Joint models for discrete longitudinal outcomes in ageing research

Ardo van den Hout

Department of Statistical Science, UCL

LSE, December 2013

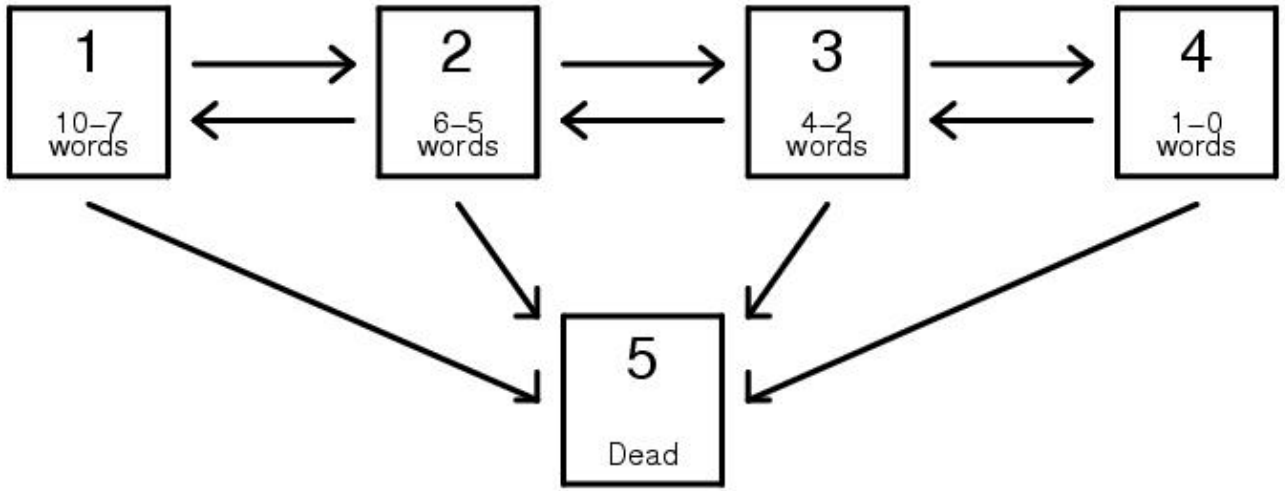
Joint models for discrete longitudinal outcomes in ageing research

1. Research interest & data
2. Multi-state survival models
3. Joint models using a shared-parameter approach
4. Conclusion

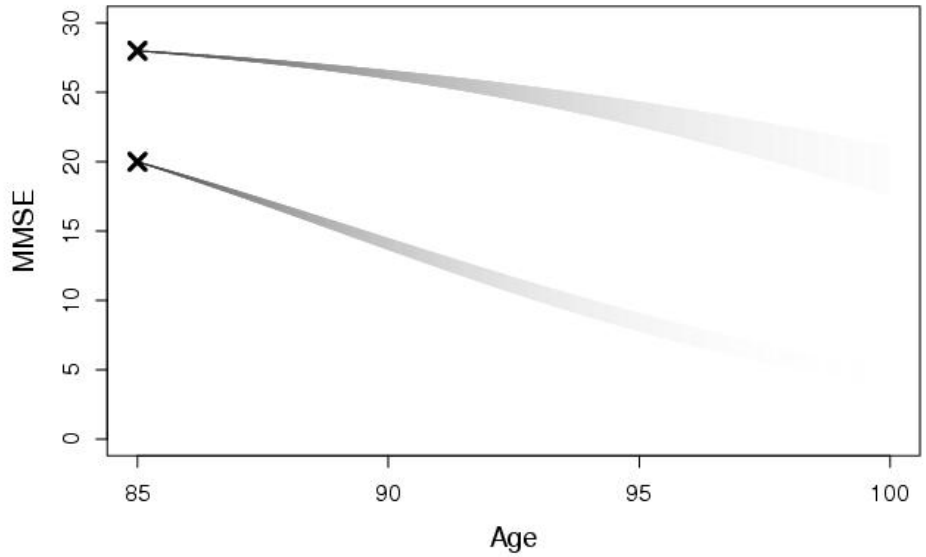
1. Research interests & data

- UK population is ageing
- Statistical modelling of the ageing process is of interest: understanding and prediction
- Wider scope of modelling: planning future health care
- Today: cognitive function. Response often discrete

- Analysis of longitudinal data for ageing processes cannot ignore dropout due to death
- One option: use a multi-state survival model
- Example for number of words remembered:



- Another option: link a model for the process of interest with a model for survival
- Example: cognitive function using the Mini-Mental State Examination (MMSE). Ceiling effects very common



MMSE prediction

Fade out for survival probability

Random effects link measurement model and survival

Longitudinal data

- MRC Cognitive Function and Ageing Study (CFAS, `www.cfas.ac.uk`)
- Cambridge City over-75s Cohort Study (CC75C, `www.cc75c.group.cam.ac.uk`)
- English Longitudinal Study of Ageing (ELSA, `www.ifs.org.uk/ELSA`)

2. Multi-state survival models

- Using theory for continuous-time stochastic process
- Regression models for transition intensities (hazards)
- Approximate time-dependency by piecewise-constant intensities
- 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)

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

$$h_{rs}(t_i) = h_{rs.0}(t_i) \exp(\boldsymbol{x}_i^\top \boldsymbol{\beta}_{rs})$$

