) \qquad \widehat{c}_{22} = 0.922 \ \ (0.903, 0.938)$$

Survival curves for models without and with misclassification



Life expectancies (age = 65 years, education = average)

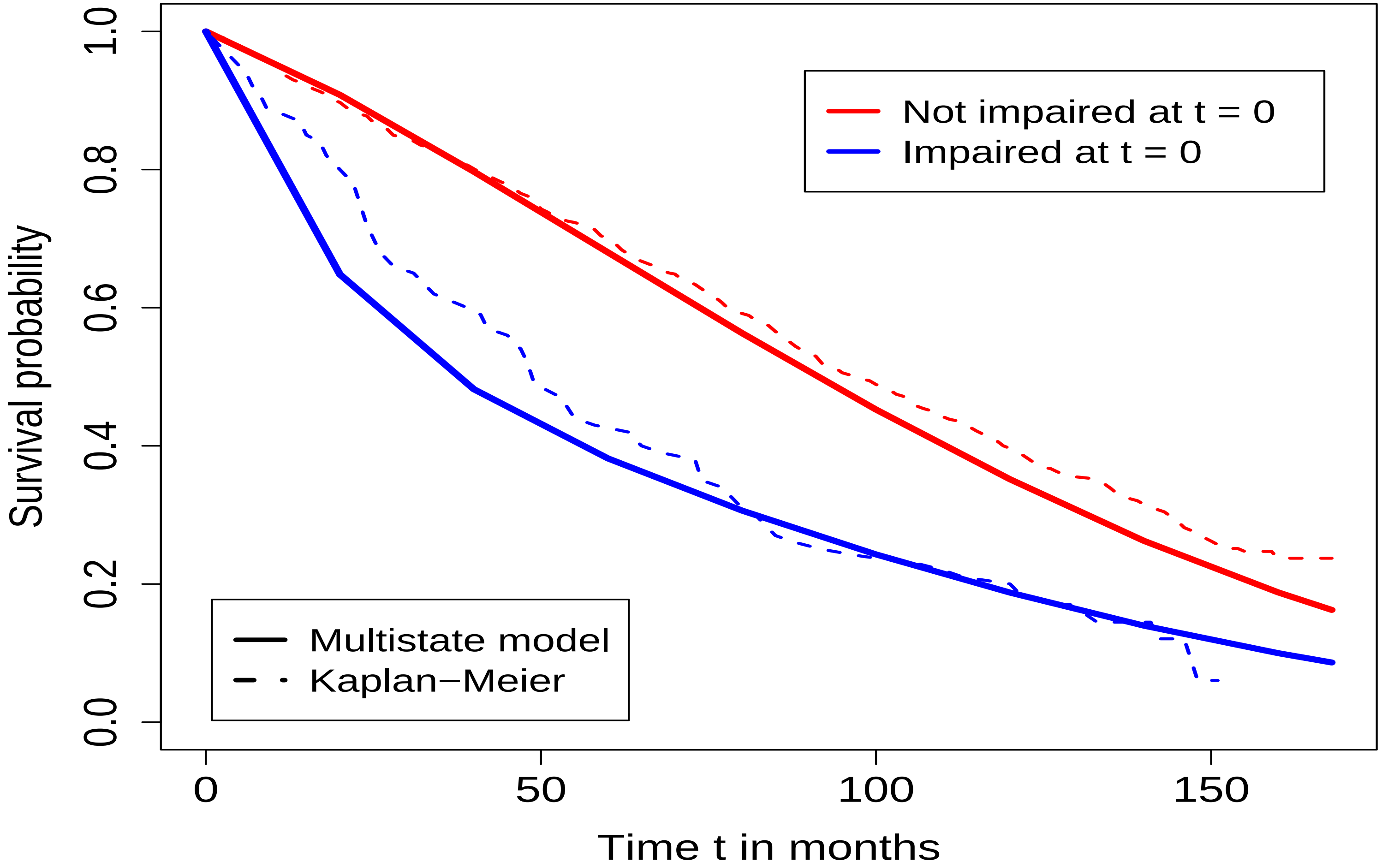
	Men		**Women**	
	Not impaired	*Impaired*	*Not impaired*	*Impaired*
Model without misclassification	**12.1** (11.1; 12.8)	**1.2** (0.8; 1.6)	**13.5** (13.0; 14.3)	**2.7** (2.2; 3.2)
Model with misclassification	**11.4** (11.2; 12.4)	**1.1** (0.7; 1.4)	**13.4** (12.8; 14.2)	**1.8** (1.4; 2.2)

(95%-confidence intervals by the bootstrap percentile method, $B = 100$*.)*

Goodness of fit

Comparing Kaplan-Meier survival curves with model-based curves.

