Supplementary material to

Estimating dementia-free life expectancy for Parkinson's patients using Bayesian inference and micro-simulation

Ardo van den Hout, Fiona E. Matthews

MRC Biostatistics Unit, Institute of Public Health
Robinson Way, Cambridge CB2 OSR, UK

Introduction

This is supplementary material to Section 5 that describes the application. Section A discusses the model comparison that led to the choice of Model 1 and Model 2. Section B describes the sensitivity analyses. In Section C, we present the ANOVA that was used to check the micro-simulation for Model 1. Section D describes the details of the Bayesian χ^2 test. Even though this test cannot be used for the data at hand, it might still be of interest to other applications.

A Model selection

For t the number of years since baseline, we consider specifications of a model that describes the baseline state and the transition intensities. For individual i, the model is given by

$$\begin{aligned}
\text{logit}(\theta^i) &= \alpha_0 + \alpha_a \text{Age}^i(0) + \alpha_s \text{Sex}^i + \alpha_d \text{Duration}^i \\
\log[q^i_{rs}(t)] &= \beta_{0.rs} + \beta_{a.rs}\text{Age}^i(t) + \beta_{s.rs}\text{Sex}^i + \beta_{d.rs}\text{Duration}^i + \tau^i_{rs}(t) \\
\tau^i_{rs}(t) &= \gamma^i_{0.rs} + \gamma^i_{a.rs} Age^i(t)
\end{aligned}$$

$$\boldsymbol{\gamma}^i = (\gamma^i_{0.12}, \gamma^i_{0.13}, \gamma^i_{0.23}, \gamma^i_{a.12}, \gamma^i_{a.13}, \gamma^i_{a.23})^\top \sim MVN(\mathbf{0}, \boldsymbol{\Sigma}).$$

Table 1 presents the models that are considered, where Model 1 and Model 2 are the models that are discussed in the paper. The order of the models in the table is not a ranking. In order to take the time-dependency into account, age was always included as a covariate for the intensities. We started the model comparison with the fixed-effect model, i.e., Model 3. This model has DIC = 4145, but some of the parameters are weakly identified: $\beta_{0.13}$ and $\beta_{s.13}$ have large 95% credible intervals and an almost perfect posterior correlation of -0.98. This means that with the data at hand it is not possible to identify the effect of sex on the transition from state 1 to death. Restricting $\beta_{s.13}$ to zero whilst leaving the rest of the model unchanged, yields Model 4 with DIC = 4145. Possible further restrictions on covariate effects are indicated by 95%

Table 1: Models for the Norwegian study of Parkinson's patients.

	Parameter restrictions	DIC
Model 1	$\beta_{s.13} = \beta_{d.13} = \beta_{d.23} = \alpha_s = 0$ $\tau^i_{rs}(t) = 0$	4142
Model 2	$\beta_{s.13} = \beta_{d.13} = \beta_{d.23} = \alpha_s = 0$ $\gamma^i_{a.12} = \gamma^i_{a.13} = \gamma^i_{a.23} = \gamma^i_a$	4142
Model 3	$\tau^i_{rs}(t) = 0$	4145
Model 4	$\tau^i_{rs}(t) = 0$, $\beta_{s.13} = 0$	4145
Model 5	$\beta_{s.13} = \beta_{d.13} = \beta_{d.23} = \alpha_s = 0$	4141
Model 6	$\beta_{s.13} = \beta_{d.13} = \beta_{d.23} = \alpha_s = 0$ $\gamma^i_{a.12} = \gamma^i_{a.13} = \gamma^i_{a.23} = 0$	4139
Model 7	$\beta_{s.13} = \beta_{d.13} = \beta_{d.23} = \alpha_s = 0$ $\gamma^i_{a.12} = \gamma^i_{a.13} = \gamma^i_{a.23} = 0$ $\gamma^i_{0.12} = \gamma^i_{0.13} = \gamma^i_{0.23} = \gamma^i$	4144

credible intervals that include zero. Model 1 is the model with restrictions $\beta_{s.13} = \beta_{d.13} = \beta_{d.23} = \alpha_s = 0$. This model has DIC = 4142. The restrictions yield a more parsimonious model and a smaller DIC.

Next we investigate whether the addition of random effects leads to a better model. In Model 5, the restrictions on the fixed effects are maintained, but there are no restrictions on the random effects structure. This model has DIC = 4141. Looking at the posterior means of Model 5, it seems reasonable to use the restriction $\gamma^i_{a.12} = \gamma^i_{a.13} = \gamma^i_{a.23} = \gamma^i_a$, i.e., one random slope for all three transitions. This is Model 2 and it has DIC = 4142. Model 6 is the model without the random slopes and has DIC = 4139. Model 7 is a shared random effect model. This model assumes that random effect γ^i describes a general susceptibility to ill-health which affects all three possible transitions.

This model has DIC = 4144.

DICs of the models are close. Model 1 is the most parsimonious, but we think that Model 2 is interesting as it includes random effects to capture possible heterogeneity not captured by the covariates and it allows random effects to change over time by regressing the effects on time-dependent age. Assuming that random effects do not change over time is a strong assumption and in that sense Model 1 and Model 6 might be too restrictive.

B Sensitivity Analysis

For the fixed effects in Model 1, we specified alternative univariate normal distributions with mean zero and variance equal to either 100 or 10000. For both these choices of variance the DIC stays 4142 and the prior means of the effects are very similar to the means given the previous priors for Model 1. We can conclude that Model 1 is robust regarding specifications of vague priors.