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

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

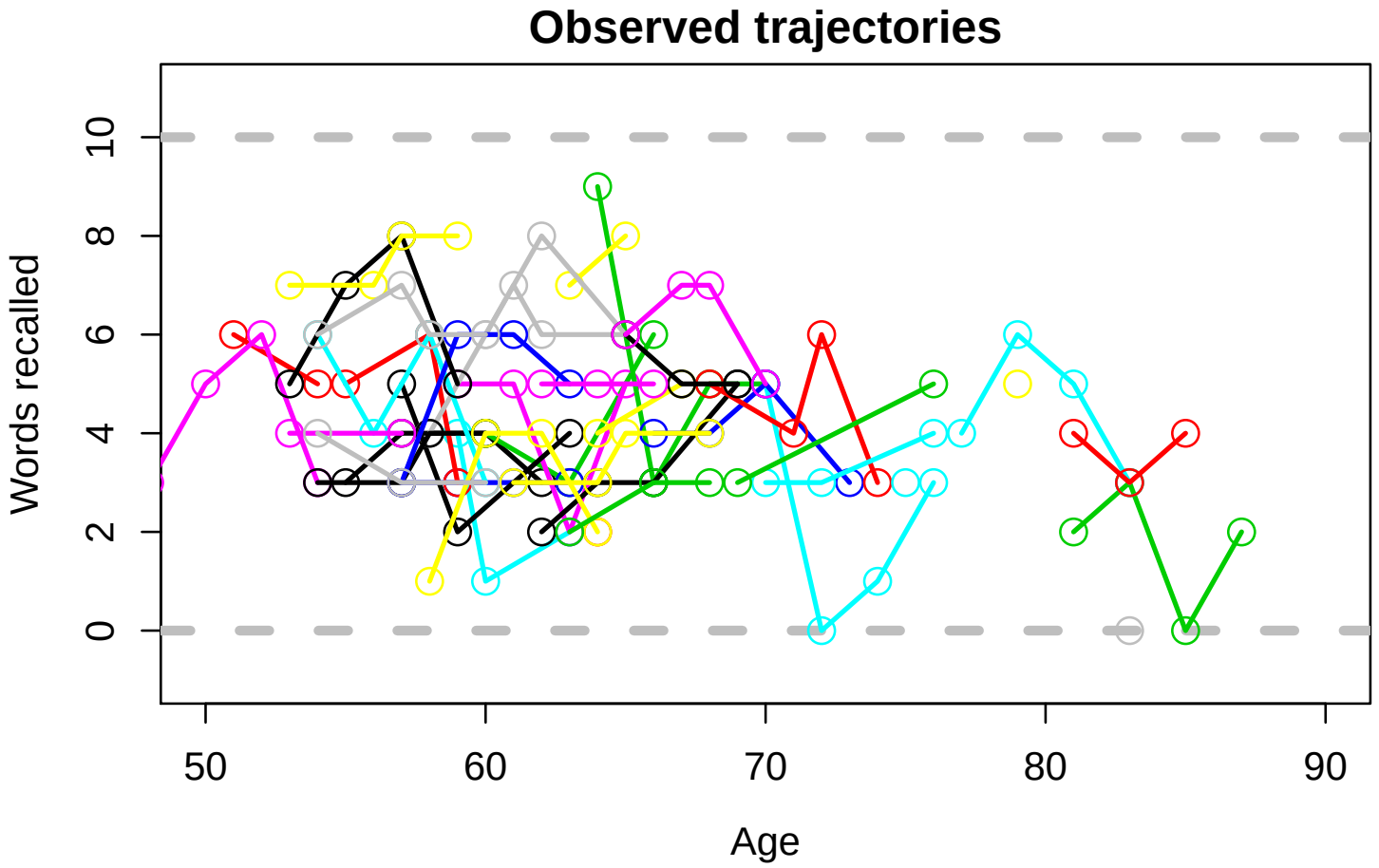
  - Choose parametric shape because of prediction

- Approximate shape by piecewise-constant intensities
  - Example: for observation times $t_{i1}, t_{i2}, t_{i3}, t_{i4}$ hazards are held constant within $(t_{i1}, t_{i2}]$, $(t_{i2}, t_{i3}]$, and $(t_{i3}, t_{i4}]$
  - When time between end of follow-up and death is long, impose grid (Van den Hout & Matthews 2009)
- For maximum likelihood, transition probabilities derived from transition intensities using the Kolmogorov differential equations (Cox & Miller 1965)
- Use scoring algorithm (extension of Kalbfleisch & Lawless 1985)