Multistate model without misclassification (men aged 75, averaged education).



5. Conclusion & future work

- Package `msm` is user-friendly and can accommodate both censoring and misclassification.

- Computation of life expectancies is based on the multistate model.

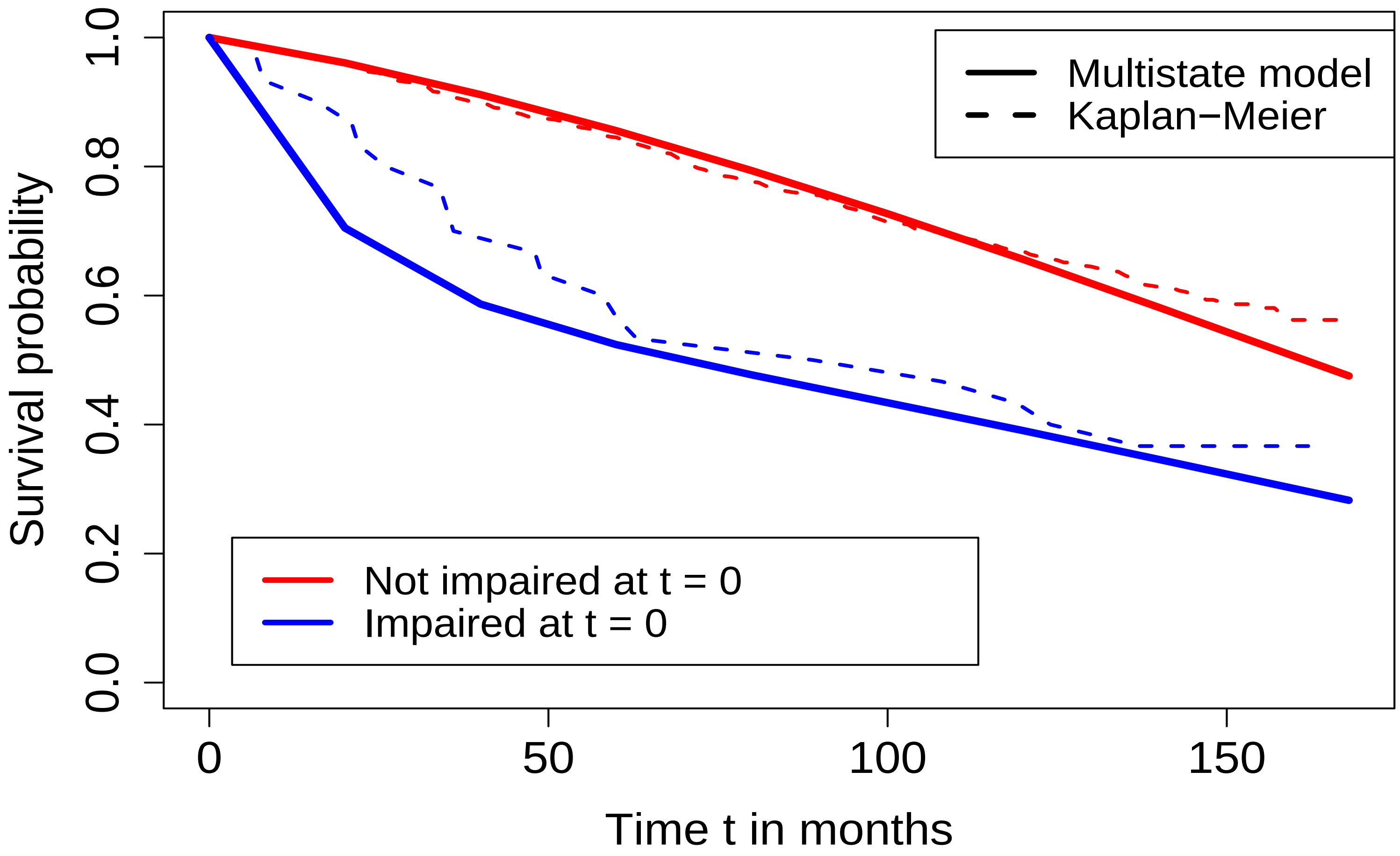
- Current & future research:
  - Semi-Markov models
  - Pearson-type goodness-of-fit test statistic (Aquirre-H. et al., 2002)
  - Ad hoc approaches to goodness of fit (Bureau et al., 2003).

(a.vandenhout@mrc-bsu.cam.ac.uk)

Goodness of fit

Comparing Kaplan-Meier survival curves with model-based curves.

Multistate model without misclassification (men aged 65, averaged education).



Results:

Survival curves for the model without misclassification