For Model 2, we investigate alternative priors for the precision matrix $\boldsymbol{T}$. Denote the diagonal entries of $\boldsymbol{R}$ for the random intercepts r_1 and the diagonal entries for the random slopes r_2. The DICs varied slightly for different specifications: for $(r_1, r_2) = (1, 0.001)$, DIC = 4140, for $(r_1, r_2) = (2, 0.01)$, DIC = 4140, and for $(r_1, r_2) = (0.25, 0.001)$, DIC = 4143. We do not consider this variation to be significant. Again, estimation of fixed effect was robust across the specifications. The estimation of $\boldsymbol{\Sigma}_{22}$, that is the variance of the

distribution for the random intercept for the transition from state 1 to state 3, increases slightly for larger r_1. For $(r_1, r_2) = (1, 0.01)$, the posterior mean is 0.75 with 95% CI (0.349, 1.583), for $(r_1, r_2) = (2, 0.01)$, the posterior mean is 1.07 with CI (0.478, 2.665). We prefer the specification $(r_1, r_2) = (1, 0.01)$ for the prior as this reflects a sensible range for the random effects.

For Model 7, we used a half-normal distribution as the prior for the variance of the shared random effect σ. Choosing the variance of this half-normal equal to 0.5 leads to a posterior mean of σ equal to 0.16. We specified alternative priors using the half-normal with variance equal to 1 and equal to 2. Although the posterior mean of σ varied (0.20 and 0.16, respectively), the DIC stayed the same and the estimation of the fixed effects was very similar. Another alternative prior for the variance of a random effect can be formulated using the Gamma distribution. We specified a prior for the precision by using the Gamma distribution with parameters (0.01, 0.01) and (0.001, 0.001). Results for DIC and the posterior mean of σ were 4144 and 4143, and 0.19 and 0.13, respectively. Again, estimation of the fixed effects was very similar.

C ANOVA for the micro-simulation

Using the one-way analysis of variance technique proposed by O'Hagan *and others* (2007), we checked results by comparing estimated standard errors.

Within-individuals and between-individuals sums of squares are defined by

$$S_{\text{within}} = \sum_{b=1}^{B} \sum_{c=1}^{C} (L_b^c - \overline{L}_b)^2 \quad \text{and} \quad S_{\text{between}} = C \sum_{b=1}^{B} (\overline{L}_b - \overline{L})^2.$$

And the estimator of the second-order uncertainty is

$$v_A = \frac{1}{C} \left(\frac{S_{\text{between}}}{B-1} - \frac{S_{\text{within}}}{B(C-1)} \right),$$

where the definition of B, C, L_b^c, and $\overline{L}_b$ are given in Section 3. As explained in Section 3, the second-order uncertainty can also be estimated using stored sample means $\overline{L}_b^c$ and variances V_b^c. Let v_M denote this second estimator. The uncertainty is overestimated by v_M if C is not large enough. The bias arises because variability in the $\overline{L}_b$ inflates their variance, over and above the variability of the true means that is represented by the second-order uncertainty (O'Hagan *and others*, 2007). Comparison of standard errors $\sqrt{v_A}$ and $\sqrt{v_M}$ is used to assess the bias in v_M and hence the bias in the approximation of the posterior of L. For Model 1 and the life expectancies reported in the paper, the standard errors are given for both the ANOVA and the micro-simulation in Table 2. The figures in the table show that for $C = 1000$ there is indeed some consistent overestimation, but that the bias is negligible for practical purposes.

Table 2: ANOVA standard errors ($\sqrt{v_A}$) and standard errors directly from the micro-simulation ($\sqrt{v_M}$) for the estimated life expectancies (LEs) for men aged 60 and 70 with eight years of Parkinson's disease at baseline (Model 1).

	For LEs men aged 60		For LEs men aged 70	
	$\sqrt{v_M}$	$\sqrt{v_A}$	$\sqrt{v_M}$	$\sqrt{v_A}$
e_{11}	0.907	0.888	0.493	0.478
e_{12}	0.511	0.497	0.353	0.338
e_{21}	1.037	1.028	0.496	0.483
e_1	0.890	0.871	0.457	0.444
e_2	0.520	0.506	0.356	0.344
e_{tot}	0.947	0.925	0.526	0.508

D Bayesian χ^2 test for validation

Bayesian χ^2 test for discrete random variables (Johnson, 2004, p. 2366) could be applied using a χ^2 test that is a sum of a series of χ^2 tests. Observation times other than death, take place at years $t_0, t_1, ..., t_6 = 0, 4, 8, 9, 10, 12$. For example, for t_1 we have a discrete distribution for individuals in state 1 at t_0 with sample space state 1, 2 and 3, and we have a discrete distribution for individuals in state 2 at t_0 with sample space states 2 and 3. This would induce two χ^2 tests. In total, for $t_1, ..., t_6$, there would be $6 \times 2 = 12$ χ^2 tests with a total sum of 18 degrees of freedom. But as can be seen from the observed frequencies reported in the paper, data are sparse for the later time points regarding individuals previously observed in state 1. Collapsing the cells for state 2 and state 3 for t_3, t_4, t_5 and t_6 regarding individuals previously observed in state 1, would bring the degrees of freedom down to 14. However, the quantile-quantile plot (not reported) of the statistic shows considerable

deviation from the 45-degree reference line. The test cannot be used in our situation. This is not a complete surprise as the definition of the test relies on asymptotic properties regarding sample size and number of bins - both of which are not met in our situation.