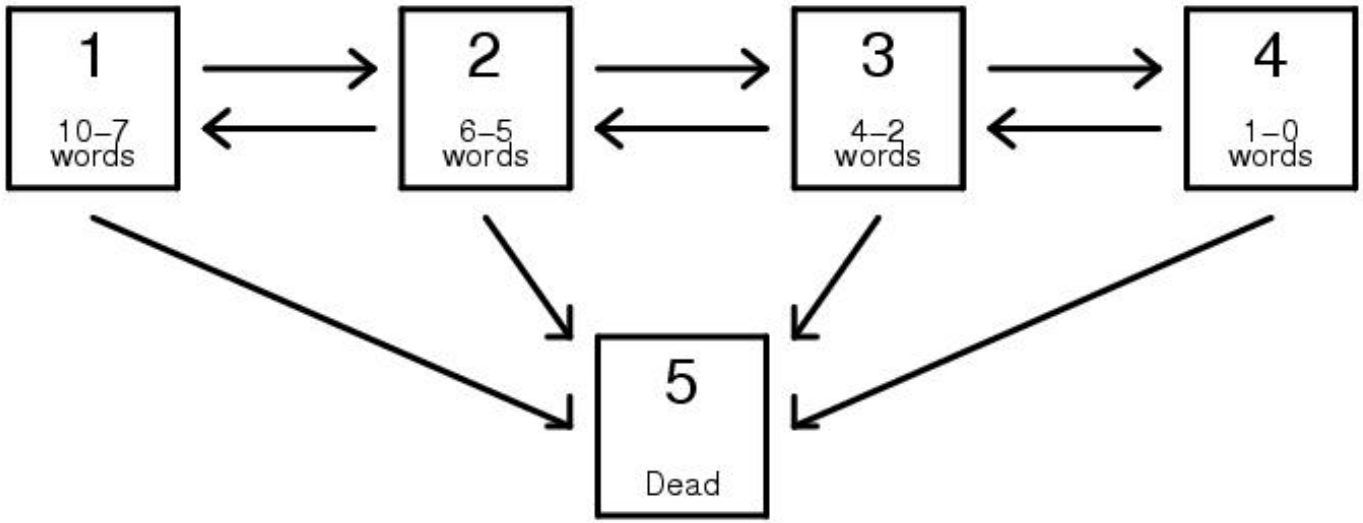
Application

- English Longitudinal Study of Ageing (ELSA)
- Info on health, economic position, and quality of life. Data in waves 1 - 4 (2002-2009). Age 90^+ is censored
- Subsample $N = 1000$, with 6 individuals with censored age, and 157 with only one observation
- Sample size in application 837. With 463 women and 374 men. Dropout due to death is about 11%

- Response is the number of words remembered in a recall from a list of ten



- Living states defined using number of words remembered in the recall



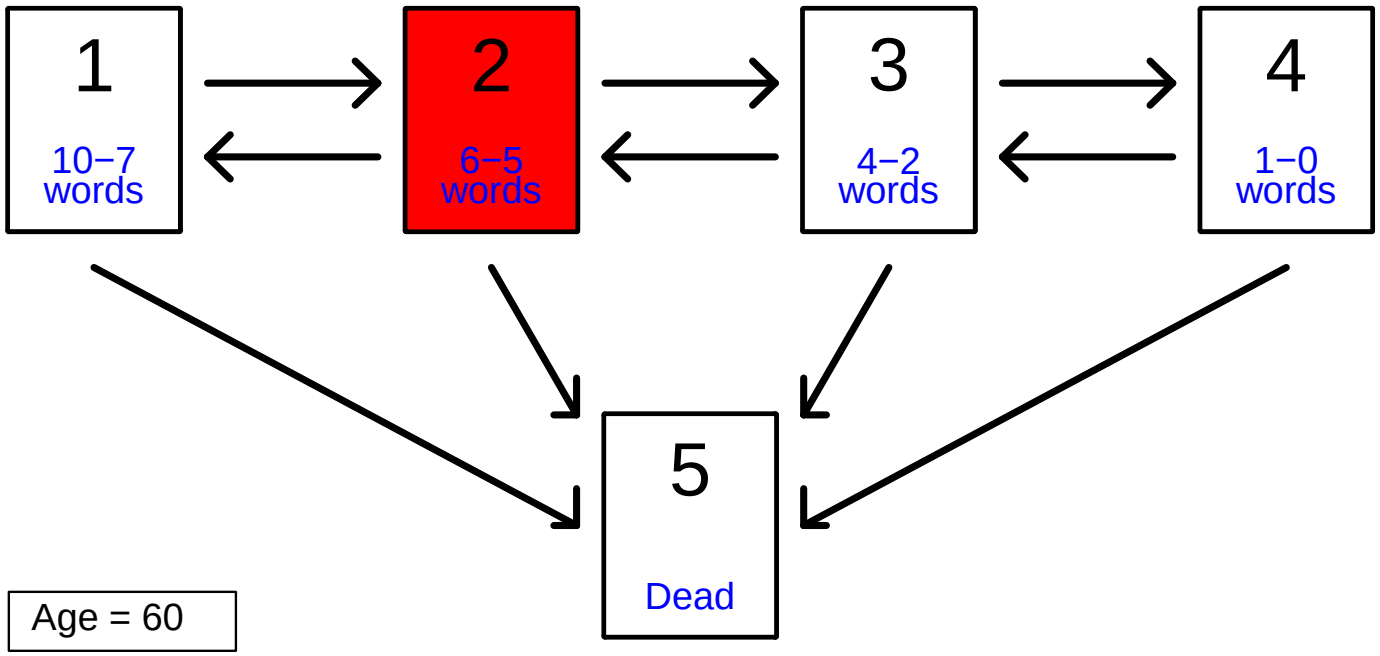
- AIC prefers Gompertz hazards for the forward transitions between living states and Weibull for death:

$$q_{rs}(t) = \exp\left(\beta_{rs.0} + \xi_{rs}t + \beta_{rs.1}Sex + \beta_{rs.2}Education\right) \quad \text{for } (r,s) \in \{(1,2),(2,3),(3,4)\}$$

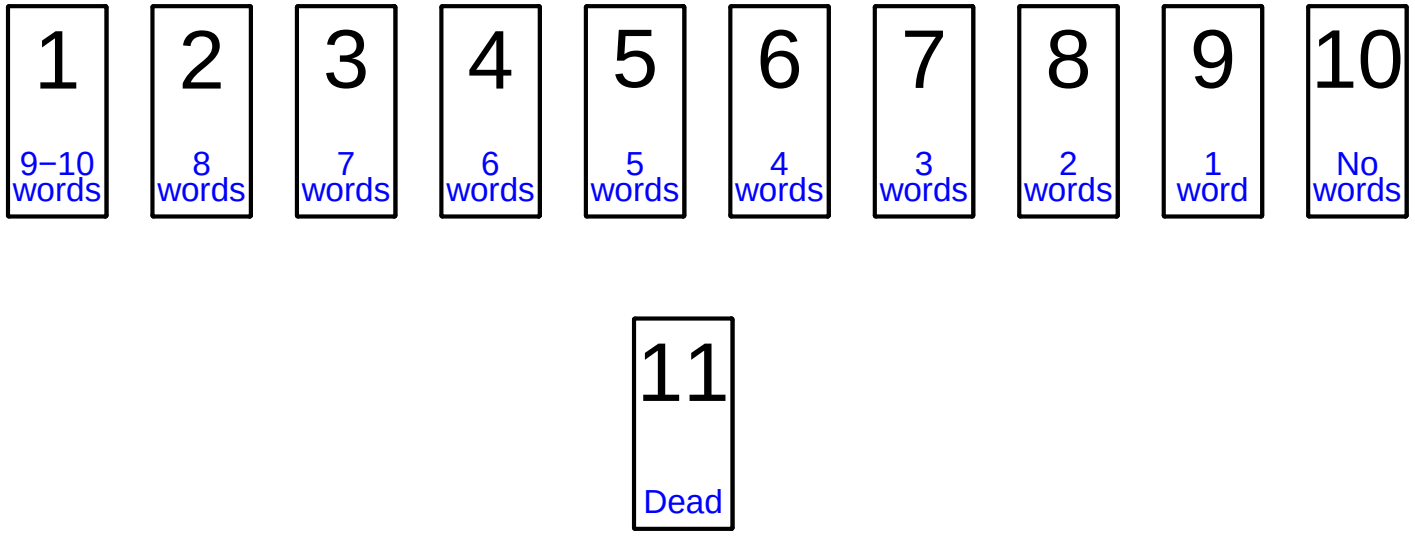
$$q_{rs}(t) = \exp\left(\beta_{rs.0}\right) \quad \text{for } (r,s) \in \{(2,1),(3,2),(4,3)\}$$

$$q_{rD}(t) = \tau_D t^{(\tau_D - 1)} \exp\left(\beta_{rD.0} + \beta_D Sex\right), \quad \text{for } r \in \{1,2,3,4\}$$

- Ageing associated with decline of cognitive function
- More education associated with a lower risk of moving to states 2 or 3
- Parametric model:
  - Can estimate transition probabilities for any start state and specified time interval
  - Can estimate duration of stay per state (residual life expectancies)

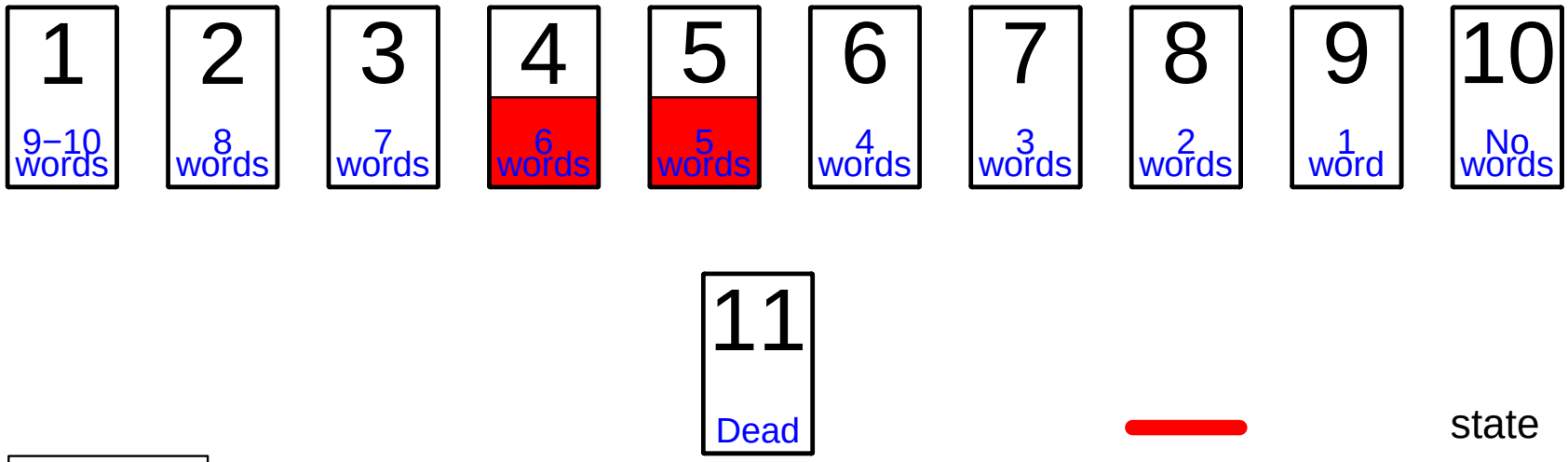

1
10−7
words
2
6−5
words
3
4−2
words
4
1−0
words
5
Dead
Age = 60

- Extend number of states:



- Example of a model:

$$
\begin{aligned}
q_{rs}(t) &= \exp(\alpha_0 + \xi t + \alpha_1 r) && \text{for } (r, s = r + 1) \\
q_{rs}(t) &= \exp(\beta) && \text{for } (r, s = r - 1) \\
q_{rD}(t) &= \tau_D t^{(\tau_D - 1)} \exp(\gamma)
\end{aligned}
$$


1
9–10 words
2
8 words
3
7 words
4
6 words
5
5 words
6
4 words
7
3 words
8
2 words
9
1 word
10
No words
11
Dead
Age = 60
state
last living state

3. Joint models using a shared-parameter approach

With Graciela Muniz

MRC Unit for Lifelong Health and Ageing, London

- Joint models used for continuous longitudinal measurements in the presence of dropout during follow-up
- Today's joint models builds upon an established framework (Rizopoulos 2012; Fitzmaurice et al. 2009, Ch. 13-16)
- Extension to deal with specific features of modelling of cognitive function in the older population

- Typically, cognitive function is measured with a questionnaire test with an integer scale

  In the applications: MMSE sum score $\in \{0, 1, 2, ..., 30\}$, and recall of words $\in \{0, 1, ..., 10\}$

- Specific features of our joint models:

  i) Non-linear mixed-effects model instead of linear

  ii) Parametric survival model instead of semi-parametric

  iii) Age as time scale: delayed entry (left truncation)

- For individual i, with response y_{ij} at time t_{ij} the **measurement model** has linear predictor

$$\begin{aligned} \eta_{ij} &= \beta_{0i} + \beta_{1i} t_{ij} + \boldsymbol{X}_i \gamma \\ \beta_{0i} &= \beta_0 + b_{0i} \\ \beta_{1i} &= \beta_1 + b_{1i} \qquad \boldsymbol{b}_i = (b_{0i}, b_{1i}) \sim N(\mathbf{0}, \boldsymbol{\Sigma}) \end{aligned}$$

- Discrete outcome: The logit link for binomial distribution or for beta-binomial distribution

- Beta-binomial has an extra variance parameter θ — here assumed to be the same for all individuals

- Beta-binomial distribution for Y defined by probability of success μ, variance parameter θ, and m trials:

$$\mathbb{E}[Y] = m\mu \quad \text{and } \mathrm{Var}[Y] = m\mu(1-\mu)\left(1 + \frac{(m-1)\theta}{(1+\theta)}\right)$$

- Observation-specific random effect in a binomial regression model: don't switch to beta-binomial

- Cluster-specific random effects: switching may work

- The binomial regression model is generalised linear mixed model, the beta-binomial is not (Agresti 2002)

- Given $\boldsymbol{\beta_i} = (\beta_0 + b_{0i}, \beta_1 + b_{1i})$, the **hazard model** is

$$h_i(t) = h_0\left\{t | g_0(\alpha_0, \boldsymbol{\beta}_i, t)\right\} \exp\{g(\alpha, \boldsymbol{\beta}_i, t) + \boldsymbol{x}_i^* \gamma^*\}$$

- Examples of parametric shapes for the baseline are

$$\begin{aligned} \text{Weibull:} \quad h_0(t) &= \lambda \tau t^{\tau - 1} \\ \text{Gompertz:} \quad h_0(t) &= \lambda \exp(\xi t) \end{aligned}$$

- g_0 and g are is functions of fixed and random effects

- Random effects $\boldsymbol{b}_i$ are *shared* by measurement model and hazard model

- The **hazard model** (again) is

$$h_i(t) = h_0\{t|g_0(\alpha_0, \boldsymbol{\beta}_i, t)\}\exp\{g(\alpha, \boldsymbol{\beta}_i, t) + \boldsymbol{x}_i^*\boldsymbol{\gamma}^*\}$$

- Using $g(\alpha, \boldsymbol{\beta}_i, t) = \alpha(\beta_{0i} + \beta_{1i}t)$

  – and the Gompertz defines a Gompertz model:

$$\begin{aligned} h_i(t) &= \exp\{\gamma_0^* + \xi t + \alpha(\beta_{0i} + \beta_{1i}t)\} \\ &= \exp\{(\gamma_0^* + \alpha\beta_{0i}) + (\xi + \alpha\beta_{1i})t\} \end{aligned}$$

  – and the exponential defines a Gompertz model:

$$h_i(t) = \exp\{(\gamma_0^* + \alpha\beta_{0i}) + \alpha\beta_{1i}t\}$$

- Marginal likelihood assuming independence given the random effects. For individual i with age at baseline t_{i1} and death or right censoring at t_i

$$L_i(\boldsymbol{\omega}|t_{i1}) = \int p(t_i|\boldsymbol{b}_i, t_{i1}, \delta_i, \boldsymbol{\omega}) p(\boldsymbol{y}_i|\boldsymbol{b}_i, \boldsymbol{\omega}) p(\boldsymbol{b}_i|t_{i1}, \boldsymbol{\omega}) d\boldsymbol{b}_i$$

$$= \int h(t_i|\boldsymbol{b}_i, \boldsymbol{\omega})^{\delta_i} P(T \geq t_i|T > t_{i1}, \boldsymbol{b}_i, \boldsymbol{\omega}) \left\{ \prod_j p(y_{ij}|\boldsymbol{b}_i, \boldsymbol{\omega}) \right\} p(\boldsymbol{b}_i|t_{i1}, \boldsymbol{\omega}) d\boldsymbol{b}_i$$

$$= \int h(t_i|\boldsymbol{b}_i, \boldsymbol{\omega})^{\delta_i} P(T \geq t_i|\boldsymbol{b}_i, \boldsymbol{\omega}) \left\{ \prod_j p(y_{ij}|\boldsymbol{b}_i, \boldsymbol{\omega}) \right\} \frac{p(\boldsymbol{b}_i|\boldsymbol{\omega})}{P(T \geq t_{i1}|\boldsymbol{\omega})} d\boldsymbol{b}_i$$

where $\boldsymbol{\omega}$ is vector with population parameters

(*cf.* Jensen *et al.* 2004; Rondeau *at al.* 2006)

- For Gompertz with λ_i and ξ_i

$$P(T \geq t_i | \boldsymbol{b}_i, \boldsymbol{\omega}) = \exp\Big[- \lambda_i \xi_i^{-1} \{\exp(\xi_i t_i) - 1\}\Big]$$

- Maximisation in R using Gaussian quadrature and a multi-purpose optimiser

- For estimation of random effects: maximum aposterior (MAP) estimation

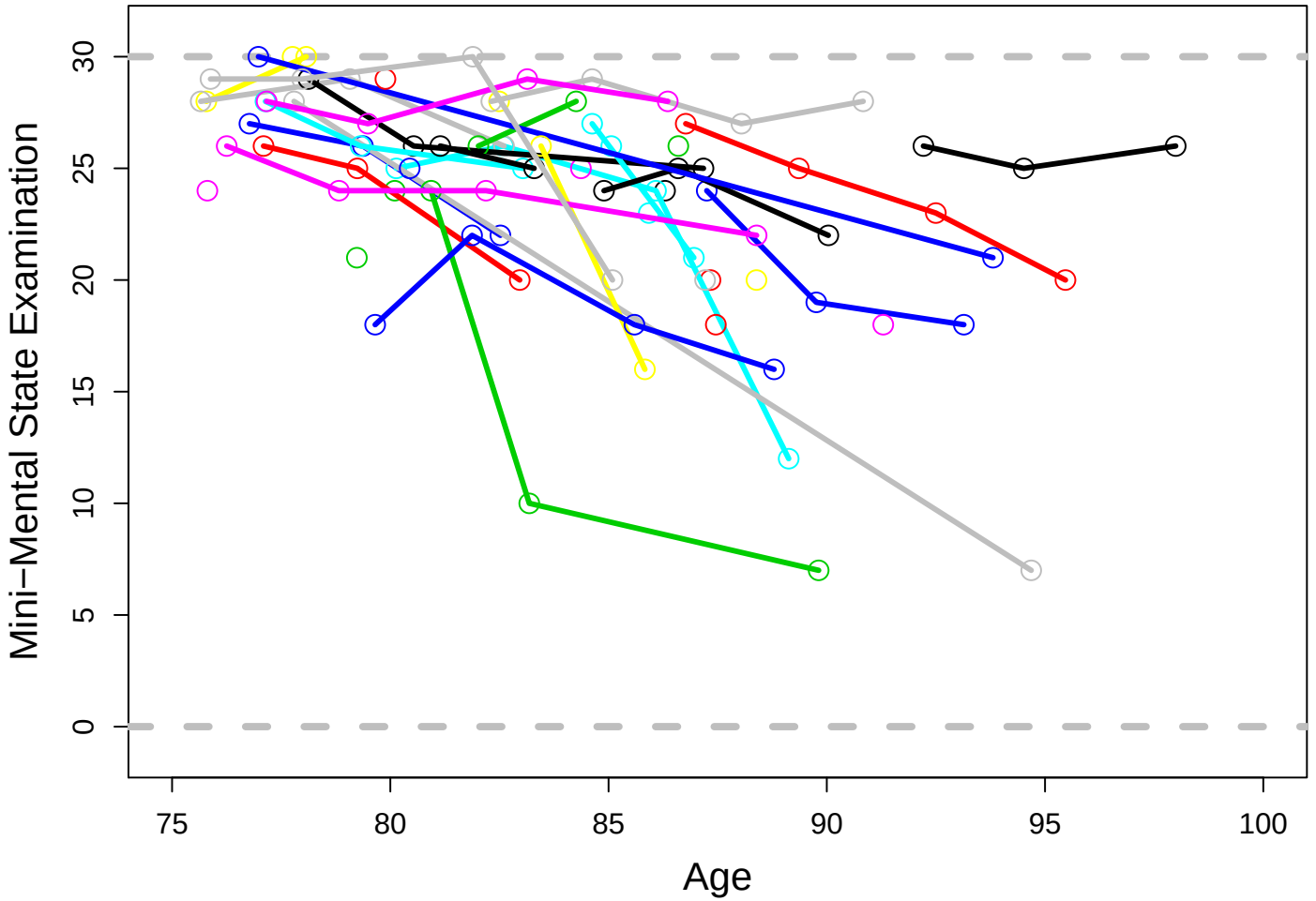
- Alternative: Bayesian inference using a Gibbs sampler, with Metropolis steps to draw values from the conditional distributions. (Can't do this in WinBUGS)

Two applications

- CC75C with $N = 1932$. All death times available. Of interest: how change of cognitive function over time affects survival

- ELSA. Dropout due to death is 11%. Of interest: effect of gender on cognitive change over time when controlling for education

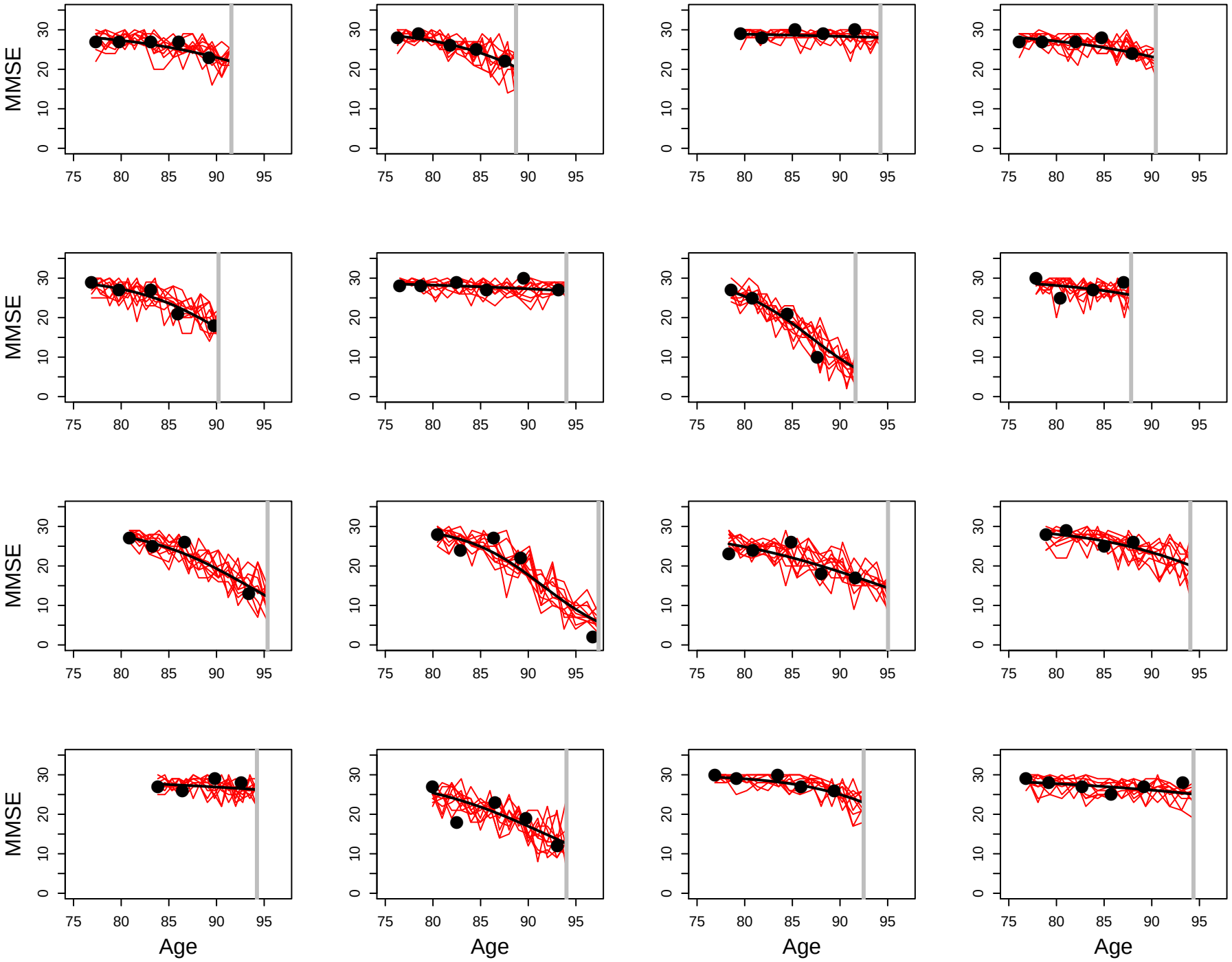
CC75C


Mini−Mental State Examination
0
5
10
15
20
25
30
75
80
85
90
95
100
Age

- 698 men and 1234 women. Mean age at baseline is 81
- Covariates `sex` (0/1 for women/men) and `educ` (0/1 for less than 10yrs of education/10yrs or more)
- AIC prefers beta-binomial regression and Gompertz hazard

$$\begin{aligned} \eta_{ij} &= \beta_{0i} + \beta_{1i} t_{ij} \\ h_i(t) &= \exp\{\gamma_0^* + \alpha\beta_{0i} + (\xi + \alpha\beta_{1i})t + \gamma_1^* \texttt{sex}_i + \gamma_2^* \texttt{educ}_i\} \end{aligned}$$

- Next slide: Fit to individual data for a random subset of 16 individuals (including simulated trajectories)


MMSE
0
10
20
30
75
80
85
90
95
Age

- Model (again) with $\beta_{0i} = \beta_0 + b_{0i}$ and $\beta_{1i} = \beta_1 + b_{1i}$

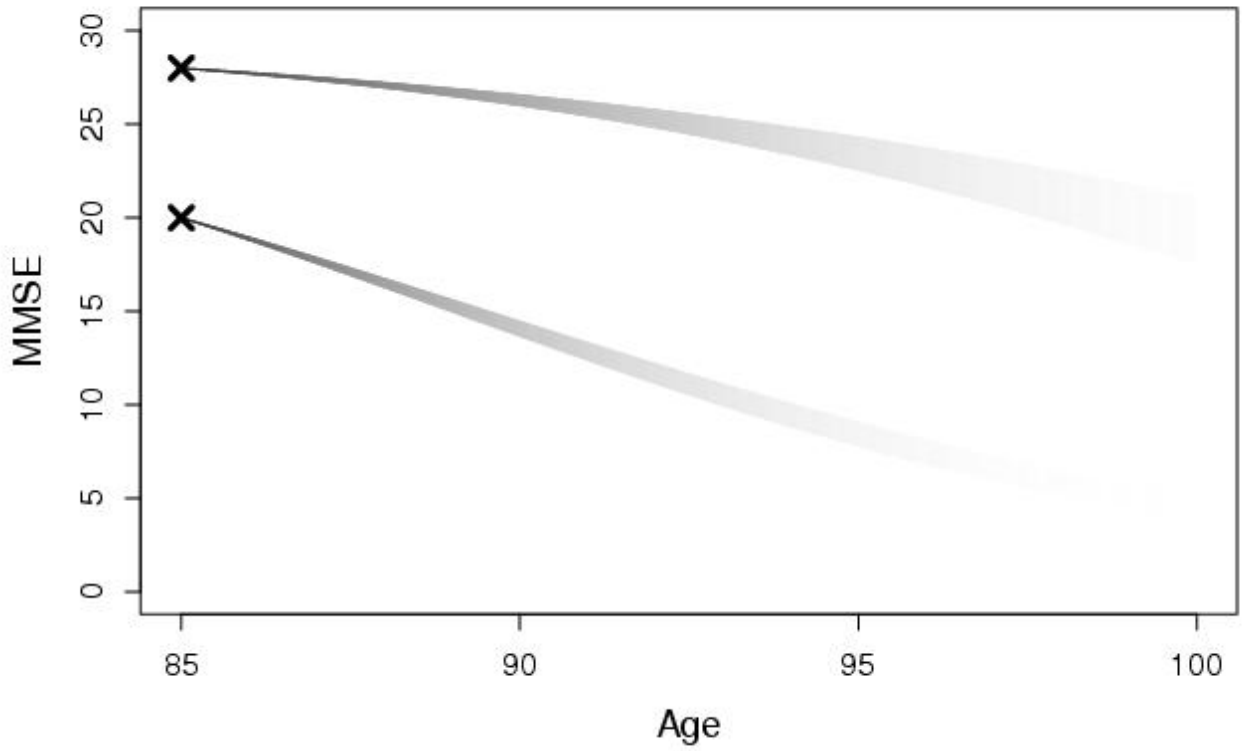
$$\begin{aligned} \eta_{ij} &= \beta_{0i} + \beta_{1i} t_{ij} \\ h_i(t) &= \exp\{\gamma_0^* + \alpha\beta_{0i} + (\xi + \alpha\beta_{1i})t + \gamma_1^* \texttt{sex}_i + \gamma_2^* \texttt{educ}_i\} \end{aligned}$$

- Time scale t is age minus 70 years

- Some results: $\widehat{\beta}_1 = -0.16$ $(-0.17; -0.15)$, $\widehat{\gamma}_1^*, \widehat{\gamma}_2^* > 0$, $\widehat{\alpha} = -0.42$ $(-0.48; -0.35)$

- Negative α implies that having better cognitive function as measured by the MMSE is associated with a smaller hazard and thus with better survival

- Prediction given MMSE at a specified age is possible
- Example



ELSA

- Data in waves 1 - 4 (2002-2009)
- The 167 individuals with censored age (90^{+}) are removed. Resulting sample size is 11627
- Dropout due to death is around 11%
- Baseline: 6533 women and 5094 men
- Response $\in \{0, 1, ..., 10\}$ is number of words remembered in a recall from a list of ten

- Binomial regression and Gompertz hazard

$$\begin{aligned}\eta_{ij} &= \beta_{0i} + \beta_{1i}t_{ij} + \gamma_1 \texttt{sex}_i + \gamma_2 \texttt{educ}_i + \gamma_3(\texttt{sex}_i \times t_{ij}) + \gamma_4 \texttt{yob}_i \\ h_i(t) &= \exp\{\gamma_0^* + \alpha\beta_{0i} + (\xi + \alpha\beta_{1i})t + \gamma_1^* \texttt{sex}_i + \gamma_2^* \texttt{bmi}_i\},\end{aligned}$$

  where men $\equiv$ 1, more education $\equiv$ 1. Year of birth minus 1900, BMI coded -2, -1, 0, 1, 2, 3

- Time scale t is age minus 30 years

- Model has AIC = 148333. Switching to beta-binomial regression: AIC = 148233

- Model (again), with men $\equiv 1$, more education $\equiv 1$:

$$\begin{aligned}\eta_{ij} &= \beta_{0i} + \beta_{1i}t_{ij} + \gamma_1 \texttt{sex}_i + \gamma_2 \texttt{educ}_i + \gamma_3(\texttt{sex}_i \times t_{ij}) + \gamma_4 \texttt{yob}_i \\ h_i(t) &= \exp\{\gamma_0^* + \alpha\beta_{0i} + (\xi + \alpha\beta_{1i})t + \gamma_1^* \texttt{sex}_i + \gamma_2^* \texttt{bmi}_i\}\end{aligned}$$

$$\begin{aligned}\widehat{\beta}_1 &= -0.002\ (0.002) & \widehat{\gamma}_1 &= -0.067\ (0.051) \\ \widehat{\gamma}_3 &= -0.004\ (0.002) & \widehat{\gamma}_1^* &= \ \ 0.539\ (0.062) \\ & & \widehat{\alpha} &= -0.161\ (0.043)\end{aligned}$$

- Fit based on MAP OK. Extrapolation not so good

- Age range is quite wide relative to length of follow-up

4. Conclusion

- Using sum score: Varying success probabilities across questions is OK for the binomial. Dependence is not
- Possible alternative: Item Response Theory models (with J-P Fox)
- Alternatives for discrete sum score and ceiling effects:
  - Proust-Lima et al (2009) use transformation
  - Hutton & Stanghellini (2009) use censored skew-normal distributions

- For longitudinal discrete outcome data: multi-state survival models and joint models can take dropout due to death into account

- Distribution of test sum scores is often skewed: use of the normal distribution for the outcome is problematic

- Presented framework is flexible w.r.t. other distributions or random-effects structures